HOW STUDENTS´ DISCIPLINARY ATTITUDES AND BELIEFS AFFECT LEARNING IN INTRODUCTORY STATISTICS COURSES

Dissertation

for Conferral of a Doctoral Title

by the Faculty of Social Sciences

at Georg-August-University of Goettingen

submitted by

Florian Berens M.A., M.Ed.

born in Nordhorn

Goettingen, 2024

First assessor:

Prof. Dr. Steffen-M. Kühnel, Chair of quantitative Methods in the Social Sciences, University of Goettingen

Additional supervisors:

Prof. Dr. Stefan Halverscheid, Chair of Mathematics Education, University of Goettingen

Prof. Dr. Thomas Kneib, Chair of Statistics, University of Goettingen

Date of the oral examination: August 09, 2024

Publication of this Thesis unter DOI:10.53846/goediss-10964

Abstract:

Attitudes and beliefs are frequently studied concepts in statistics education research. A main reason for this is that both are associated with learning success. However, very little is known about the mechanisms that induce these relationships between beliefs about statistics, attitudes towards statistics and learning success. Some studies find a mediating role of self-perceived learning engagement and applied learning strategies. However, studies finding such a relationship with more objective behavioral measures than self-reports are lacking.

To provide this evidence, this thesis first develops a conceptualization and a measurement instrument for beliefs about statistics. To objectively record learning behavior, it develops a digital learning platform that in particular is designed for a scientific use of the resulting digital behavioral traces, tests their operationalization in the field and documents the connection between the measured constructs and learning success. Following these preliminary studies, this thesis tests numerous structural equation models to estimate the mediating effect of various learning behavior dimensions.

The main results show that the association between attitudes towards statistics and learning success is indeed partially mediated by learning engagement and the distribution of learning. However, a smaller part of the association is also due to a spurious correlation that can be explained by the high-school graduation average. Such mediating relationships cannot be directly identified for beliefs about statistics. However, there is a relationship between beliefs and attitudes, so that beliefs are also linked to learning behavior through this mediation.

Further analyses indicate, however, that attitudes towards statistics (can) change during a statistics course. This suggests further research into how current attitudes are related to current learning behavior. At the same time, this limitation strengthens the relevance of the effect found, as already the initial attitude is linked to learning behavior throughout the course. A further investigation in addition shows that the relationships found do not necessarily remain stable when interventions are made in teaching, which demonstrates that intervention studies should always investigate all possibly induced effects.

Abstract (in German):

Einstellungen und Überzeugungen sind vielfach untersuchte Konzepte in der Didaktik der Statistik. Grund dafür ist insbesondere, dass beide mit Lernerfolg verbunden sind. Über die Mechanismen, die diese Beziehungen zwischen Überzeugungen über Statistik, Einstellungen zu Statistik und Lernerfolg induzieren, ist allerdings sehr wenig bekannt. Einige Studien finden dazu eine mediierende Rolle des selbstwahrgenommenen Lernengagements und der angewandten Lernstrategien. Studien die eine solche Beziehung mit objektiveren Verhaltensmessungen als Selbstberichten belegen, fehlen allerdings.

Um diesen Beleg zu führen, entwickelt diese Arbeit zunächst eine Konzeptionalisierung und ein Messinstrument für Überzeugungen über Statistik. Zur objektiven Erfassung von Lernverhalten entwickelt sie eine digitale Lernplattform, die insbesondere für die wissenschaftliche Verwendung der entstehenden digitalen Verhaltensspuren ausgelegt ist, erprobt deren Operationalisierung im Feld und dokumentiert den Zusammenhang zwischen den gemessenen Konstrukten und dem Lernerfolg. Nach diesen vorbereitenden Studien testet diese Arbeit zahlreiche lineare Strukturgleichungsmodelle, um die mediierende Wirkung verschiedener Lernverhaltensdimensionen zu schätzen.

Die Hauptergebnisse zeigen, dass die Beziehung zwischen Einstellungen zur Statistik und Lernerfolg tatsächlich teilweise durch das Lernengagement und die Verteilung des Lernens mediiert wird. Ein kleinerer Teil der Beziehung geht aber auch auf eine Scheinkorrelation zurück, die durch den Abiturdurchschnitt aufgeklärt wird. Für die Überzeugungen über Statistik lassen sich solche mediierenden Beziehungen nicht direkt feststellen. Es zeigt sich aber eine Beziehung zwischen den Überzeugungen und den Einstellungen, sodass durch diese Mediation auch die Überzeugungen mit Lernverhalten verbunden sind.

Weitere Analysen weisen aber darauf hin, dass die Einstellungen zur Statistik sich im Kursverlauf verändern (können). Dies legt weitere Forschung dazu nahe, wie die jeweils aktuelle Einstellung mit dem je aktuellen Lernverhalten in Beziehung steht. Zugleich stärkt diese Limitation aber die Bedeutung des gefundenen Effekts, da bereits die initiale Einstellung mit dem Lernverhalten im gesamten Kursverlauf in Verbindung steht. Eine weitere Untersuchung zeigt zudem, dass die gefundenen Beziehungen bei Intervention in die Lehre nicht zwangsläufig stabil bleiben müssen. Dieser Umstand mahnt, in Interventionsstudien immer alle möglicherweise induzierten Effekte zu untersuchen.

Content

List of Figures

List of Tables

1 First Part: Introduction: Rationale and Outline of This Dissertation: Non-cognitive Factors in Statistics Education

1.1 Introduction

In 1994, Gal and Ginsburg (1994) reflect on the then current state of statistics education research in a review article. They conclude that

> *“while statistics educators have focused on improving the cognitive side of instruction, i.e., the skills and knowledge that students are expected to develop, little regard has been given to non-cognitive issues such as students' feelings, attitudes, beliefs, interests, expectations, and motivations” (Gal & Ginsburg, 1994, p. 7)*

As a reason why non-cognitive factors of learning (NCF) should have a larger role in statistics education research, Gal and Ginsburg theorize that non-cognitive factors influence both students´ learning behavior and learning success. In 1997, Gal et al. (1997) add that non-cognitive factors not only affect learning success in a statistics course, but also influence whether students will take voluntary follow-up courses in statistics and whether and/or how much students are inclined to use statistics in their later careers and everyday lives.

Because the importance of non-cognitive factors this way extends beyond the individual statistics course, Gal and Ginsburg (1994) argue that non-cognitive factors not only are conditional factors for successful statistics education, but declare the improvement of students' non-cognitive factors to be a goal of statistics education itself. This claim, in turn, implies that non-cognitive factors are not immutable and consequently raises the question of how and under which conditions non-cognitive factors change. Closely related again is the question on the sources of the students' existing attitudes, beliefs, emotions, etc. Thus, non-cognitive factors influence learning, but are also themselves influenced (by learning). Gal and Ginsburg (1994) use this dual relationship to call for more research on the sources, transformations, and consequences of non-cognitive factors in statistics education research, especially since the approaches they describe are theoretically developed, but not empirically grounded.

In addition to Gal and Ginsburg´s (1994) call for increased research on non-cognitive factors , the paper established the term "non-cognitive factors" and the way to define it in statistics education research. Looking at the educational sciences as a whole, there is no consensus on what should be included under the term non-cognitive factors. While definitions are rarely truly disjunctive, they differ considerably in the breadth of what is included as a non-cognitive factor. Some authors and disciplines in the educational sciences subsume all characteristics of a person under the term non-cognitive factor that do not directly describe the person's cognitive abilities but can nevertheless have an impact on academic performance. As a consequence, besides e.g. motivation, attitudes and emotions, also e.g. soft skills or social capital are included in the term non-cognitive factor (see e.g., Sommerfeld, 2011).

Other authors and disciplines in the educational sciences have a narrower understanding of the term non-cognitive factors. They exclude, for example, all skills and (experiential) knowledge, as well as all characteristics that do not relate exclusively to the person themselves, but to their social environment. Thus, non-cognitive factors are personal characteristics that are not (purely) cognitive, but have an affective component as well (see e.g., Rosen, Glennie, Dalton, Lennon, & Bozick, 2010).

Statistics education researchers agree with this view and have - primarily on the basis of Gal and Ginsburg´s article (1994) - developed the understanding of non-cognitive factors as an umbrella term for concepts such as motivation, attitudes, beliefs, emotions, epistemological beliefs, and other individual personality traits (see quote above), which are characterized by beeing distinguishable from cognitive factors, such as intelligence, on the one hand, and situational feelings on the other (Chiesi & Primi, 2010; Gal & Ginsburg, 1994; Spencer, Griffith, Briska, Post, & Willis, 2023). Non-cognitive factors understood this way can contain cognitive components, but they always contain affective components as well. They thus move on a continuum between pure affect and pure cognition (Rosen et al., 2010). Other authors use the term affect as a synonym for noncognitive factors, as conceptualized for example by Hannula (2004, 2006).

As the most important non-cognitive factors in statistics education, Gal and Ginsburg (1994) identify attitudes, beliefs, and statistics anxiety. In this regard, they note that attitudes toward statistics are the most widely studied non-cognitive factor at their time. In particular, two measurement instruments for attitudes toward statistics available at the time, the SAS (Roberts & Bilderback, 1980; Roberts & Saxe, 1982) and the ATS (S. L. Wise, 1985), have been used widely to demonstrate that students in introductory statistics courses often have negative attitudes toward statistics. But also the high prevalence of statistics anxiety could be shown several times using the Statistics Anxiety Rating Scale (STARS) (Cruise, Cash, & Bolton, 1985). Regarding beliefs, on the other hand, Gal and Ginsburg (1994) identify little previous research but attempt to demonstrate that beliefs about statistics, understood as responses to the question "What do beginning students understand the term `statistics´ to mean?", are relevant to student learning in the classroom by describing own experiences and quoting own students.

However, Gal and Ginsburg (1994) are deficient in definitions of their terms; in particular, attitudes, beliefs, and statistics anxiety are neither defined nor distinguished from one another. Since this thesis, like Gal and Ginsburg (1994), focuses in particular on attitudes, beliefs, and anxiety, and since these are also not adequately defined in many other empirical studies, subchapter 1.2 of the first part of this thesis follows with definitions of the terms attitudes, beliefs and anxiety as they will be used in this thesis. Subchapter 1.3 then presents a review of the literature on attitudes, beliefs, and anxiety published after Gal and Ginsburg (1994) in order to derive the research question of this thesis.

1.2 Attitudes, Beliefs and Anxiety – Definitions and Delimitations

1.2.1 Attitudes

The term "attitudes" does not originate from educational research, but is to be understood in a more general sense and is therefore discussed in many fields as, for example, in psychology, political science, and sociology. In the field of health psychology, for example, one is interested in attitudes toward certain health-relevant behaviors such as smoking, or in public opinion research one is interested in the attitudes of eligible voters toward political parties or politicians (Aronson, Wilson, & Sommers, 2021). In this sense, empirical research interested in the attitudes of individuals has existed for a very long time. However, the concept of attitudes has only been conceptualized since the 1960s. In particular, two phases of definition can be distinguished (Briñol & Petty, 2012). Before the turn of the millennium, attitudes were usually understood as dispositions of persons. To this end, the seventies and eighties were characterized by a debate as to whether these dispositions refer to three or rather only two dimensions (Briñol & Petty, 2012). Some authors understand attitudes as a disposition toward an affect, cognition, or behavior related to a person, object, or idea (Breckler, 1984; Rosenberg & Hovland, 1960). Other authors separate action and, relatedly, the intention to act from the concept of attitudes and see attitudes more as a linkage of dispositions to affects and cognitions (Chaiken & Stangor, 1987).

In more recent conceptualizations of attitudes, the term is understood somewhat more liberally, in which the focus on the disposition and thus a focus on the consequence of the attitude is dropped, and attitudes are defined somewhat more generally as evaluations of persons, objects, or ideas that lie within an individual (Aronson et al., 2021; Petty, Briñol, Fabrigar, & Wegener, 2019). These evaluations have the possibility of leading to affects, cognitions, or behaviors, directly or indirectly, but they are considered attitudes independent of these consequences. These attitudes may be conscious or unconscious (Aronson et al., 2021).

Although current definitions of the concept attitudes, thus, differ somewhat from older conceptions, ideas about the underlying structures of attitudes within the definition have not changed significantly. The first property to be assigned to an attitude is valence (Petty et al., 2019). Attitudes can be positive or negative in their evaluation of their subject. Thus, they create a continuum from absolutely positive to absolutely negative on which the individual classifies its evaluation of the subject. However, such a classification in the continuum can also be omitted in favor of ambivalent evaluations.

A second property of an attitude, also independent of the explicit definition, is the simultaneous presence of both affective and cognitive elements in the evaluation. This property is connected with a much-discussed method of classifying different attitudes. Attitudes are thereby classified according to their source (Aronson et al., 2021; Chaiken & Stangor, 1987). In this context, this is understood to mean whether an attitude is rather derived from affective or from cognitive sources. For example, an attitude about the value of science would be considered a more cognitive attitude, since in evaluating their attitude an individual would usually evaluate the rational benefits and uses of science for society or the individual. At the same time, however, an attitude about the value of science would also contain affective elements in that the individual would, for example, include personal moments of happiness or frustrations that he or she attributes to science in the evaluation. An example of an attitude derived more affectively would, for example, be the individual's attitude about whether he or she likes mathematics. Here, too, cognitive considerations about the general usefulness of mathematics could play a role, but the majority of individuals will evaluate personal emotions in their contact with mathematics. Both examples show that the classification of different attitudes with regard to their main source may well depend on the individual respondent's subjectivity, but at the same time it is assumed that certain attitudes usually address one or the other source more strongly and are, therefore, to be assigned to one or the other category (Aronson et al., 2021).

Another dimension on which attitudes are classified is the strength of the attitude. Attitudes can be very strong, anchored, and stable, or they can be very soft, fluctuating, and unstable (Krosnick & Petty, 1995; Petty et al., 2019). Theory assumes that the strength of the attitude has an influence on the temporal stability and changeability of an attitude as well as on the probability that an attitude will be reflected in behavior. The strength of the attitude is related to the intensity with which the attitude is reflected affectively and cognitively in the individual.

Another approach to structure different attitudes lies in the classification of functions that these attitudes fulfill. Katz (1960) describes a utilitarian function in which attitudes provide practical benefits to the individual, a knowledge-related function in which attitudes help to structure new knowledge and to classify it into existing knowledge, a self-defensive function in which attitudes are intended to maintain self-confidence, and a value-expressive function in which attitudes represent general values and beliefs.

The sources of personal attitudes lie at different levels. While some research even finds correlations between attitudes and genetic predispositions (Aronson et al., 2021), a link that is thought to be mediated by more stable personality traits, the main source of attitudes probably lies in various (learning) processes. If affects and cognitions are the two main components on which attitudes are based, it is also these two that cause changes in attitudes. Cognitive processes that relate to the subject of the attitudes can, thus, cause a change in the attitude in addition to a change in the cognitive structure. Likewise, an affective experience related to the subject of the attitudes can change the attitude (Aronson et al., 2021; Petty et al., 2019). Overall, it can be concluded that new experiences with the subject of the attitudes are the main driver behind changes in attitudes.

Educational research largely refrains from using its own definitions of the concept attitudes, but rather restricts itself to grasping its own types of attitudes that are relevant to it. Many of these are attitudes of learners towards learning objects. These are often broken down into general or subject-specific categories and then combined into theoretical models of relationships between different attitudes or to other variables. Eccles et al. (1983), for example, suggest to distinguish expectancy-related and value-related attitudes. They further distinguish expectations about the difficulty of the learning object from expectations about one's own abilities in relation to the learning object. They also differentiate attitudes about the value of the objects of learning from the interest in the object of learning, the usefulness of the object of learning for the individual, and the value of achieving the learning goal for the individual. In a similar way, distinctions and systems of attitudes can be found at many other authors which are usually very purposeful in their construction oriented to own empirical research goals (e.g., Bandura, 1977; Deci & Ryan, 1985).

Particularly relevant for this thesis is the categorization of attitudes toward statistics developed by Schau, Stevens, Dauphinee, and Del Vecchio (Schau, 2003; 1995), who closely follow Ecccles' Expectancy-Value Theory (Eccles et al., 1983; Eccles & Wigfield, 2020).Schau et al. (Schau, 2003; 1995) define six dimensions of attitudes. Under the term Affect they subsume regularly occurring (positive) feelings towards statistics. Following Eckel's two main categories of expectancy, Schau et al. (1995) define difficulty as the personal perception about the general/abstract difficulty of statistics and cognitive competence as the perceived own ability to learn statistics. Regarding the value of statistics, Schau (2003) concideers the personal interest in statistics and the perceived value of statistics for the individual. As a final category of attitudes, Schau (2003) adds something to Eccles et al. (1983) considering the intended effort that individuals plan to invest in a statistics course.

Although there is criticism on this theoretical conception (Whitaker, Unfried, & Bond, 2022), as well as on the respective empirical measurement instrument Survey of Attitudes Toward Statistics (SATS) (Vanhoof, 2010; Whitaker et al., 2022), this thesis is oriented towards Schau's conception of attitudes. Guiding for this is that Schau's concept and especially also the measurement instrument must be considered as the most widespread in the field of statistics education research (Emmioğlu & Çapa Aydın, 2012; Whitaker et al., 2022). The desired connection of this thesis to the research described in chapter 1.3 therefore requires referencing back to Schau (2003; 1995).

1.2.2 Beliefs

Stronger than the concept of attitudes, the concept of beliefs faces the challenge that a variety of everyday understandings of the concept of beliefs are in circulation, some of which fit well with scientific conceptions of beliefs, but some of which are beyond the scope of scientific theorizing. Thus, from everyday life, phrases like "I believe that I can do this!" or "I believe that the weather will get worse." or "I believe that Washington was the first president of the United States." are well known. The use of the word "belief" is thereby commonly understood from the context, although already the three examples mentioned require different facets or understandings of the term. While the first statement presents itself more as an evaluative belief about the person himself, the other two are statements under uncertainty. The latter are also to be distinguished again into a first statement, which is to be regarded as uncertain in general and hence also for others, while the last statement is only perceived as uncertain by the person himself, while others understand Washington's presidency as a mere fact.

The discussion of the concept of beliefs within the scientific community is at least as heterogeneous. Even if one focuses on the discussion of the concept within educational research, there remains a great heterogeneity of ideas about how to conceptualize the term beliefs. In particular, different conceptions and definitions of beliefs differ in terms of the broadness of the term. Conceptions that take a particularly broad view of belief often place many affective concepts, such as attitudes, values, or self-efficacy expectations, under an umbrella term belief. Pajares (1992), for example, only distinguishes knowledge from beliefs and considers almost all concepts that contain an affective component as beliefs. This concept of beliefs then basically corresponds to the umbrella concepts of non-cognitive factors by Gal and Ginsburg (1994) or affects by Hannula (2004, 2006).

An equally broad but quite different definition of beliefs is given by Bar-Tal (1990), who understands all knowledge and additionally all subjectively perceived facts, all opinions and all faith convictions as beliefs. In contrast to Pajares (1992), he includes knowledge in his conception of beliefs and makes the cognitive component the sole criterion for deciding whether a concept belongs to the umbrella of beliefs. Thus he understands everything as belief that can be understood as cognitively dominated, while he excludes affectively dominated non-cognitive factors from his conception of beliefs (Bar-Tal, 1990).

For both, for very broad definitions of beliefs, which make them quasi synonymous with non-cognitive factors, and for broad definitions focusing on cognitive elements, it can be noted that they find application as a framework for theoretical considerations, but are difficult to use empirically, at least in their entirety. Empirical use then always refers to only a subset of beliefs. For broad definitions of beliefs, therefore, a categorization of types of beliefs is necessary. A step in this direction is taken by Op't Eynde, Corte, and Verschaffel (2002), for example, who pursue an equally broad definition of beliefs, but by categorizing beliefs into students' beliefs about a domain, students' beliefs about the self, and students' beliefs about the social context contribute to a more concrete idea about possible concepts behind the term beliefs.

Taking up this distinction, other authors formulate narrower understandings of the concept beliefs. In particular, they take up what Op't Eynde et al. (2002) call students' beliefs about a domain. Rokeach (1972), for example, was an early proponent of this idea and describes beliefs as a disposition toward a person or thing that has primarily cognitive, but to a lesser extent affective, components. Following this approach, V. Richardson (1996) also describes beliefs as a construct between purely cognitive and purely affective concepts and additionally distinguishes attitudes and emotions within this spectrum. McLeod (1992) follows this approach and adds that beliefs are not only the most cognitive of these three concepts but also the most stable over time. This approach is also followed by Philipp (2007), who defines beliefs as "lenses that affect one's view of some aspect of the world or as dispositions toward action" (Philipp, 2007, p. 259). What remains important here is the high cognitive component of beliefs, which may even lead individuals who hold these beliefs to consider them as pure cognitions or knowledge. At the same time, however, beliefs retain an affective component because they are merely subjective and do not cross the boundary into objectively shared knowledge (Philipp, 2007; Rokeach, 1972).

Even more narrowly defined beliefs continue to exhibit a certain broadness, which is used in various ways in the literature to structure different sorts of beliefs. For example, beliefs are structured according to how central or poriferous they are anchored in the person (Törner, 2002). Furthermore, the object or its size can be used for classification, e.g. global beliefs, domain-specific beliefs, and beliefs about concrete content are conceptualizable. As a third dimension to distinguish beliefs, Törner (2002) describes how general or concrete they make a statement about their object.

If one introduces such classifications of beliefs, it follows that different beliefs on different levels of these classifications do not stand isolated, but are related to each other. If not only two beliefs are related to each other, but a larger cluster of beliefs forms an overall statement about an object, this cluster of beliefs is called a belief-system (Philipp, 2007; Törner, 2002). Some authors go even further and call a belief-system that combines a multitude of beliefs of different categories into an overall model about an object and makes it connectable to other cognitive or affective factors a "conception" (Philipp, 2007).

Because of the clear definitions and the emerging hierarchy of constructs, each of which is concrete and thus measurable, this thesis follows the definitional work of Philipp (2007). It conceptualizes beliefs in their narrower sense as lenses that determine a person's perspective on an object. Following Philipp (2007), complex systems of various beliefs of this kind are referred to as conceptions.

To follow Philipp´s (2007) definitions means to be compatible with many other empirical studies, which also refer explicitly to Philipp (e.g., Geisler & Rolka, 2021; Whitaker, 2020). At the same time, it can be stated that Grigutsch´s (1996) definition of worldviews and Op't Eynde et al. (2002) definition of disciplinary beliefs are so similar to Philipp´s (2007) definition of conceptions. For this thesis these terms are, therefore, used synonymously, as possible slight differences in details cannot be distinguished empirically.

1.2.3 Emotions and Statistics Anxiety

If it can be said of the term beliefs that it is not only used in science but also in everyday language, this applies all the more to the term emotions. In everyday language, there are many different ideas and conceptions of this term. Similarly, with regard to scientific literature, it can be stated that emotions are studied in many different ways and across many disciplines. This also means that it is even less possible to assume a uniformly understood definition than with attitudes and beliefs (Hannula, 2015; Patulny et al., 2019).

The only thing all approaches have in common is a strong affective element in the conceptons (Patulny & Olson, 2019; Pekrun, 2021).

In terms of educational research, McLeod (1992) describes emotions as primarily affective states of the individual that only contain a small proportion of cognitive elements. In contrast to attitudes and beliefs, he ascribes a high level of intensity to emotions, but at the same time a comparatively low level of stability over time. This general approach is usually followed by concepts developed in educational research (Pekrun & Stephens, 2010; Schukajlow, Rakoczy, & Pekrun, 2023).

What is important for understanding the role and significance of emotions in education, however, is more than their general definition, their systematization and their theoretical embedding in learning theories. There are numerous approaches to both aspects. However, Pekrun's work is probably the most important, as it combines comprehensive theoretical work on emotions (e.g. Pekrun, 2006, 2021; Pekrun, Frenzel, Goetz, & Perry, 2007; Pekrun & Stephens, 2010) with the development of measurement instruments (e.g. Bieleke et al., 2023; Pekrun, Goetz, Frenzel, Barchfeld, & Perry, 2011; Pekrun, Goetz, Titz, & Perry, 2002) and with empirical research on the role of emotions in learning (e.g. Pekrun, Lichtenfeld, Marsh, Murayama, & Goetz, 2017; Schubert, Pekrun, & Ufer, 2023).

To systematize emotions, Pekrun and colleagues (Pekrun et al., 2002; Pekrun et al., 2007) use three dimensions of differentiation: They consider the first dimension to be the reference point of the emotion. This can be either the activity itself or the outcome of the activity. As a second dimension, Pekrun and colleagues consider the valence of the emotion, i.e. how positive or negative it is. The third dimension looks at the consequence of the emotion and distinguishes between emotions that tend to have an activating effect (e.g. hope) and those that tend to have a deactivating effect (e.g. relief).

The appearance of these emotions Pekrun and colleagues, however, do not understood as a random process. Rather, various triggers favor the occurrence of these emotions or even directly cause their occurrence (Pekrun et al., 2002; Pekrun et al., 2007). Pekrun therefore developed the Control-Value Theory (CVT) to describe how different conditional contexts affect emotions (Pekrun, 2006; Pekrun et al., 2007; Pekrun, 2021).

In addition to modelling the origins of emotions, it is important that the control-value theory also conceptualizes why emotions are not just a volatile momentary state of affective feelings. Rather, control-value theory identifies cognitive decisions - and in particular the perceived control over the situation - as the most important conditioning factors for emotions (Pekrun, 2006; Pekrun et al., 2007; Pekrun & Stephens, 2010). This introduces a cognitive component into the primarily affective construct of emotions.

As a consequence, the measurability of emotions also changes. In addition to asking about the current emotional status (e.g. through experience sampling method), it is possible to use classical surveys to ask about frequently experienced emotions and the factors that determine emotions (Pekrun et al., 2017; Pekrun & Stephens, 2010). This allows the view of emotions to be freed to a certain extent from the permanent fluctuation of affects and transferred into permanent dispositions that regularly affect learning. For emotions understood in this way, numerous studies have shown that they have relevant effects on learning across different areas (e.g. Pekrun et al., 2017; Schubert et al., 2023; Schukajlow et al., 2023).

While various educational sciences have used different and varying numbers of emotions for such empirical studies, research in statistics education has so far essentially been limited to the emotion of statistics anxiety, which, according to Pekrun et al. (2007), is classified as a negative output-oriented activating emotion. As a phenomenon frequently observed by many lecturers in statistics, especially in service courses for other departments (Esnard, Alladin, & Samlal, 2021; Onwuegbuzie, 2004; Zeidner, 1991), statistics anxiety has been anchored as a concept in the domain of statistics education at least since Cruise et al. (1985) conceptualized it in a multi-facetted way and developed the measurement instrument Statistics Anxiety Rating Scale (STARS).

Although this instrument is still the most widely used instrument in statistics education research (Chew & Dillon, 2014; Esnard et al., 2021), the theoretical concept of statistics anxiety has evolved since then. For Cruise et al. (1985) and also a little later for Zeidner (1991) and Onwuegbuzie, Da Ros, and Ryan (1997), statistics anxiety is understood as any intense worry that occurs in connection to contact with statistics. More recent authors have criticized this general definition in particular for not distinguishing anxietie specific for statistics from anxieties that arise in a person regardless of the discipline and thus also, but not only, in contact with statistics (Baloglu, 2004; Chew & Dillon, 2014; Chiesi & Primi, 2010). For example, Cruise et al.´s (1985) fear of asking the statistics lecturer for help is criticized and subsequently excluded from the concept of statistics anxiety, as it seems plausible that there is an underlying general aversion to approaching lecturers/authorities, which is also evident outside of statistics (Vigil-Colet, Lorenzo-Seva, & Condon, 2008). Since there is at least no evidence against this objection, these components of anxiety are excluded from more recent conceptions of statistics anxiety and statistics anxiety now only encompasses those negative emotional states that are specific to statistics (Chew & Dillon, 2014; Vigil-Colet et al., 2008). However, the existence of such statistics-specific anxieties, which are different from general anxiety and math anxiety in particular, has been shown conceptually as well as empirically (Baloglu, 1999; Chew & Dillon, 2014; Gibeau et al., 2023).

1.2.4 Systematization of the Concepts Attitudes, Beliefs, and Statistics Anxiety

The discussions above of the concepts attitudes, beliefs and emotions show that these three concepts have a lot in common. Putting aside the fact that all three concepts are not uniformly defined across scientific disciplines, it remains that the definitions developed here, on which this thesis is based, have in common that attitudes, beliefs and emotions all incorporate cognitive and affective components. Understanding them as three levels of non-cognitive factors therefore does not mean that they are without cognitive components, but that they are not purely cognitive. McLeod (1992), among others, points out this fact and derives an overarching concept for attitudes, beliefs and emotions from it, which has since become the widely accepted basis for theoretical work in statistics education research and to a large extent also in mathematics education research.

McLeod (1992) describes how attitudes, beliefs and emotions can be differentiated in terms of stability, intensity, awareness and development. Beliefs and, to a lesser extent, attitudes are stable over time. Changes are gradual, if they occur at all. Emotions, on the other hand, are volatile and can change very quickly, even to a considerable extent. On the other hand, emotions are experienced more intensely and therefore enter consciousness more easily, while attitudes and, to an even greater extent, beliefs remain "cold". Attitudes are therefore rarely fully conscious, even if respondents find it easy to recognize them in a survey. Beliefs, on the other hand, usually remain subconscious, to the extent that their affective and subjective components also remain unconscious to many people. With regard to the development of attitudes, beliefs and emotions, McLeod (1992) assumes that beliefs develop slowly and therefore sometimes have origins dating back a long time, while emotions can arise quickly and in the moment. Attitudes lie between these two poles. Although they exhibit stability, they develop, at least potentially, in observable periods of time, especially when intensive interaction between the person and their object produces many new experiences that could be assimilated into the attitudes.

McLeod (1992) thus spans a continuum whose poles are stable, long-term, unconscious and unintensive on the one hand and volatile, short-term, conscious and intensive on the other (see

Figure 1). Beliefs lie close to the first pole, attitudes in the middle area and emotions close to the second pole. This categorization fits in with the insight that attitudes, beliefs and emotions all have cognitive and affective components, but with varying degrees of intensity. McLeod (1992) conceptualizes his continuum mentioned above also as a continuum between cognitive and affective character. Beliefs are close to the cognitive pole, but also contain something affective. Emotions, on the other hand, are close to the affective pole, but have cognitive elements. Attitudes, on the other hand, lie in the middle range and combine cognitive and affective elements in more or less equal parts, depending on the specific attitude dimension.

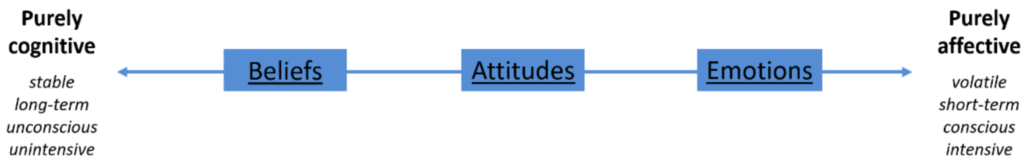

Figure 1: Beliefs, attitudes, and emotions on a continuum between cognitive and affective nature (own visualization based on McLeod (1992))

This thesis follows this conceptualization of levels of non-cognitive factors. This way, the full range of non-cognitive factors can be addressed, as called for by Gal and Ginsburg (1994) at the beginning, while the concepts are not listed unconnected as in Gal and Ginsburg (1994), but are conceptually interrelated. The model of these three levels of non-cognitive factors thus helps to examine attitudes, beliefs and emotions not only individually, but in combination. Although this combination will not be possible in every chapter and in every study in this thesis, it is the aim of several parts of the thesis.

1.3 Current State of Research and Deriving of Own Research Questions

After introducing the concepts of attitudes, beliefs and emotions and relating them to each other, the current state of research on all three of these concepts will be outlined. However, this overview does not claim to be a complete literature review on all research on attitudes, beliefs and emotions in statistics education. In particular, it only focuses on the attitudes, beliefs and emotions of learners. Non-cognitive factors of other stakeholders, especially those of teachers, are also studied widely, but were not what Gal and Ginsburg (1994) had in mind and are not part of this thesis. Therefore, this review of literature on attitudes, beliefs and emotions in statistics education only looks at learners.

This thesis moreover argues that research on attitudes, beliefs and emotions of statistics learners can be grouped into four clusters: The first cluster combines work conceptualizing non-cognitive factors with work developing measurement instruments for these concepts. This includes all work that defines, describes and differentiates factors, incorporates them into theories or models and theoretical work on the relevance of these factors, as well as all measurement instrument developments, validations or adaptations. The second cluster consists of work that investigates the causes of the realizations of non-cognitive factors in learners. This includes different origins, antecedents, reasons and conditioning factors of the factors. The third cluster focuses on the consequences of the factors and the investigation of the mechanisms by which these consequences are achieved. Consequences for learning success as well as for learning behavior and for other non-cognitive factors are considered. The fourth cluster combines work that looks at changes in learners in factors across time and work that attempts to deliberately produce changes in factors through interventions.

These four clusters can be represented by the following broad general research questions:

> General Research Question 1:
> *Which non-cognitive factors can be conceptualized at the three levels attitudes, beliefs, and emotions? How can these concepts be measured?*
> *(GRQ 1a: **Conceptualization**, GRQ 1b: **Measurement**)*

General Research Question 2:
What origins, antecedents, causes and conditions can be identified that can explain a learner's realization in the non-cognitive factors?
*(GRQ 2: **Causes**)*

General Research Question 3:
What consequences do the non-cognitive factors have for learning, especially for learning success, but also for learning behavior and for other non-cognitive factors? Why do they have these consequences and which mechanisms help to explain how they achieve their effect?
*(GRQ 3a: **Consequences**, GRQ 3b: **Mechanisms**)*

General Research Question 4:
How and why do non-cognitive factors change over time? How can a change be consciously induced?
*(GRQ 4a: **Development**, GRQ 4b: **Manipulation**)*

The following overview over the literature is provided for attitudes, beliefs and emotions seperately and in order of the four general research questions. To this end, answers or attempts to answers that have so far been found are presented and it is identified which of the four research questions (or seven sub-questions) has so far been answered resp. which question has already been worked on. The research questions of this thesis are then derived from the remaining gaps. In this context, it is important to note that this thesis cannot provide a complete literature review for every non-cognitive factor. In addition to limiting the literature review to learners, this review is limited in such a way that it examines the extent to which an answer to the respective research question exists for the three levels of non-cognitive factors. If such research exists, it will not be presented in full, but will be illustrated by its key elements or by examples that are as representative as possible. Such an approach takes into account the limited space intended here and still allows to justify the selection of the most relevant research gaps. Specific literature reviews on the research questions arising from this work follow within the individual studies.

1.3.1 Empirical Research on Attitudes Toward Statistics

1.3.1.1 Conceptualization and Measurement of Attitudes Toward Statistics

Even at the time of Gal and Ginsburg (1994), there has already been some research on attitudes towards statistics. In particular, several authors had already worked on the measurement of these attitudes. The first were Roberts and Bilderback (1980). In response to similar developments in mathematics education research (e.g. Aiken, 1970, 1976; Dutton, 1951), they argue that instruments developed for mathematics need to be adapted for statistics in order to do justice to the very nature of statistics. They then modified Dutton´s (1951) scale on arithmetic for statistical purposes and published the result as the Statistics Attitude Survey (SAS) with a total of 34 items (Roberts & Bilderback, 1980). A little later, Roberts and Saxe (1982) provided further psychometric validation for the instrument. However, the Statistics Attitude Survey is one-dimensional and does not come with a conceptualization of different (dimensions of) attitudes.

S. L. Wise (1985) was the first to provide a categorization of attitudes with two dimensions. He criticizes Roberts and Bilderback´s (1980) measurement as being too related to learning success and not differentiating enough between a concrete statistics course and the field of statistics as an abstract entity. In his Attitudes Toward Statistics (ATS) instrument, he therefore measures *attitudes towards the course* (with 9 items) and *attitudes towards the field* (with 20 items) separately (S. L. Wise, 1985).

Even if this distinction between a concrete course and the abstract field is important and will also be discussed again below with Schau (2003), this does not provide the multidimensional view of attitudes towards statistics as a discipline that is actually of interest. The first to provide this was Auzmendi (1991). In her Multifactorial Scale of Attitudes toward Statistics (MSAS), she measures the five dimensions *motivation*, *enjoyment*, *anxiety*, *confidence*, and *usefulness* with a total of 25 items (Auzmendi, 1991). Auzmendi (1991) justifies the selection of these five dimensions of attitudes with their high relevance for statistics and statistics education. However, Auzmendi (1991) does not provide a theoretical model that would frame and further justify this selection or link it to other research or theories.

Such an integration of a model of attitudes towards statistics into an existing, established theory was first provided by Schau et al. (1995). They discuss Eccles' Expectancy-Value

Theory (EVT) (Eccles et al., 1983; Eccles & Wigfield, 2020) in detail and derive four dimensions of attitudes towars statistics: Following Eckel's two main categories of expectancy, Schau et al. (1995) define *difficulty* as the personal perception about the general/abstract difficulty of statistics ("Statistics is a complicated subject.") and *cognitive competence* as the perceived own ability to learn statistics ("I will find it difficult to understand statistical concepts."). Regarding the value aspect in Eccles´ EVT, Schau et al. (1995) ask about the perceived utility *value* of statistics for the individual ("Statistics is irrelevant in my life."). As the more affective component of Eccles´ value aspect in EVT they use the dimension *affect* to subsume regularly occurring (positive) feelings towards statistics ("I will like statistics."). To measure these four dimensions of attitudes towards statistics, Schau et al. (1995) propose the Survey of Attitudes Toward Statistics (SATS) with 28 items.

Schau later further developed the SATS and the underlying theoretical concept on her own. She argues that in the area of the value aspect of EVT, the very cognitively dominated evaluation of the abstract value of statistics and the very affectively dominated measurement of affect leave too much space between them. She therefore includes the dimension of personal *interest* in statistics ("I am interested in learning statistics.") in the conceptualization and (with four items) in the measurement of the SATS (Schau, 2003). In addition, she draws on Wise's approach of including the specific statistics course in the survey and considers the intended *effort* that individuals plan to invest in a statistics course ("I plan to work hard in my statistics course.") as a final category of attitudes (Schau, 2003). Her extended instrument comprises of a total of 36 items and is therefore referenced as SATS-36 (Schau, 2003). Later, Schau deepened the theoretical link between the effort dimension and the five EVT-based dimensions in the Model of Students' Attitudes Toward Statistics (SATS-M), in which Ramirez, Schau, and Emmioglu (2012) model the planned effort as a consequence of expectancy and value, which in turn has an effect on learning behavior.

The psychometric structure of the SATS-28 and SATS-36 instruments has since been tested many times and in different contexts. Many authors have confirmed very good fit for the four- respectively six-dimensional structure, e.g. Tempelaar, van der Loeff, and Gijselaers (2007), Coetzee and van der Merwe (2010) and Stanisavljevic et al. (2014). Other authors, including e.g. Cashin and Elmore (2005), Vanhoof, Kuppens, Castro Sotos, Verschaffel, and Onghena (2011) and Hommik and Luik (2017), find only three or four dimensions in their use of the instrument and therefore criticize the instrument and the associated conceptualization.

However, this criticism must acknowledge the fact that the three- or four-dimensional structures they found hardly contradict the basic concept of the SATS. The course-related effort dimension remains a separate dimension in all studies. Difficulty and cognitive competence, however, can not be distinguished by the aforementioned critics. However, these two dimensions are designed as sub-dimensions of the main category *expectancy* anyway. Not being able to separate these dimensions therefore does not necessarily indicate a fundamental problem, but shows that the two are actually related and belong to a meaningful main category. A similar argument can be made for value, interest and affect. These three belong to the main category *value* of EVT. The fact that not all three can be separated from each other in some studies on SATS therefore does not fundamentally argue against the concept of SATS. For this reason, also Nolan, Beran, and Hecker (2012) in their review on the validity and reliability of the instrument argue that the SATS should be left unchanged. In line with this argument, a preparatory study by the author of this thesis in the context in which the studies in this thesis will take place also finds that either a six-dimensional measurement with the SATS or a reduction to the three basic dimensions of *expectancy*, *value* and *effort* make sense and are psychometrically sound (Berens, 2018).

Since their publication, SATS-28 and SATS-36 can be considered the most widely used instruments for measuring attitudes towards statistics (Emmioğlu & Çapa Aydın, 2012; Whitaker et al., 2022). Their importance for the field of statistics education research is also underlined by the fact that no other instrument for measuring attitudes towards statistics has been published in the more than twenty years since their publication.

Only recently criticism of the SATS has emerged that goes beyond the evaluation of the factor structure, in particular raised by Whitaker et al. (2022). This research group has therefore been working for some time on a new and more comprehensive conceptualization of attitudes towards statistics (Batakei, Bolon, & Bond, 2018; Whitaker, Unfried, & Batakei, 2018) and is also developing corresponding measurement instruments (Unfried et al., 2022). However, final results of this work have not yet been published.

This thesis is therefore based on the conception of the SATS by Schau (2003) and thus indirectly on Eccles' Expectancy-Value Theory (Eccles et al., 1983; Eccles & Wigfield, 2020). With this decision, it ties in with the most theoretically elaborated as well as empirically most widespread form of conceptualization and measurement of attitudes towards statistics. Depending on the specific question and modeling, either the SATS-36 with all six dimensions is used or a reduction to the three core areas of expectancies, values and effort is made.

1.3.1.2 Causes of Attitudes Toward Statistics

Applications of SATS-28 and SATS-36 in statistics education research have included extensive research into the antecedents and causes of attitudes towards statistics. Gal and Ginsburg (1994) had already hypothesized in the paper discussed at the very beginning of this thesis that attitudes towards statistics are shaped in particular by experiences with mathematics (at school).

This assumption is by far the most studied topic in terms of the antecedents and causes of attitudes towards statistics. Experience with mathematics, however, has been interpreted very differently in this context. For example, some studies find a correlation between the last math grade and attitudes towards statistics (e.g. Carmona, Martínez, & Sánchez, 2005; Chiesi & Primi, 2010; Emmioğlu, 2011; Zimprich, 2012) in the direction that better grades in mathematics are associated with more positive attitudes towards statistics. In other studies, mathematical skills are assessed using standardized tests. Here, too, there is a positive relationship with attitudes towards statistics (e.g. Auzmendi, 1991; Roberts & Saxe, 1982; Sesé, Jiménez, Montaño, & Palmer, 2015).

Other authors consider more affective constructs with regard to previous experience with mathematics as causes for attitudes towards statistics. Studies that relate self-perceived ability in mathematics or perceived success in mathematics to attitudes towards statistics are still relatively close to the cognitive dimension. Here, too, the positive relationship is confirmed (e.g. Coetzee & van der Merwe, 2010; Dempster & McCorry, 2009; Stanisavljevic et al., 2014). Mills (2004), for example, goes even further in the direction of Bandura´s (1977) concept of self-efficacy expectations and also finds a positive relationship. Nasser (2004) and La Hernández de Hera, Morales-Rodríguez, Rodríguez-Gobiet, and Martínez-

Ramón (2023), for example, consider attitudes towards mathematics as a cause of attitudes towards statistics. The positive relationship between the two is reproduced there as well. Nasser (2004) also shows that anxiety about mathematics is negatively related to attitudes towards statistics.

Overall, this gives a pretty clear picture that performance, attitudes and experiences in mathematics are related to attitudes towards statistics. Even if in all these studies direction and causality of the relationship are difficult to prove, the fact that learners usually come into contact with mathematics much earlier than with statistics suggests that contact with mathematics may be a causal factor in attitudes towards statistics.

Despite the importance of mathematics for attitudes towards statistics, mathematics is not the only antecedent or cause of attitudes towards statistics. As a further bundle of causes for attitudes towards statistics, some authors deal with special characteristics of statistics itself. In Griffith, Adams, Gu, Hart, and Nichols-Whitehead (2012) and Hannigan, Gill, and Leavy (2013), for example, pre-service mathematics teachers state that statistics has a very different character than mathematics and that they therefore have more negative attitudes, and in particular find statistics more difficult. A possible specification of this different character is provided by Leavy, Hannigan, and Fitzmaurice (2013). Their interviewees see statistical thinking and statistical reasoning as thought processes that are different from those previously known, especially since uncertainty, room for interpretation and context play such an important role.

Hood, Creed, and Neumann (2012) add that a distinctive feature of statistics is its close connection to research methods as tools of science. They argue that many study programs include very general and abstract introductions to research methods, which many students find difficult and boring. These negative attitudes then rub off on attitudes towards statistics.

The author of this thesis takes up the argument of linking research methods and statistics in a preparatory paper for this thesis. Inspired by Lalonde and Gardner (1993), who see statistics as the language of science and research, Berens (2018) argues that attitudes towards research should be reflected in attitudes towards statistics, as the usefulness of the scientific language statistics increases if research is seen as useful. In fact, a slight correlation in this sense could be found (Berens, 2018). In addition, Yuhai Zhang et al. (2012) confirm that attitudes towards statistics improve with years of research experience.

In addition to these characteristics of statistics, some studies also show that characteristics of the statistics course and characteristics of the person are related to attitudes towards statistics. For example, Roberts and Saxe (1982) and Auzmendi (1991) show that the statistics instructor has an influence on attitudes towards statistics. Chiesi and Bruno (2021) find correlations between the big five components emotional stability, openness and conscientiousness and attitudes towards statistics.

Numerous studies find more negative attitudes among female respondents (e.g. Coetzee & van der Merwe, 2010; Roberts & Saxe, 1982; Zimprich, 2012). Interestingly, however, this has mainly been found outside the United States (Ramirez et al., 2012), while studies in the United States often find no gender difference (e.g. Carnell, 2008; Cashin & Elmore, 2005; Pierce, 2006; Watson et al., 2003). Differences between races in the United States are sometimes found (e.g. in Watson et al., 2003; Watson, Lang, Thomas, R., & Kromrey, 2002), sometimes not (e.g. in Pierce, 2006; van Es & Weaver, 2018). A migration background, on the other hand, has a negative effect (e.g. Ruggeri et al., 2008; Tempelaar, Gijselaers, & van der Loeff, 2006; Tempelaar & Nijhuis, 2007). Studies on the effect of age did not find any consistent results (e.g. Coetzee & van der Merwe, 2010; Scott, 2001; Verhoeven, 2009)

1.3.1.3 Consequences and Mechanisms of Attitudes Toward Statistics

Research on the consequences of attitudes towards statistics is almost exclusively limited to their correlation with learning success. S. L. Wise (1985) had already found a positive correlation between his Attitudes Toward Statistics and the final course grade in his instrument development. At the beginning of the millennium, for example, Finney and Schraw (2003), Cashin and Elmore (2005), Evans (2005), Sizemore and Lewandowski (2009), Chiesi and Primi (2010) and Emmioğlu (2011) came to the same conclusion. Other authors use standardized tests instead of the course grade, but also find a positive relationship between attitudes towards statistics and achievement (e.g. Chiesi & Primi, 2009; Dempster & McCorry, 2009; Estrada & Batanero, 2008; Tempelaar et al., 2006).

In their meta-analyses of the relationship between attitudes towards statistics and achievement, Emmioğlu and Çapa Aydın (2011, 2012) identify 17 studies that have investigated this relationship as of August 2012. Of these, 15 found a significant positive relationship between attitudes and achievement (Emmioğlu & Çapa Aydın, 2011). A closer look at the four dimensions of the SATS-28 shows that the effects of attitudes on success vary in magnitude. Emmioğlu and Çapa Aydın (2012) find a Hedges g of .3 for *affect* and *cognitive competence*, which according to Cohen´s (1988) classic rule of thumb marks the boundary between a small and a moderate effect. The dimensions of *value* and *difficulty*, on the other hand, are significantly lower with effects of Hedges g = .21 and .20 respectively (Emmioğlu & Çapa Aydın, 2012). However, all four effects are highly significant. The effects of interest and planned effort were only examined in a few studies and could therefore not be considered in the meta-analyses, but are generally of a similar size as the four dimensions mentioned (Emmioğlu & Çapa Aydın, 2011, 2012).

In an analysis of variance across the studies, Emmioğlu and Çapa Aydın (2012) furthermore find that the effects of all four alalyzed dimensions of attitudes on success are significantly larger in studies conducted in the United States than in the other studies. In the eight studies conducted in the United States, the effect is on average about twice as large as in the other nine studies conducted in Italy, the United Kingdom, Turkey, Spain, Israel, the Netherlands, and Switzerland, with effects of .39 vs. .18 for affect, .39 vs. .19 for cognitive competence, .27 vs. .14 for value, and .26 vs. .13 for difficulty.

Since the meta-analyses by Emmioğlu and Çapa Aydın (2011, 2012), no further meta-analyses or systematic reviews have been conducted on the relationship between attitudes towards statistics and learning success. However, later studies do not deviate significantly from the results of Emmioğlu and Çapa Aydın (2011, 2012) (e.g. Abbiati et al., 2021; Kiekkas et al., 2015; Sesé et al., 2015; Stanisavljevic et al., 2014). However, some of these studies supplement the findings by indicating that attitudes at the end of the course shortly before the exam are even more predictive of exam success than attitudes at the beginning of the course (e.g. Chiesi & Primi, 2010; Whitaker et al., 2022). In some studies, changes in attitudes between the start and end of the course are even more predictive of exam performance (Chiesi & Primi, 2018; Whitaker et al., 2022).

One limitation of these fairly consistent findings, however, is that all the studies mentioned take a fairly short-term view, in which attitudes relate to success in the same course. The long-term influences of attitudes towards statistics are largely unexplored, although Gal et al. (1997) explicitly called for research on this in their review. Only Vanhoof et al. (2006) consider a longer period of time. They find that attitudes towards statistics at the beginning of a doctoral program in educational research correlate with the final grade of the doctorate to about the same extent as they correlate with the grade in the statistics course in the first semester of the doctoral program. This is an association of attitudes towards statistics with an outcome that is achieved almost five years later.

In addition to the relationship between attitudes towards statistics and achievement, some studies have investigated the relationship between attitudes towards statistics and the likelihood of dropping out of statistics courses. Both Zimmerman and Austin (2018) and Spencer et al. (2023) find higher drop-out rates among students with negative attitudes. This means that an increased dropout rate is one of the negative consequences of negative attitudes towards statistics. Lower short-term and long-term success and an increased drop-out rate are thus documented effects of negative attitudes towards statistics. However, whether negative attitudes towards statistics also reduce the enrollment in voluntary statistics courses and reduce the use of statistics in everyday and professional life, as Gal et al. (1997) assume, has not yet been shown.

Research on the mechanisms by which these consequences of attitudes towards statistics are achieved, however, is very scarce. Although both Expectancy-Value Theory (Eccles et al., 1983; Eccles & Wigfield, 2020) and Control-Value Theory (Pekrun, 2006; Pekrun et al., 2007) suggest changes in learning behavior as an explanatory mechanism for the relationship between attitudes towards statistics and success in statistics, little attention has been paid to this even in other domains of educational research outside of statistics education (for exceptions see Artino & Jones, 2012; Dabas, Muljana, & Luo, 2021; Dayupay, Velos, Go, Cababat, & Golbin, 2022; Yingbin Zhang et al., 2022).

Even fewer such studies are known within statistics education research. Lavidas, Barkatsas, Manesis, and Gialamas (2020) use the intuitive finding that engagement in statistics courses, for example demonstrated by regular attendance and active participation, has a positive effect on learning success. This has also been shown empirically (e.g. by Budé et al., 2007; Reyes, Brackett, Rivers, White, & Salovey, 2012). Lavidas et al. (2020) combine this finding with the Model of Students´ Attitudes Toward Statistics (SATS-M) (Ramirez et al., 2012) to develop a causal model in which prior achievement in mathematics and the four dimensions of attitudes towards statistics measured with the SATS-28 each have a direct effect on engagement and achievement. They find that *affect* and *cognitive competence*, controlled by perceived competence in mathematics, have a positive effect on engagement, which in turn has a positive effect on the final grade (Lavidas et al., 2020). This provides evidence that increased learner engagement may be a mechanism that can explain the positive relationship between attitudes and achievement. However, it must be said that only two self-report items were used to measure engagement here, which relate to the frequency of attendance and the activity during attendance. However, it is known from various studies that such self-reports correspond only slightly with other, more objective measures (Ye & Pennisi, 2022; Yingbin Zhang et al., 2022). Further research is therefore needed to validate this idea.

At first glance, B. D. Jones and Carter (2019) pursue a similar idea. They also link learners' attitudes regarding a statistics course to their engagement and this in turn to their success. As Lavidas et al. (2020), they find evidence for their model in their sample (B. D. Jones & Carter, 2019). Here too, however, engagement is measured by a small number of self-report scales (four in this case). Another problem with B. D. Jones and Carter (2019), however, is that they do not measure attitudes with one of the common instruments for measuring attitudes toward statistics, but use a self-developed instrument by B. D. Jones (2009), which has a very different conceptualization and a very different focus than the other instruments. In particular, it is not specific to domain characteristics, but primarily focusses the pedagogy of the statistics course.

Chiesi and Primi (2015, 2017, 2018), on the other hand, use the classic conceptualization of attitudes towards statistics and the SATS-28, but broaden the view of engagement a little. They do not look at the self-reported amount of engagement, but rather the self-perceived intensity, type and intention of engagement. They use Marton and Säljö´s (1976) approach of distinguishing between surface learning and deep learning with the addition of Entwistle and Ramsden´s (1983) conceptualization of strategic learning as a third category. Using self-reports through instruments from the ASSIST family of instruments (Chiesi, Primi, Bilgin, Lopez, & Del Fabrizio, 2014; Tait, Entwistle, & McCune, 1998), Chiesi and Primi (2015, 2017, 2018) find that students with negative attitudes towards statistics are more inclined to pursue surface approaches to learning, while students with more positive attitudes towards statistics are more inclined to pursue deep or strategic approaches.

Thus, this other focus on learning behavior also shows that attitudes towards statistics influence learning behavior and thereby have an impact on learning success. However, all studies only ever look at a specific aspect of learning behavior and have to rely on self-reports to describe it. This is a major limitation of previous research that needs to be adressed.

1.3.1.4 Development and Manipulation of Attitudes Toward Statistics

Research on the development of attitudes towards statistics has so far mainly focused on changes between the beginning and end of the students´ (first) statistics course. However, this has not yet resulted in a consistent picture. Some authors find more positive attitudes towards statistics on average after the end of the course compared to the beginning of the course (e.g. Chiesi & Primi, 2010; Herman & Kerby-Helm, 2022). Other authors, on the other hand, find more negative attitudes towards statistics on average at the end of the course (e.g. Gundlach, Richards, Nelson, & Levesque-Bristol, 2015; Schau & Emmioğlu, 2012). Still others find significant changes only in a few dimensions of attitudes towards statistics (e.g. Bateiha, Marchionda, & Autin, 2020; Cladera, Rejón-Guardia, Vich-I-Martorell, & Juaneda, 2021) or find no significant change at all (e.g. Bond, Perkins, & Ramirez, 2012; Whitaker et al., 2022).

As a possible explanation, Chance, Wong, and Tintle (2016) and C. Xu, Peters, and Brown (2020) examine the influence of the lecturer on the change in attitudes towards statistics. Chance et al. (2016) find a more favorable development with male and more experienced instructors. C. Xu et al. (2020) find an effect of the attitudes that the lecturers themselves have towards their course.

All these observations focus on course-wide averages. However, several authors note that this development of averages does not adequately reflect what is happening in statistics courses, because many students do change their attitudes towards statistics, but positive and negative developments cancel each other out when looking at the averages (e.g. Chance et al., 2016; Millar & White, 2014; Whitaker et al., 2022). Chance et al. (2016) note that in their statistics courses a tendency towards an increase in heterogeneity of attitudes towards statistics can be observed, in the sense that students who already hold positive attitudes towards statistics at the beginning have even more positive attitudes towards statistics after the course, while students who already hold negative attitudes become even more negative.

However, in-depth findings on the exact development processes and their reasons are widely lacking. This is also due to the fact that all the studies mentioned only look at two measurement points and are therefore blind to intermediate steps and the exact course of development. Only Kerby and Wroughton (2017) add a further measurement point in the middle of their course. They found only a few differences between the attitudes at the beginning and at the end of the course. However, the attitudes at the middle of the course are by no means an average between the initial and final measurement, but in some dimensions of attitudes towards statistics are very similar to the post-measurement and more extreme than pre- and post-measurement in others. For the value dimension, for example, there is a V-shaped progression in which a kind of shock in the first half of the course is followed by a slight recovery in the second half (Kerby & Wroughton, 2017).

The studies presented show that the development of attitudes towards statistics is neither a linear nor a monocausal process. However, while more detailed research into the causal structures has made little progress to date, many authors have attempted to cause a (positive) change in attitudes towards statistics through interventions or have investigated how given pedagogical differences in teaching affect attitudes.

Studies investigating different ways in which statistics courses are delivered include studies comparing purely online delivery of the course with a flipped classroom format and with traditional higher education teaching. The comparison of purely online teaching with traditional teaching shows an advantage for traditional teaching, which supports a more positive development of attitudes towards statistics (DeVaney, 2010; Gundlach et al., 2015). However, when comparing the flipped classroom method with more traditional teaching formats, there is no clear picture regarding possible advantages for attitudes towards statistics. While Carlson and Winquist (2011) find significantly better attitudes at the end of a flipped classroom format compared to traditional teaching, Gundlach et al. (2015) find that traditional teaching has a slight advantage.

Mixed findings also occur with regard to the effects of adding digital elements to traditional teaching. Suanpang, Petocz, and Kalceff (2004) supplement their traditional teaching for an experimental group with additional digital teaching material and digital exercises for individual and group work. They find a significantly more positive development of attitudes towards statistics in the experimental group. Wiberg (2009) replaces the execution of statistical procedures with a calculator with the use of software on the computer and thereby also achieves a positive effect on attitudes. Huynh (2018), on the other hand, lets students work on a computer with the digital learning environment "The Islands" (Bulmer & Haladyn, 2011) and finds very mixed effects with strongly negative effects on attitudes in some parts of the study. These findings show that digitalization is not good (or bad) per se, but that the underlying technology and implementation play a decisive role.

The same applies to non-digital teaching interventions. Nevertheless, the published research in this field paints a relatively clear positive picture overall. Posner (2011) shows that regular formative assessment and feedback can improve attitudes towards statistics. Herman and Kerby-Helm (2022) even achieved an improvement in attitudes through a question of the week. Smith (2017) uses an extensive gamification of her teaching, which includes a fictional accompanying story, challenges and interactive elements for the students, and also finds clearly positive effects on attitudes towards statistics.

Various other studies are more strongly inspired by learning theory. They rethink the entire curriculum of statistics courses in terms of constructivist learning and, for example, replace a strong teacher orientation with a strong activation of the students. Mvududu (2003) and Bateiha et al. (2020) find very positive effects for different types of activating teaching. Paul and Cunnington (2017) go in a very similar direction and completely redesigns teaching based on the Guidelines for Assessment and Instruction in Statistics Education (GAISE) (GAISE College Report ASA Revision Committee, 2016). They also find positive effects of this change on attitudes towards statistics (Paul & Cunnington, 2017).

However, it is not only the chosen pedagogical approach to teaching that can have an impact on attitudes towards statistics. Various authors show that the chosen content and the didactical approach can also have an influence on attitudes towards statistics. R. C. Jones (2019), for example, finds that embedding the statistical concepts to be learned in application-related contexts has a positive effect on learners' attitudes towards statistics.

Carnell (2008) and C. Xu et al. (2020) both find positive effects on attitudes towards statistics when dealing with real data rather than theoretical examples in statistics courses. Wiberg (2009), Bayer (2016) and Elder (2023) each find positive effects that project-based work has on attitudes towards statistics.

This positive picture for many pedagogical and didactical interventions tested is less evident when looking at interventions that are intended to directly address individual attitude components. While Acee and Weinstein (2010) achieve slightly positive effects on attitudes towards statistics with an intervention that uses examples to demonstrate the value of statistics to the students, Andrews (2021) can hardly achieve any effect with three different interventions to highlight the value of statistics.

Overall, therefore, a mixed picture emerges with regard to all the interventions tested. Many can improve learners' attitudes towards statistics, but quite a few cannot. It must also be noted that statistics education research lacks a larger theoretical framework and a strategy for selecting promising interventions. Rather, the interventions presented are ideas of individual researchers or derivations from various theories, in which the mechanism by which the hoped-for effect is achieved usually remains unclear. This also fits in with the finding that little is known about the reasons for observed changes in attitudes between the start and end of a statistics course. Further research should help to understand such mechanisms in order to design more targeted interventions.

1.3.2 Empirical Research on Beliefs about Statistics

1.3.2.1 Conceptualizations of Beliefs about Statistics

Looking at the three levels attitudes, beliefs and emotions, beliefs are the most recent field of research, certainly in the area of statistics education research, and also the one in which the least work has been done to date. The existing research is inspired by mathematics education research. Inspired by Schoenfeld (1985), among others, research into beliefs about mathematics gained greater popularity in the 1990s (e.g. K. Crawford, Gordon, Nicholas, & Prosser, 1994; Grigutsch, 1996; Köller, Baumert, & Neubrand, 2000; Pehkonen & Törner, 1999; Törner & Pehkonen, 1996).

In the early 2000s, this research was reflected in the field of statistics education research, where Reid and Petocz (2002), Gordon (2004) and Rolka and Bulmer (2005) each proposed a domain-specific conceptualization of beliefs about statistics.

Reid and Petocz (2002) use a phenomenographic approach to identify the beliefs of 20 first-year undergraduate students about statistics using in-depth interviews. They find six conceptions of statistics, which they classify on a spectrum between fragmented and inclusive. At the fragmented end of the spectrum, they find three conceptions of statistics that differ only slightly in their degree of fragmentation and sophistication. Some students see *statistics as individual numerical activities*, some see *statistics as the use of individual statistical techniques* and some see *statistics as a collection of statistical techniques*. As still slightly fragmented, but already somewhat more integrated, they categorize the view of *statistics as analysis and interpretation of data* and the view of *statistics as a way of understanding real-life using different statistical models*. As the most integrated form, they identify a view of *statistics as an inclusive tool used to make sense of the world and develop personal meanings*.

Gordon (2004) also uses a phenomenological approach and examines texts from an open-ended written survey of 279 psychology students. She identifies five categories of conceptions about statistics and follows Reid and Petocz (2002) in classifying them on a spectrum from fragmented to inclusive. Like Reid and Petocz (2002), she considers three groups to be particularly fragmented: A group for whom *no concept of statistics* could be identified at all (4% of her sample), one that understands *statistics as processes and algorithms* (24%), and a group that understands *statistics as a mastery of statistical concepts and methods* (33%). In the area of partial integration, she categorizes *statistics as a tool for getting results in real life* (25%). The highest category is *Critical Thinking* (3% of her sample). Gordon (2004) further finds that in these five groups there are different proportions of students who take their statistics course voluntarily or compulsorily, with voluntary participants tending to have more integrated beliefs about statistics.

Rolka and Bulmer (2005) took another approach to identify learners beliefs about statistics. They had 164 life science students in an introductory statistics course paint/draw/sketch their picture of statistics on a piece of paper. These pictures were then classified into five groups. Again, the groups follow a hierarchical structure from very simple to more advanced beliefs. On the very simplistic side of the spectrum are pictures of statistics that *identify the statistics course with statistics* and thus make statistics a thing (about 12% of their sample). A little more advanced is the perception of *statistics as a collection of numbers, tools, visualizations and statistical parameters* (about 46%). A view of statistics that *combines the tool character of statistics with use in a (real) context* is rated as somewhat more advanced (18%). As the most advanced form of beliefs about statistics, they consider images that show *statistics as a process of understanding in a complex world, of understanding relationships and presenting a unified picture of data* (19% of their sample). As the fifth group of beliefs Rolka and Bulmer (2005) consider some images that understand *statistics as a surreal world of their own or a game* (5% of their sample), but classify this to be outside their hierarchy of groups.

Thus, by the mid-2000s, there were three somewhat parallel and independently developed proposals for conceptualizations of beliefs about statistics, which have some rough similarities but also some differences. None of them can necessarily be considered superior based on method or breadth of research. A suggestion for a measurement instrument, on the other hand, was not available. Bond et al. (2012) react to this situation by choosing a mixed-methods design for their research on the relationship between beliefs and attitudes, in which they first qualitatively develop their own conceptualization. They survey 47 undergraduate students using a number of open and closed questions. In the phenomenographic analysis of their data, however, Bond et al. (2012) find a result that they perceive as very similar to Reid and Petocz´s (2002) conception. For the quantitative part of their research, Bond et al. (2012) therefore decide to switch completely to Reid and Petocz´s (2002) conception in order to be more compatible.

Despite this additional empirical evidence for Reid and Petocz´s (2002) conception, Özmen and Baki (2018) also decide to develop their own qualitative conception due to their different cultural context. They interview 50 undergraduate students from various disciplines in Turkey. From their data, they identify seven clusters of beliefs about statistics: *focus on inferential statistics* (21 students), *numerical or calculation-based focus* (22 students), *statistics as research* (5 students), *focus on use in daily life* (8 students), *statistics as probability* (5 students), *focus on some statistical terminology* (15), *focus on collecting, organizing, and analyzing data* (11 students).

Given this even greater heterogeneity of conceptual proposals, Justice, Morris, Henry, and Brondos Fry (2020) decide to develop their own conceptualization using a phenomenographic approach. This was done in parallel to the work on this thesis, which will be addressed in more detail in chapter 2 on study 1. For their study, Justice et al. (2020) survey 44 undergraduate students from different years of study and disciplinary backgrounds. They identify four clusters of beliefs and described them in metaphors of painting. By *pant-by-numbers*, they mean a view of statistics that understands statistics as a step-by-step execution of tests and procedures through which one can get to the final result. Uncertainty and applicability in the real world play no role here. As a second group, Justice et al. (2020) describe the *step-by-step instruction to paint*. These beliefs likewise focus on the execution of tests and procedures, but recognize uncertainty and imprecision in the results. However, real-world application hardly plays a role here either. In particular, it is this use of statistics that *realist painters* focus on. They use statistics in the real world to solve problems, but do not see any uncertainty or imprecision in it. *Picasso painters*, as the last group, also see statistics as a way of solving real-world problems, but always regard the solutions as uncertain and possibly containing flaws.

Like Reid and Petocz (2002), Gordon (2004) and Rolka and Bulmer (2005), Justice et al. (2020) see a scale from novice conceptions to more advanced conceptions in their approach. They regard the paint-by-numbers cluster as particularly novice and the Picasso painters as closest to the expert consensus. The other two clusters lie equally in between. As an innovation in the discussion about beliefs about statistics, Justice et al. (2020) introduce a two-dimensional model of their clusters. They describe the acknowledgement or non-acknowledgment of uncertainty as one dimension in which the various conceptions differ. As a second dimension, they describe the anchoring of statistics as solving real-world problems or on the other side of the scale its isolation from the real world.

This approach by Justice et al. (2020) is the first theoretical model of beliefs about statistics that goes beyond an evaluation of beliefs between novice and advanced. This new approach makes it possible to better understand learners' beliefs in their components and may in the future allow to address individual facets in interventions.

However, an instrument for a quantitative measurement of beliefs about statistics in the sense of this thesis is still not available.

1.3.2.2 *Causes and Consequences of Beliefs about Statistics*

This lack of a quantitative instrument considerably limits the possibilities for working on the other general research questions outlined above. Accordingly, there are very few studies that deal with the causes, consequences, mechanisms, developments or manipulation of beliefs about statistics.

Among the few exceptions are two studies that deal with mathematics, the experience of mathematics, and beliefs about mathematics as antecedents of beliefs about statistics (Findley, 2022; Findley & Kaplan, 2018; Hedges & Harkness, 2017). Hedges and Harkness (2017) focus more on the lived experiences with mathematics, especially at school, and on the pedagogy and didactics experienced in mathematics lessons. In their case reconstructions, they identify in particular the teacher and the didactical design of the lessons as influential factors for the development of beliefs about statistics. In his case reconstructions, Findley (2022; Findley & Kaplan, 2018) focuses more on the relevance of mathematical and pedagogical beliefs of individuals as antecedents of beliefs about statistics. In addition, he finds that different disciplinary backgrounds (in his case a degree in mathematics vs. a degree in the life sciences) have an influence on beliefs about statistics.

With regard to the research question about the consequences of beliefs about statistics, Bond et al. (2012) find the theoretically expected connection between beliefs and attitudes towards statistics. Not only do they find that attitudes are not distributed equally across the different belief groups, but Bond et al. (2012) actually find evidence that beliefs classified by them and Reid and Petocz (2002) as more advanced are associated with more positive attitudes than more novice beliefs are.

In a study that was conducted after most of the work on this thesis, the author of this thesis and colleagues find evidence that beliefs about statistics of incoming statistics majors influence how they appropriate statistics and in consequence how easy it is for them to integrate into their new field of study (Berens, Findley, Justice, & Kinson, 2023).

However, there are no studies to date on the effect of beliefs about statistics on learning success in statistics. Findings from mathematics education research suggest that such an effect exists (see e.g. K. Crawford, Gordon, Nicholas, & Prosser, 1998; Kaldo & Hannula, 2012; Köller, 2001; Liebendörfer & Schukajlow, 2017), but this has yet to be confirmed for statistics education.

Accordingly, there are no studies that take a closer look at the mechanisms by which beliefs about statistics could contribute to success. The author of this thesis is also not aware of any studies on the possible modifiability of beliefs about statistics through intervention. This shows that the lack of quantitative measurement instruments for measuring beliefs about statistics is a massive obstacle that is considerably slowing down progress in statistics education research on beliefs. Only such a development could allow comprehensive studies on the mechanisms and interventions that would be of high relevance for educational practice.

1.3.3 Empirical Research on Statistics Anxiety

1.3.3.1 Conceptualization and Measurement of Statistics Anxiety

While, as described above, (many) different dimensions of attitudes and beliefs have already been considered in the literature, research on emotions in statistics education is almost exclusively limited to the study of statistics anxiety. There are no subject-specific conceptualizations or measurement instruments for any other emotion. And even a subject-unspecific analyses of emotions in contexts of statistics education are very rare (for exceptions see e.g. Heinzman, 2022; Niculescu et al., 2015).At the level of emotions, this thesis follows this implicit classification of relevance and presents conceptualizations and empirical findings only for statistics anxiety.

Apart from a few forerunners, the topic of statistics anxiety found its way into statistics education research in the mid-1980s. In addition to the perceived prevalence of the phenomenon, this was probably due to the extensive literature on mathematics anxiety within mathematics education research at the time. In 1972, F. C. Richardson and Suinn had developed the Mathematics Anxiety Rating Scale (MARS), giving new impetus to an already existing debate on mathematics phobia (Dreger & Aiken, 1957), in which the prevalence, consequences and modifiability were now widely discussed (see for example Betz, 1978; Dew, Galassi, & Galassi, 1984; Hendel, 1980; Kelly & Tomhave, 1985).

Cruise et al. (1985) take up this debate, but change it in two ways: Firstly, they argue that a separate concept of statistics anxiety is needed, which is meaningfully distinguishable from mathematics anxiety, as they do not understand statistics as a sub-discipline of mathematics but as a neighboring discipline. Secondly, they argue that statistics anxiety is not a one-dimensional construct, but has several facets (Cruise et al., 1985). While mathematics education research has worked with mathematics anxiety as a one-dimensional construct for a relatively long time due to the use of the MARS, it has become common practice in statistics education research following Cruise et al. (1985) to take a multidimensional approach to statistics anxiety.

Cruise et al. (1985) propose a six-dimensional approach. They distinguish between *Interpretation Anxiety* (“I experience anxiety when interpreting the meaning of a table in a journal article.”), *Test and Class Anxiety* (“I experience anxiety when waking up in the morning on the day of a statistics test.”), *Fear of Asking for Help* (“I experience anxiety when I am going to ask my statistics teacher for individual help with material I am having difficulty understanding.”), *Worth of Statistics* (“I wonder why I have to do all these things in statistics when in actual life I will never use them.”), *Computation Self-Concept* (“I could enjoy statistics if it were not so mathematical.”), and *Fear of Statistics Teachers* (“Statistics teachers are so abstract they seem inhuman.”).

To measure these constructs, Cruise et al. (1985) propose the Statistics Anxiety Rating Scale (STARS) with a total of 51 items. Studies on the reliability, validity and robustness of the instrument are available from Hanna, Shevlin, and Dempster (2008) and from Papousek et al. (2012). While Papousek et al. (2012) confirm the overall structure of the instrument, Hanna et al. (2008) suggest some minor changes, but also confirm the robustness of the instrument's structure.

Criticism on the Statistics Anxiety Rating Scale, however, is provided on a conceptual level, for example by Vigil-Colet et al. (2008). They criticize that the STARS does not sufficiently distinguish statistics anxiety from attitudes towards statistics (as measured by worth of statistics), from attitudes towards mathematics (as measured in the computation self-concept), and from fear for authorities (as measured in fear of statistics teachers). They therefore propose the Statistical Anxiety Scale (SAS), which measures the constructs Examination Anxiety, Asking for Help Anxiety and Interpretation Anxiety with a total of 24 items (Vigil-Colet et al., 2008). Studies on the validity of this instrument are available from Steinberger (2020), Lavidas, Manesis, and Gialamas (2021), and O'Bryant, Natesan Batley, and Onwuegbuzie (2021), among others.

Vigil-Colet et al.´s (2008) criticism of Cruise et al.´s (1985) work has since been widely accepted in further research. Some projects have therefore worked with the SAS since then (e.g. Durak & Karagöz, 2021; Saidi & Siew, 2022), other authors remain with the STARS for reasons of compairability to pevious research, but only use the items on Interpretation Anxiety, Test and Class Anxiety, and Fear of Asking for Help, and in some cases also Fear of Statistics Teachers. Worth of Statistics and Computation Self-Concept are then excluded (e.g. Chew, Dillon, & Swinbourne, 2018; Chiesi & Primi, 2010; Esnard et al., 2021).

In addition to STARS and SAS, there are four other instruments for measuring statistics anxiety: the Statistics Anxiety Inventory (40 items, 2 dimensions) (Zeidner, 1991), the Statistics Anxiety Scale (10 items, 1 dimension) (Pretorius & Norman, 1992), an unnamed instrument by Zanakis and Valenzi (1997) (36 items, 6 dimensions) and the Statistics Anxiety Measure (44 items, 5 dimensions) (Earp, 2007). However, all four instruments are not based on any significant conceptual innovation. Presumably for this reason, and because there is less reference literature available for them, all four are not widely used.

This thesis follows the majority of the literature of recent years and will therefore use the STARS in a reduced form. This way, the highest compatibility with existing literature on statistics anxiety is achieved. Worth of Statistics is omitted due to its conceptual proximity to attitudes towards statistics, Computation Self-Concept is omitted due to its proximity to attitudes towards mathematics. Fear of Statistics Teachers is retained, as there is no empirical evidence to date that would confirm that this fear does not have a subject-specific element, but is instead absorbed by a general fear of authority.

1.3.3.2 Causes of Statistics Anxiety

One reason why the STARS (and the SAS) have been frequently used is that many authors have returned to the original conceptual argument of Cruise et al. (1985) and empirically tested the extent to which statistics anxiety can be separated from mathematics anxiety. Zeidner (1991), Baloglu (1999, 2004), and Gibeau et al. (2023), among others, find that although statistics anxiety correlates with mathematics anxiety, this correlation is so weak (below 0.5 in each case) that, in conjunction with the conceptual argument, a separation and separate consideration of mathematics anxiety and statistics anxiety makes sense.

Baloglu (1999) also looks at the comparison of statistics anxiety with general anxiety and again finds low correlations, which for him suggest a conceptual separation and at the same time a causal relationship between the two. Thus, Baloglu (1999, 2004) uses his findings on a separation of anxiety, mathematics anxiety and statistics anxiety to theoretically model that general anxiety and mathematics anxiety are conditions that causally favor statistics anxiety. He is thus an early, albeit not the first, advocate of analyzing the causes and origins of statistics anxiety.

A broader look at research on the causes and antecedents of statistics anxiety, i.e. a look at the general research question 2 mentioned above, reveals extensive literature, so that this research question, unlike in the case of attitudes and beliefs, is probably the most researched in relation to statistics anxiety. For an overview of the work on this cluster of research, Onwuegbuzie and Wilson (2003) suggest dividing the antecedents studied into three groups: Situational antecedents, dispositional antecedents, and environmental antecedents. Cui, Zhang, Guan, Zhao, and Si (2019) follow this approach and develop it a bit further. By situational antecedents, they mean all antecedents that are directly related to the situation, in particular previous experiences and emotions derived from them, such as mathematics anxiety. By dispositional antecedents, they mean all aspects that the individual brings to the situation, in particular their psychological structure. Environmental antecedents, on the other hand, relate to the interaction between the individual and their environment, including socio-demographic factors in particular.

With regard to situational antecedents, in addition to mathematics anxiety and general anxiety, previous experiences and cognitive abilities have been investigated as antecedents of statistics anxiety. As expected, positive experiences and cognitive resources (such as prior knowledge) have a mitigating effect on statistics anxiety, while negative experiences and a lack of cognitive resources increase statistics anxiety (see, e.g., Faber & Drexler, 2019; Keeley, Zayac, & Correia, 2008; Saidi & Siew, 2022; Wilensky, 1997). In addition, specifics of statistics courses have also been discussed as antecendents of statistics anxiety, such as the abstract nature of the subject matter and an associated perceived fast pace of the course (Chew & Dillon, 2014) and the inherent occurrence of uncertainty in statistics (Williams, 2013). Of course, the actual implementation of the statistics course and the lecturer also play a role (Lesser & Reyes III, 2015; Trassi, Leonard, Rodrigues, Rodas, & Santos, 2022; Williams, 2010).

Attitudes towards statistics, which are categorized differently in Onwuegbuzie and Wilson (2003) and Cui et al. (2019), have also been examined many times. There is agreement on a high correlation between the two. While some authors are agnostic regarding a possible causal direction of effect (e.g. Saidi & Siew, 2022), this correlation is usually interpreted as the effect of attitudes towards statistics on statistics anxiety (e.g. Faber & Drexler, 2019; Macher et al., 2013; O'Bryant et al., 2021), which fits in with the theoretical conceptualization from Chapter 1.2 in that it assumes that the more stable and cognitive level of attitudes is reflected in the emotion of anxiety. In addition to attitudes towards statistics, other attitudes are also considered in some cases. Lalonde and Gardner (1993) and Onwuegbuzie and Wilson (2003), for example, argue that learning statistics has not only connections to learning mathematics but also similarities to learning foreign languages and argue that this simultaneous demand for mathematical and language competencies makes statistics particularly difficult and therefore also requires positive attitudes towards language learning.

Dispositional antecedents of statistics anxiety are also studied widly in statistics education research. For example cognitive characteristics, general personality traits as well as anxiety and stress-related personal parameters are taken into account. Daley and Onwuegbuzie (1997), for example, found that intelligence has a negative effect on statistics anxiety. Hong, Chew, and Dillon (2014) find that neuroticism and extraversion increase statistics anxiety, while openness and agreeableness lower it. In Williams (2015), people with higher believe that worry is beneficial, those who take a negative approach to problems, and who utilize cognitive avoidance as a coping strategy show higher statistics anxiety. McIntee et al. (2022) find that cognitive emotion regulation strategies, and satisfaction of psychological needs affect the statistics anxiety of learners.

With regard to the environmeltal antecedents, statistics education research has so far mainly investigated gender and, to a lesser extent, age, race and cultural background. Regarding the influence of gender on statistics anxiety, many authors find higher statistics anxiety among female learners (e.g. Baloğlu, Deniz, & Kesici, 2011; Gibeau et al., 2023;

Ralston, 2020). However, it is not uncommon to find no significant difference (e.g. Bui & Alearo, 2011; Mji, 2009), and sometimes even higher statistics anxiety among male learners (e.g. Ralston, Gorton, MacInnes, Gayle, & Crow, 2021). Trassi et al.´s (2022) meta-analysis therefore shows no significant gender effect, but suspects that this may be due to cultural differences between countries and that in many Western societies female learners can be expected to have higher statistics anxiety.

Regarding age, no general effect can be found (Trassi et al., 2022). Regarding race Onwuegbuzie (1999) finds increased statistics anxiety in African Americans while Bui and Alearo (2011) find no difference for Hispanics compared to Caucasian peers. J. A. Bell (2008) finds higher statistics anxiety among international students in the United States and Liu, Onwuegbuzie, and Meng (2011) find differences between students in the United States and China in that interpretation anxiety and test/class anxiety are more pronounced in the United States, whereas fear of asking for help and fear of statistics teachers are more pronounced in China.

Further insights into the antecedents of statistics anxiety are discussed in the reviews and meta-analyses by Cui et al. (2019) and Trassi et al. (2022), among others. For this thesis, it should be noted that although not all questions about the antecedents of statistics anxiety may be answered, very comprehensive literature on the topic exists.

1.3.3.3 Consequences and Mechanisms of Statistics Anxiety

Much less literature exists on the above outlined research question 3 regarding the consequences of statistics anxiety and the mechanisms through which it operates. Most of this literature focuses on the effect of statistics anxiety on success in statistics courses, usually measured by the course exam(s). As expected, it has been shown that statistics anxiety has a negative effect on course success (e.g. Hanna & Dempster, 2009; Hoegler & Nelson, 2018; Paechter, Macher, Martskvishvili, Wimmer, & Papousek, 2017), which is also confirmed by Trassi et al. (2022) in their meta-analysis. Studies that do not use the course exam but validated measurement instruments for statistical reasoning also come to the conclusion that statistics anxiety has a negative effect on learning success (e.g. Saidi & Siew, 2022; Yusuf, Suyitno, Sukestiyarno, & Isnarto, 2019). Only Keeley et al. (2008) describe that particularly low levels of statistics anxiety can also have negative effects on learning success and therefore assume a curvilinear, hill-shaped relationship.

Apart from effects on learning success, only a few other consequences of statistics anxiety have been investigated to date. The second most frequently investigated consequence of statistics anxiety is procrastination. Here, for example, Onwuegbuzie (2004), Macher et al. (2013), and Paechter et al. (2017) find comparatively large effects of statistics anxiety on procrastination. This is also confirmed by Trassi et al.´s (2022) meta-analysis. In addition, Esnard et al. (2021) show how statistics anxiety influences the expectations towards a statistics course in many ways. In connection with this, Papousek et al. (2012) demonstrate that statistics anxiety considerably increases the experience of stress during a statistics course.

While at least some answers are available for research question 3a on the consequences of statistics anxiety, less can be said about the mechanisms that lead to it. Only the (self-reported) use of different learning strategies is considered by a few authors as a mediator between statistics anxiety and learning success. González, Rodríguez, Faílde, and Carrera (2016) consider the use of self-regulatory strategies and the use of deep processing strategies and find empirical evidence for their argument that the effect of statistics anxiety on learning success is mediated by the learning strategies used. Macher et al. (2013) also find a correlation between statistics anxiety and deep level learning strategies. Kesici, Baloğlu, and Deniz (2011) and Vahedi, Farrokhi, Gahramani, and Issazadegan (2012) also find correlations between statistics anxiety and the use of several learning strategies, but interpret the learning strategies as antecedents of statistics anxiety.

Cui et al. (2019) and Trassi et al. (2022) agree with this view and include all these works in their reviews under the discussion of possible antecedents of statistics anxiety. While the correlations found between learning strategies and statistics anxiety on the one hand and learning success on the other can indeed also be interpreted in the direction of learning strategies being an antecedent, this categorization ignores the potential of González et al.´s (2016) theoretical model to include the findings on the relationships of statistics anxiety with learning strategies in the discussion of the mechanisms of statistics anxiety. Thus, the classification of the findings in Cui et al. (2019) and Trassi et al. (2022) leads to the fact that the question of the mechanisms of statistics anxiety remains largely under the radar and receives little attention.

1.3.3.4 Development and Manipulation of Statistics Anxiety

With regard to the above-mentioned research question 4 on the change and changeability of statistics anxiety, there are no studies depicting developments of statistics anxiety across many time pints, but some ideas for reducing statistics anxiety have already been tested at various levels. At the level of teacher communication, for example, Lesser and Reyes III (2015) show that humor has a mitigating effect on statistics anxiety. Williams (2010) shows that immidiacy, i.e. teachers being open and approachable, can reduce statistics anxiety. Also, the level of the course and the form of delivery can make a difference. DeVaney (2010) and Hedges (2017), for example, find increases in statistics anxiety in online courses that offer little direct contact with instructors. Intepe and Shearman (2020) show that open consultation hour formats with low-threshold help can reduce statistics anxiety. The form of examination can also make a difference. Kapitanoff and Pandey (2018), for example, find that collaborative examination formats can reduce statistics anxiety.

However, not all interventions investigated to reduce statistics anxiety were successful. In addition to a possible publication bias at this point, Andrews (2021), for example, reports on three intervention projects in which the value of statistics was to be explained using practical examples, none of which were able to achieve a demonstrable effect.

In addition to these individual failed attempts to reduce statistics anxiety, however, it must be noted in particular that there is no conceptual framework that would theoretically show which types and levels of intervention might be effective. Thus, the studies mentioned remain isolated test balloons that may be based on encapsulated good ideas. However, it has not yet been possible to understand the mechanisms of interventions in statistics anxiety in such a way that generalizable findings or even a coherent further research program on interventions against statistics anxiety would have emerged. This may also be related to the fact that, as explained above, little is known about the mechanisms driving the consequences of statistics anxiety.

Thus there is a fundamental lack of insight into the engine room of statistics anxiety. Due to its considerable consequences for research questions 3 and 4, and as indicated to a lesser extent also for research question 2, the question about the mechanisms may be considered the biggest problem in researching statistics anxiety at present.

1.3.4 Deriving of Own Research Questions

The literature review in the previous chapter 1.3 first shows that extensive research on non-cognitive factors in statistics education exists. A very substantial part of this research was conducted after the call for more research by Gal and Ginsburg in 1994 cited at the beginning of this thesis. It seems that the most research has been conducted on statistics anxiety, followed at a small distance by attitudes towards statistics. In contrast, there are far fewer studies on beliefs about statistics.

This impression is supported by a look at statistics in Google Scholar. As a proximal indicator for the amount of existing research, the number of search results for the three levels attitudes, beliefs and statistics anxiety can be looked at. Based on the cut-off date of May 1, 2024, the most hits are indeed for statistics anxiety (search term: "statistics anxiety" OR "anxiety about statistics") with a total of 6930 hits, 6640 of which are younger than 1994. Google Scholar found 4950 hits for attitudes towards statistics (search string: "attitudes towards statistics" OR "attitudes towards statistics" OR "statistics attitudes"), 4680 of them more recent than 1994. In contrast, only 584 hits were found for beliefs about statistics, 557 of them more recent than 1994 (search string: "statistics beliefs" OR "statistics conceptions" OR "beliefs about statistics" OR "conceptions about statistics").

While these numbers of search results up to 1994 should certainly be taken with caution due to the limited digitization, this look at the search results on Google Schoolar firstly supports that a lot of research on non-cognitive factors in statistics education was conducted after the 1994 review by Gal and Ginsburg. Secondly, it supports the impression that statistics anxiety has been (slightly) the most studied, while beliefs about statistics remain in the background.

A following comparison of the chapters above on attitudes towards statistics, beliefs about statistics and statistics anxiety shows where these differences lie and which general research questions are currently pressing. The research questions and the resulting research agenda for this thesis are derived from this comparison.

A look at the first general research question on conceptualizations and measurements of non-cognitive factors shows that there is already some research on all three levels of non-cognitive factors. In the case of both attitudes towards statistics and statistics anxiety, this research has led to established conceptualizations and associated measurement instruments that dominate and shape further research. For the Survey of Attitudes Toward Statistics (SATS) (Schau et al., 1995; Schau, 2003) in the area of attitudes as well as for the Statistics Anxiety Rating Scale (STARS) (Cruise et al., 1985) and Statistical Anxiety Scale (SAS) (Vigil-Colet et al., 2008) in the area of statistics anxiety, it can on the one hand be stated that conceptual as well as psychometric criticisms of these measurement instruments have been raised. In the area of attitudes towards statistics, this has led to efforts to develop new measurement instruments (Unfried et al., 2022; Whitaker et al., 2022), but not to published work yet. No such efforts are known in the area of statistics anxiety, although the theoretical connection to domain-independent learning theories is even less given there. On the other hand, however, it must be noted that these instruments have been evaluated positively often and have resulted in extensive and productive use.

Knowing that other authors are already working on the redesign of measurement instruments and that at the same time measurement instruments are already available which, although not very good, work reasonably acceptable to well, the decision is therefore made for this thesis to take the SATS and STARS measurement instruments and not to strive to develop own instruments for attitudes towards statistics or statistics anxiety. This inevitably means that their conceptual background is also adopted to a certain extent. How exactly conceptualization and measurement instruments are used is discussed within the individual studies.

For the area of beliefs about statistics, the situation is quite different. As the overview of the literature has shown, there is currently no measurement instrument for beliefs about statistics, let alone a validated, widely used, established one. One reason for this is that there is no consensus on a conceptualization of beliefs about statistics. The first general research question about conceptualization and measurement described above must therefore be regarded as largely unanswered. As could be shown, this circumstance has the consequence that only a few answers could be found to the other research questions, as the necessary tools are lacking. If one wishes to conduct empirical research on the other research questions in relation to beliefs about statistics, the literature discussed shows that such work is difficult to implement without a quantitative measurement instrument. Since this thesis aims to work on all three levels of non-cognitive factors, it is almost inevitable that it will have to develop a measurement instrument for beliefs about statistics.

With regard to the conceptualization of beliefs about statistics, the use of an existing model would be theoretically thinkable. As the literature review has shown, previous phenomenographic descriptions of types have produced little more theoretical modeling than simply distinguishing between novice beliefs and more advanced beliefs. A theoretically somewhat broader model only emerged in parallel to this work (Justice et al., 2020). Therefore, apart from adopting a given set of types of beliefs, there would be little conceptualization to adopt.

Adopting an existing conceptualization for which no qualitative data is available also makes the development of a measurement instrument considerably more difficult, as the design of possible items depends solely on the published underlying study and its reception by the developers of the items. If own qualitative data are available, these can be used as a basis for many draft items. This increases the possibilities in item development and at the same time helps to stay close to the target population in the formulations of the items. This work therefore aims to develop its own conceptualization and a corresponding measurement instrument, whereby both parts naturally also reflect existing work. Research question 1 of this thesis therefore is general research question 1 from above specified to beliefs:

> Research Question 1:
> *Which beliefs about statistics can be conceptualized? How can these concepts be measured?*
> *(RQ 1a: **Conceptualization of beliefs**, RQ 1b: **Measurement of beliefs**)*

With regard to general research question 2 above, the literature review has shown that this has already been examined many times. The body of literature on the causes and antecedents of non-cognitive factors is probably the most extensive when comparing the four general research questions. This applies in particular to statistics anxiety. Numerous studies are available that examine very different causes, levels and groups of factors. There are also numerous findings at the level of attitudes towards statistics. Even at the level of beliefs about statistics, the area of causes is the largest, after attempts of conceptualization.

The fact that much research has already been done on the causes and antecedents of the non-cognitive factors certainly does not mean that all research questions on these topic have already been answered. The comparison of the investigated causes of attitudes towards statistics and statistics anxiety shows that the effects on both levels have not yet been examined for all these causes. There is furthermore a lack of analyses that would attempt to quantify the significance of the individual causes in comparison to each other. Nevertheless, due to the already comparatively extensive literature on the causes of non-cognitive factors, this thesis will refrain from investigating general research question 2 and its research area further.

While very extensive research has been conducted on the causes of non-cognitive factors, general research question 3 on the consequences and their mechanisms is clearly the least researched research question. For attitudes towards statistics and statistics anxiety, it has been shown in many studies that these are associated with learning success. This is also assumed for beliefs about statistics, as this correlation has been found in other disciplines, particularly in mathematics.

This is a highly relevant assumption, as accepting such a link between non-cognitive factors and learning success is the basis for many arguments as to why non-cognitive factors should be considered at all (e.g. Emmioğlu & Çapa Aydın, 2012; Gal et al., 1997). There are other reasons for the relevance of non-cognitive factors, for example that non-cognitive factors such as attitudes can themselves be an educational goal (see, for example, the competence definitions of the PISA studies, OECD, 2019a). However, the link to (classical) learning success as measured in a statistics test is certainly a very strong argument for the relevance of non-cognitive factors. Nonetheless, empirical evidence for such a connection between beliefs about statistics and learning success in statistics is still lacking in statistics education research. Whether this connection can be empirically proven will therefore be the first part of the second research question of this thesis.

As a second consequence of beliefs about statistics, an influence on attitudes towards statistics and thus an effect from one level of non-cognitive factors to the other, less stable and more affective level of non-cognitive factors is often assumed (Liebendörfer & Schukajlow, 2017; McLeod & McLeod, 2002). Bond et al. (2012) find evidence for this in a qualitative study. Since this thesis does not want to look at one level of non-cognitive factors alone, but is also interested in the interplay of non-cognitive factors, the question of the consequences of beliefs for learners' attitudes is the second part of the second research question. Research question 2 therefore deals with the overall consequences of beliefs about statistics and is thus a form of general research question 3a that is specific to beliefs:

> Research Question 2:
> *What consequences do the beliefs about statistics have for learning? What are the consequences of the beliefs for learning success? What are the consequences of the beliefs for the attitudes of the learner?*
> *(RQ 2a: **Consequences of beliefs for success**, RQ 2b: **Consequences of beliefs for attitudes**)*

Longer-term consequences of non-cognitive factors were only examined in a few cases with regard to attitudes towards statistics. There is therefore great potential for further research in this area. Due to the considerable time span required for data collection for such studies, this thesis refrains from such projects.

Other, more short-term consequences of non-cognitive factors were investigated primarily in the area of learning behavior. Some studies find effects for both attitudes towards statistics and statistics anxiety on learning engagement, learning organization and the learning strategies used. In most cases, these consequences of non-cognitive factors for learning behavior are understood as mechanisms through which the described effects on learning success are achieved.

Despite this research on mechanisms behind the effects on learning success, the area of mechanisms (GRQ 3b) is the least researched of all seven subcategories of the general research questions described at the beginning. The few existing results must in addition be viewed with extreme caution. They all rely on learners' self-reports of their learning behavior. While the quality of self-reports must be considered in principle, this applies all the more to self-reports about one's own learning behavior. Numerous studies in this area show that learners can be extremely poor at monitoring, reflecting and reporting their learning process (e.g. Tempelaar, Rienties, & Nguyen, 2020; Zhou & Winne, 2012). However, observations that are independent of learners' subjective self-reports and apply other, more objective measures to the learning process do not exist in the field of statistics education research, and only very few exist in other areas of educational research.

Research on general research question 3b, which investigates learning behavior with data sources other than self-reports, is therefore absolutely necessary. This work is dedicated to this goal. In research question 3, it aims to investigate the mechanisms, in particular the changes in learning behavior, that help explain why non-cognitive factors are associated with learning success. While RQ1 and RQ2 only refer to beliefs about statistics, this question is to be addressed for all three levels, beliefs, attitudes and statistics anxiety, partly in their interaction. Research question 3 of this thesis therefore corresponds with general research question 3b under the restriction that other forms of observation of learning behavior than self-reports are to be used:

> Research Question 3:
> *Which mechanisms, in particular which changes in objectively observed learning behavior, help to explain why non-cognitive factors are related to learning success?*
> *(RQ 3a: **Mechanisms of beliefs**, RQ 3b: **Mechanisms of attitudes**, RQ 3c: **Mechanisms of statistics anxiety**)*

The lack of understanding of the mechanisms by which the relationship between non-cognitive factors and learning success can be explained is also apparent when looking at the general research question 4, particularly in the intervention studies aimed at manipulating non-cognitive factors. Numerous intervention studies are available for both attitudes towards statistics and statistics anxiety, which attempt to achieve improvements in these areas. Even though there have been quite mixed successes, it is particularly evident, as shown above, that these interventions are not based on a superordinate concept that would explain which intervention could be promising when and why. Understanding the operating mechanisms behind the non-cognitive factors could be an important help here. This thesis therefore refrains from testing its own ideas for interventions, but aims to make a contribution to improving non-cognitive factors more effectively in the long term by identifying mechanisms behind the effects of non-cognitive factors.

With regard to the other part of general research question 4, the development of non-cognitive factors that occurs (unintentionally) in the course of learning, it can be seen that this has hardly been addressed at all for beliefs about statistics and also for statistics anxiety. Only in the case of attitudes towards statistics there is already some work, especially comparisons between attitudes before and after a statistics course. However, these comparisons also lack an understanding of the mechanisms behind the developments.

Since it has been shown in various studies that attitudes at the end of the statistics course are more predictive of learning success than those at the beginning of the course and that changes in attitudes are even more predictive of learning success (Chiesi & Primi, 2010, 2018; Whitaker et al., 2022), it could be important to understand the mechanisms behind the changes in attitudes in order to understand the mechanisms governing the consequences of attitudes. To complement the answers to research question 3 with the perspective of dynamic changes in attitudes and their importance for the consequences of attitudes, research question 4 of this thesis examines the mechanisms in the development of attitudes towards statistics. Research question 4 is therefore general research question 4a specified to attitudes and deepened by the aim of not only describing the development but also explaining the mechanisms behind the development:

> Research Question 4:
> *How do attitudes toward statistics change over time? What are the reasons and mechanisms driving this development?*
> *(RQ 4a: **Development of attitudes**, RQ 4b: **Mechanisms driving this development**)*

Since research question 4 is mainly selected for its potential to contribute to a more comprehensive picture with regard to research question 3, research question 4 is not addressed for beliefs about statistics and statistics anxiety in this thesis. Without reliable and replicated findings on the typical development of beliefs or statistics anxiety, it makes little sense to investigate the underlying mechanisms. To investigate developments of non-cognitive factors for their own sake is not an aim of this thesis.

Consequently, after a comparison of the current states of research on attitudes towards statistics, beliefs about statistics and statistics anxiety on the basis of the four general research questions, four research questions arise for this thesis. However, these research questions are not to be understood as four equally important objectives of this work. On the one hand, the results on the four research questions can, should and will be independent contributions to statistics education research that will advance research in the respective areas. On the other hand, the analysis has shown that by far the largest gap in research into non-cognitive factors in statistical education lies in the investigation of the mechanisms by which non-cognitive factors are linked to learning success. Research question 3 needs therefore to be seen as the main research question of this thesis: **This thesis aims to identify mechanisms that help explain why non-cognitive factors are associated with learning success** (see Figure 2)**.**

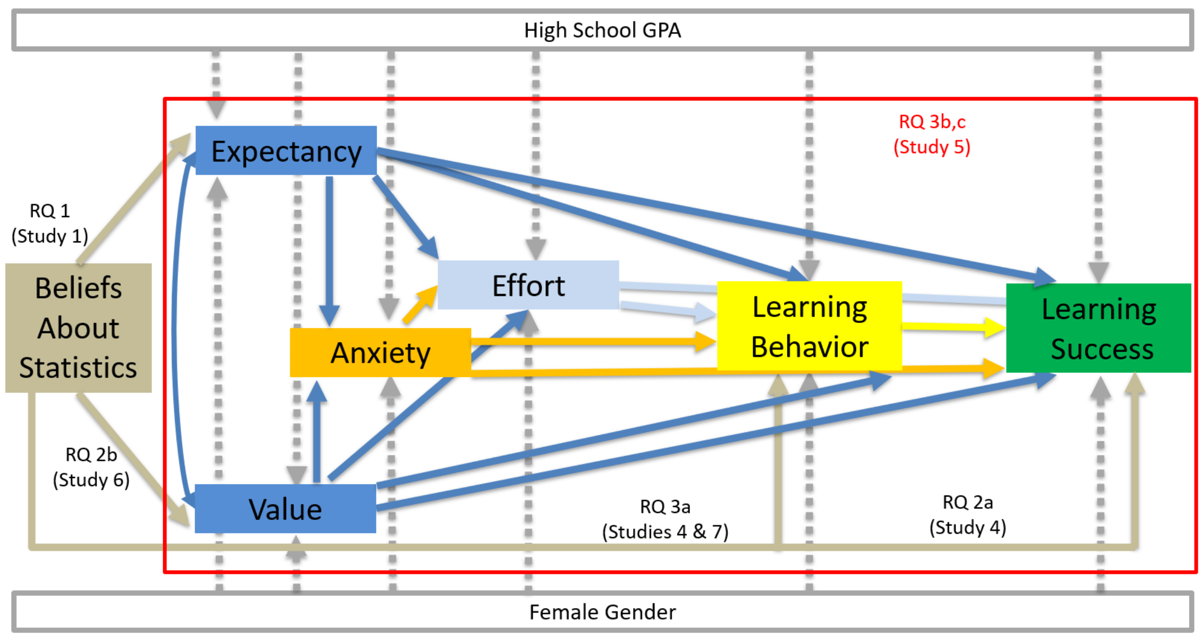

Figure 2: Theoretical model of this thesis and where it will be tested

Figure 2 serves as a hypothesis for these mechanisms based on the literature. The other three research questions support the thesis in this main objective. If the mechanisms that help explain why beliefs about statistics are associated with learning success are to be identified, such beliefs must first be described conceptually (RQ 1a) and made measurable (RQ 1b). The second step is to check whether consequences of beliefs for learning success can be determined at all (RQ 2a) or whether a bridge can at least be built via consequences of beliefs about statistics for attitudes (RQ 2b). If a suitable answer can be found to all these questions, research question 3a on the mechanisms that help to explain why beliefs about statistics are associated with learning success can be addressed in a meaningful way.

Research question 4 supplements the preliminary answer to research question 3 with a perspective of temporal dynamics in which, using the example of attitudes towards statistics, mechanisms of the development of non-cognitive factors are identified that can also be reflected in learning success.

1.3.5 Resulting Research Agenda

The four research questions in this thesis will largely be addressed in order. Study 1, which was conducted in cooperation with Kelley Findley, addresses research question 1 in chapter 2 of this thesis. Study 1 is divided into three major sections: The first section develops its own typology of beliefs about statistics and compares these in particular with the typology developed in parallel by Justice et al. (2020). The second section provides a theoretical framework for these types, resulting in a comprehensive conceptualization of beliefs about statistics. The two sections as a whole thus provide an answer to research question 1a about the conceptualization of beliefs about statistics. In the third section of Study 1, a suitable measurement instrument is developed and validated. This section thus provides an answer to research question 1b of this thesis. Overall, research question 1 is thus answered within study 1 to the extent that the main research question of interest (RQ 3) can be adequately addressed later.

This work on research question 3 is intended to be characterized by the fact that learning behavior as a possibly decisive mechanism is not to be observed by self-reports, but as objectively as possible. In this work, digital behavioral trace data is chosen as the way to achieve such an observation. Many studies have already shown that such data can be used to observe learning behavior comprehensively, in detail and comparatively objectively (e.g. Tempelaar, Nguyen, & Rienties, 2020; Ye & Pennisi, 2022; Zhou & Winne, 2012). How good such an observation through digital behavioral data can be depends decisively on the digital learning platform in which this digital behavioral data is generated. In order to have optimal data available for this purpose, a new digital learning platform is created in Study 2, which was conducted together with Sebastian Hobert, in chapter 3. This system is designed to generate added value for the learners in the context under investigation and at the same time to create optimal conditions for evaluating the digital behavioral data using learning analytics. The Stifterverband gratefully supported the development of this digital learning environment and honored it with an award for innovative higher education teaching and thereby made the development of such a system possible in the first place (grand number H120 5228 5008 32762, https://www.stifterverband.org/lehrfellowships/2018/hobert_berens).

Study 3 in chapter 4, which was also conducted together with Sebastian Hobert, serves to explore how the digital behavioral data from the system developed in study 2 can be processed, operationalized and used in research on student learning. It also provides empirical evidence using such data and the context of this thesis, that not only the quantity of learning but also its distribution is a relevant characteristic of learning behavior when it comes to learning success.

After studies 2 and 3 do not directly answer one of the research questions of this thesis, but prepare the work on the main research question, the answer to this main research question is begun in study 4. Study 4 in chapter 5 of this thesis examines the relationships between beliefs about statistics, learning behavior and learning success. In the first step, the relationship between beliefs about statistics and learning success is examined. This provides an answer to research question 2a of this thesis. Subsequently, learning behavior is examined as a mediating variable between beliefs about statistics and learning success. This provides an answer to research question 3a. As this study combines the work on beliefs about statistics with the use of digital behavioral data, this study was conducted in collaboration with both Sebastian Hobert and Kelly Findley.

Study 5 in chapter 6 is not only the middle, but in a way also the main point of this thesis. In study 5, which is again a joint work with Sebastian Hobert, a model is theoretically developed and empirically tested that simultaneously examines the extent to which engagement, learning organization and quality of learning as three dimensions of learning behavior are suitable mediating variables between learning success on the one hand and attitudes towards statistics and statistics anxiety on the other. Study 5 thus provides answers to research questions 3b and 3c and thus the main part of the answer to the main research question.

Nevertheless, this work cannot end with Study 5. One of the reasons for this is that Study 4 does not find any relationships between beliefs about statistics, learning behavior and learning success. In line with research question 2b, it must therefore be investigated whether at least an effect of beliefs about statistics on attitudes towards statistics can be found. In connection with the effects of attitudes towards statistics from Study 5, it could then be argued that beliefs about statistics do have effects on learning, but that these may be too small on learning success or too complex to be found in Study 4. Study 6 in chapter 7 therefore investigates the relationship between beliefs about statistics and attitudes towards statistics and was conducted together with Kelly Findley and Sebastian Hobert.

Study 6 finds individual correlations between beliefs about statistics and attitudes towards statistics. However, these correlations are rarely direct and linear, but mostly interaction effects or effects that only apply to a subpopulation of students. Study 6 thus provides an unsatisfactory and incomplete picture, although it can show that there are associations between beliefs about statistics and attitudes toward statistics. Study 7 therefore takes a closer look at the consequences of beliefs about statistics in three qualitative case studies in chapter 8. Co-authored with Kelly Findley and Nicola Justice, the study illustrates that beliefs about statistics are reflected both in attitudes towards studying statistics and in the approaches to studying. Thus, both research question 2b and research question 3a are enriched with examples that show a more comprehensive picture of how such effects can look like.

Study 8 in chapter 9 addresses research question 4 of this thesis. This sole-authored study uses a qualitative interview panel to investigate how attitudes towards statistics change over the course of an introductory statistics course and what reasons students give for these changes. The results not only provide answers to research question 4, but also reveal mechanisms of attitudes towards statistics that shed new light on research question 3b by revealing more subtle processes of change that are likely to play a role in the effect of attitudes towards statistics on learning success.

Chapter 10 contains Study 9, the last study of this thesis. In contrast to all other studies, Study 9 neither provides an answer to one of the four research questions nor does it prepare such an answer by laying necessary foundations. Study 9 was not even planned at the beginning of the work on this thesis and is therefore less an organic part of this thesis.

Rather, Study 9 arose from the sudden emergence of the Covid-19 pandemic during the work on this thesis and was intended to be a spontaneous reaction to the uncertainty that arose at the time regarding academic teaching under the conditions of emergency remote teaching. The study compares the first digital behavioral data collected with the system from study 2 in the summer semester of 2019 with the behavioral data from the summer semester of 2020, which was changed due to the pandemic, and combines this data with surveys on beliefs, attitudes and emotions. Even if this comparison of traditional teaching with emergency remote teaching does not directly address a research question of this thesis, its results do provide an interesting outlook. Under the conditions of the pandemic, in particular the emotional situation of students has deteriorated significantly, while at the same time learning behavior and learning success appear to have improved. This finding indicates that external disruptions, such as a pandemic, can greatly change the investigated system of relationships between non-cognitive factors and learning. The relationships between non-cognitive factors and learning therefore constitute a highly dynamic system of relationships.

After chapters 2 to 10 document the nine studies that were conducted for this thesis, chapter 11 provides an overall discussion of the results of this thesis. To this end, the results of the nine studies are first summarized. On the basis of these summaries, it is shown what contributions the studies were able to make to the four research questions of this thesis. As an overall conclusion to this subchapter, the overall contribution to the main research question of this thesis is discussed. The summarizing discussion of the results of this thesis is followed by a reflection on the ethical and data protection procedures used. This is followed by a consideration and discussion of the limitations of this thesis. Finally, suggestions are presented and discussed as to how further research could take up this work and lead to even deeper, more comprehensive findings.

For an overview over the course of this thesis, Table 1 provides a tabular summary of the studies of this theses, their methodological approach, the research questions answered and the data used:

Table 1: Overview over the studies of this thesis, their methodologies, data and research questions

	Short Title	Methodology	Course(s) from which the data originate	Question addressed
Study 1	Diamond Model	QUAL (GT) → QUAL + QUAN	Statistics I, II, III (2018, 2019, 2020, 2021) Computational Data Analysis Biostatistics (UIUC, USA)	RQ 1
Study 2	Digital Tutor	Works. → QUAL + QUAN	Statistics I (2019) Information Systems (2019)	
Study 3	Spaced Learning	QUAN (+ LA)	Statistics I (2020)	
Study 4	Beliefs and Learning	QUAN (+ LA)	Statistics I (2020)	RQ 2a
Study 5	Motivation and Learning	QUAN (+ LA)	Statistics I (2020)	RQ 3b
Study 6	Beliefs affect Attitudes	quan → QUAL	Statistics I (2020)	RQ 2b
Study 7	Beliefs and Identity Development	QUAL (Cases)	UIUC Statistics Mayors (since 2020)	RQ 3a
Study 8	Attitude development	QUAL (QCA)	Statistics I (2019)	RQ 4
Study 9	Learning during Covid	QUAN (+ LA)	Statistics I (2019, 2020)	

2 Study 1: The Diamond Model of Statistics: Framing and Measuring Students´ Conceptions about our Field

This chapter originated as a manuscript co-authored with Kelly Findley, University of Illinois at Urbana-Champaign. The author of this dissertation is first author of the manuscript. The manuscript is currently under review at a peer-reviewed journal. It is used and pre-printed in this dissertation with the kind permission of the publisher and the co-author.

Berens, F., & Findley, K. (Under review). The Diamond Model of Statistics: Framing and Measuring Students´ Conceptions about the Field.

Abstract: The conceptions students hold about a particular discipline often color their learning experiences in that discipline. While some work has examined the different ways students think about and represent the discipline of statistics, there is still work to be done to merge these conceptions with theories of cognition for data analysis. There may also be value in producing an instrument that can measure one's conceptions of statistics. In this paper, we undertake these contributions. First, we use a grounded theory approach to identify and describe four different conceptions of statistics. Second, we use models from two sources to represent these four conceptions theoretically, both as individual conceptions, and in relation to one another. In the third step, we present a validated instrument intended to measure one's relative agreement with each of these four identified conceptions. The instrument's items are available in the appendix and has shown suitability for students with at least some exposure to statistics at the college level or above. In our discussion, we also challenge the field to reflect on how different conceptions may represent the use of statistics in different contexts.

Keywords: Conceptions, Beliefs, Psychometric Instrument, Statistics Education

2.1 Introduction

As data grows more ubiquitous and central to our everyday lives, our field continues to consider the kinds of knowledge students need to succeed in data-driven careers (American Statistical Association Undergraduate Guidelines Workgroup, 2014; GAISE College Report ASA Revision Committee, 2016; Gould, 2010; Wild, 2017). According to the ASA's 2014 Curriculum Guidelines for Undergraduate Programs in Statistical Science, "Students need to see that the discipline of statistics is more than a collection of unrelated tools (or methods); it is a general approach to problem solving using data" (American Statistical Association Undergraduate Guidelines Workgroup, 2014, p. 6). A central goal of statistics education today is to offer this wider view of statistics to students—presenting statistical work as a cognitive process of investigation and inquiry through data (Bailyn, 1977; Horton & Hardin, 2015; Wild, Utts, & Horton, 2018).

But to advance this perspective of statistics in our classrooms, there is a need to understand the statistical beliefs and conceptions of our students. Beliefs and conceptions play a critical role in students' motivation for learning (Kloosterman, 2002), how students view disciplinary advancement and knowledge construction (Muis, 2004; Tsai, Jessie Ho, Liang, & Lin, 2011), and how students think about problem solving (P. Bell & Linn, 2002; Ozturk & Guven, 2016; Schoenfeld, 1992). Even with the same curriculum and the same teacher, students engage in tasks differently depending on the disciplinary beliefs and conceptions they bring to these contexts (Hammer, 1994; Presmeg, 2002). But there are ways to leverage and advance these beliefs in productive ways (Francisco, 2013; L. Mason & Scrivani, 2004). How do incoming students conceptualize the discipline of statistics, and how should we as educators leverage those prior frameworks?

While existing literature on student beliefs and conceptions in related disciplines can be insightful, there is much to be gained by exploring the topic from a distinctly statistical perspective (Hofer, 2006; Muis, Bendixen, & Haerle, 2006; Muis, Franco, & Gierus, 2011). Prospective learners of statistics begin their college statistics courses with substantively different impressions of statistics (Justice et al., 2020). Identifying the notions that students have of statistics can be valuable for educators and practitioners alike. By better understanding students' developing statistical conceptions, the field can better understand the conceptual progressions of newcomers and better prepare them for positions in statistics and data science.

In this paper, we review the existing research on students' statistical beliefs and conceptions and extend these findings with additional empirical data. We then use this data in combination with current literature to offer a theory-based framing of statistical conceptions. Finally, we present a validated instrument that assesses students' conceptions in statistics consistent with the model presented.

2.2 Background

Constructs like "beliefs" and "conceptions" are not always used consistently in educational literature. For this paper, we draw heavily on these constructs as defined by Philipp (2007). He defines beliefs as "lenses that affect one's view of some aspect of the world or as dispositions toward action" (Philipp, 2007, p. 259). These beliefs may then become more settled knowledge when they are specific enough to the individual to be known with relative certainty. Additionally, Philipp defines conception as "a general notion or mental structure encompassing beliefs, meanings, concepts, propositions, rules, mental images, and preferences" (Philipp, 2007, p. 259).

In this paper, we term beliefs as referring to the views and stances students hold toward specific elements in statistics, knowledge as the understandings they hold, and conceptions as the broader picture they have for the discipline. For example, a student might believe that a hypothesis test is a strict mathematical process resulting in an objective conclusion. That student may also have knowledge about how to complete a hypothesis test or how to interpret a p-value. These beliefs and understandings may then contribute to their broader conceptions of statistics. Their conceptions, however, will also be influenced by other topics, experiences, and the contextual spaces in which they encounter statistics. These various pieces form impressions that are brought forward when students encounter the discipline.

As one might expect, research has shown that students develop different beliefs and conceptions across different domains (Corte, Op't Eynde, & Verschaffel, 2002; Hofer, 2006; Muis, 2004; Muis et al., 2006). But researchers disagree on exactly how to frame domain-specific belief systems and even what terms to use for each notion (Hofer, 2006; Muis et al., 2006; Philipp, 2007). Beliefs has come to refer to a broad set of very different ideas; for example, students' epistemic beliefs as applied to a discipline (e.g., "Statistical knowledge is constructed", "Statistical knowledge is tentative") can be quite different than students' beliefs about learning (e.g., "I learn statistics best by applying theories to a context I understand") or beliefs about problem solving (e.g., "I think cross-validation is usually the best criteria to judge models").

Rather than study these more fine-grained beliefs, our research tackles the broader construct of students' statistical conceptions. In particular, we want to know what students think about the purpose of statistical work, what statistical experts do, and how statistical work proceeds. Since these two constructs are so often linked in research, we find it important to address how both have been studied and documented. Thus, to set up our work, we consider how disciplinary beliefs and conceptions have been explored in related disciplines like mathematics and science, as well as the limited research on the topic in statistics.

2.2.1 Belief and Conception Framings in Mathematics

A number of studies have identified different beliefs that students may express about mathematics. In her review of literature on students' mathematical beliefs, (Muis, 2004) finds many problematic beliefs that students commonly hold about mathematics: that mathematics problems may only have one correct answer, that mathematical answers reflect objective truths, that being "good" at mathematics is an inherent trait, and that formal proof does not require invention or discovery. Students may also hold more sophisticated mathematical beliefs. For example, in their respective classroom case studies, Francisco (2013) and L. Mason and Scrivani (2004), each describe what it looked like when students began to conceive of mathematics as a process of conjecture and sensemaking about structure and patterns.

These various spectrums of belief can be seen in the findings from two studies that investigated students' conceptions of mathematics. Both K. Crawford et al. (1994), as well as Törner and Grigutsch (1994) completed extensive interviews and instrument development in their work with university students. K. Crawford et al. (1994)identified five conceptions that can be summarized as follows: a) fragmented, b) fragmented with application, c) coherent system, d) coherent system with application, and e) coherent system that opens new perspectives on the world. While Törner and Grigutsch (1994) described only three views of mathematics (toolbox, system, and process), there are clear similarities between frameworks. The toolbox aspect represents mathematics as a static set of rules and formulas that culminated in calculations and black-and-white procedures. The system aspect represents mathematics as a set of existing proofs, definitions, and logical connections to be learned and applied. The process aspect is the most dynamic, as it conceives of the construction of mathematical objects and ideas.

Some themes from the literature in mathematics shows that more novice conceptions of mathematics largely see mathematics as static, disconnected procedures that may have isolated application. Intermediary conceptions begin to view mathematics as a cohesive system that may have application to the real world. More expert conceptions begin to see how mathematics itself is a constructed framework that does more than provide application, but rather provides a way of seeing and interacting with the world.

2.2.2 Belief and Conception Framings in Science

In contrast to mathematics, the term "conception" in science is rarely understood as a broad belief about the discipline, but instead as a notion toward a specific scientific concept. Broader views about the discipline of science are typically framed through the construct of "epistemology" or "nature of science." As one example, the Views of Nature of Science (VNOS) Questionnaire is a well-regarded framework of students' beliefs about science that is well connected to empirical findings (Lederman, Abd-El-Khalick, Bell, & Schwartz, 2002). The VNOS instrument probes the role of empirical evidence in constructing scientific knowledge, the tentativeness of scientific knowledge, the roles of creativity and theory-building in science, and the social and cultural influences in scientific advancement. In this instrument, each dimension is assessed from novice to expert conceptions using open-ended questions.

In addition to beliefs about science in general, some work has also looked at students' views in specific scientific disciplines. For example, Hammer (1994) discussed different perspectives that students may have about physics and how these perspectives foreground students' readiness to learn and conceptual understanding. Hammer noted that students' beliefs about the structure of physics could be characterized as a collection of isolated pieces, as a collection of concepts that had weak coherence, or as a coherent system of interlinking principles to make meaning.

In comparing beliefs research in science to that of mathematics, we see similarities with novice and expert framings. Novice framings focus on procedures, view disciplinary components as disconnected, and see learning as imparted from authority. More expert notions focus on concepts, view knowledge as a coherent framework based in logic, and see learning as constructed and requiring justification through conjecture and evidence. Key distinctions between mathematical and scientific framings of beliefs hinge on different foci on disciplinary work. Mathematical beliefs are more directly tied to perceptions of problem solving, and scientific beliefs to perceptions of knowledge.

2.2.3 Belief and Conception Framings in Statistics

The literature on beliefs and conceptions in statistics is comparatively sparse, but some tentative findings can be drawn. At a peripheral level, there are several studies that examined epistemic beliefs applied toward statistics by various groups (Diamond & Stylianides, 2017; Findley, 2022; Muis et al., 2011). For example, Muis et al. (2011) examined students' epistemic beliefs after different sequences of content, but did not directly probe students' statistical conceptions. Overall, these more epistemic-focused studies provide an interesting backdrop to framing statistical conceptions, but do not directly sketch out organized categories of conceptions.

We did find four studies that directly investigated students' broad conceptions of statistics and offered unique framings (Gordon, 2004; Justice et al., 2020; Reid & Petocz, 2002; Rolka & Bulmer, 2005), with one of these papers (Justice et al., 2020) being conducted concurrently, yet independently, to our work. In this section, we focus on the three papers that existed at the time of our own research, but return to Justice et al. (2020) when discussing our findings to point out how our respective works found key common ground.

Reid and Petocz (2002) completed interviews with 20 university students studying mathematics and asked open-ended questions on the purpose of statistics. The authors identified six conceptions in total. The first three were closely related and revolved around statistics as a static collection of techniques or numeric calculations that may (or may not) have usefulness. The fourth and fifth conceptions represented a focus on data, with one conception being statistics as the analysis and interpretation of data, and the other being statistics as a way of understanding real life using statistical models. The sixth was a combination of the previous conceptions, but with a primary focus on finding meaning with data.

Gordon (2004) posed similar questions as Reid and Petocz (2002), but through written responses with 279 students. She identified five categories of conceptions that map closely to the findings from Reid and Petocz. The first three categories (statistics as having no meaning, statistics as a thing to master, and statistics as procedural processes) can be seen as rather static, procedural views. The next category (statistics as a tool) focuses on the usefulness of statistics in applying methods to real world problems, and the third category (statistics as critical thinking) relates closely to Reid and Petocz' layer of statistics as making meaning.

Rolka and Bulmer (2005) collected drawings from 164 university students in an introductory biological statistics course to identify different ways that these students pictured statistics. In the end, they identified five categories to classify the different images of statistics conveyed. These included: a) Statistics as just a course or generic area of study, b) statistics as a collection of tools and procedures, c) statistics as tools with a context, d) statistics as a means of understanding a complex world or as part of a unified picture with understanding data, and e) statistics as integrated into the world itself. The authors do point out that, while the first four categories exist in a loose hierarchy, the fifth was difficult to place in hierarchy.

Despite different data collection approaches, there are striking similarities in the conception categories described by all three of these papers. Novice conceptions recognize nothing particularly unique or clearly unifying about statistics (e.g., statistics as disconnected procedures). Intermediary conceptions see growing cohesion around statistics in its contextual application. Expert conceptions add a clear emphasis on meaning-making, with statistics seen not simply as a tool, but also as a way to see the world differently.

2.2.4 Research Aims

Among the existing studies, there are several consensus notions about the different ways that undergraduate students may perceive of statistics. Still, there is not complete agreement in the reviewed literature on how to best identify and distinguish these conceptions. Additional empirical data may bring needed insight for triangulation. We also see need to bring a stronger theoretical structure to these conceptions by identifying how different conceptions are positioned with respect to one another. This modeling of conceptions can then aid in the development of a psychometric instrument that identifies students' conceptions.

With these needs in mind, we address three objectives in this paper: First, we examine student conceptions of statistics using grounded theory based on both focus group and one-on-one interview data. Second, we offer our identified student conceptions in a model that is connected theoretically to discussions in the literature regarding the nature of statistics and statistical work. Third, we share a validated instrument that assesses how students conceive of statistics in relation to our proposed model. We believe this instrument may be useful in several contexts, including research to document how students' conceptions of statistics may change with different learning experiences, or research on how different conceptions may be associated with different learning behavior.

We organize our paper in the order of our three research objectives. Each section will discuss the data we collected and methods of analysis used, followed by our findings.

2.3 Research Objective 1: Identifying Students' Disciplinary Conceptions

2.3.1 Methods

Our first goal was to capture students' ideas across a range of broad statistical topics in order to develop a typology of conceptions about statistics. We wanted these conceptions to grow inductively from students' statements; thus, the research process is based on a grounded theory approach (Glaser & Strauss, 1967). Our methodology was mainly guided in its implementation by Charmaz (2014) in order to do justice to the constructivist foundation of beliefs research.

A grounded theory approach is characterized by six criteria that are intended to ensure the rigor of the scientific process: "developing theory; generating concepts from data, not existing theory; using the constant comparative method; collecting and analyzing data concurrently; conducting theoretical sampling and saturation; and composing memos" (Lassig, 2022, p. 98). For this research, we developed a typology of conceptions about statistics as a grounded theory that emerged directly from students' statements in both qualitative interviews and focus group discussions. We alternated through phases of data collection and data analysis until further data and analyses were neither able to produce a new type of conception nor substantially changed the understanding of the existing types.

We first conducted two qualitative focus group discussions. The focus groups thus made it possible to observe students' emerging conceptions in depth and in contrast to other ideas shared. All participating students were enrolled in a social science program at a large German public university and had completed at least two statistics courses. The first focus group consisted of ten students who had elected to take a relatively hard advanced statistics course (Focus Group Participant 1 (FGP01) to FGP10). In contrast, the second focus group included nine students who had not elected to take that course but took part in an easier follow-up course (FGP11-FGP19). Participants in both focus groups had been asked in advance to bring some written notes to the discussion on the following question (in German): "What is statistics for you? Try to gather your ideas of what statistics is about. Trying to come up with your own definition can be a helpful element of reflection." Each focus group lasted for approximately one hour.

The focus groups yielded helpful insights, but were often dominated by two or three students doing most of the speaking. For this reason, we also conducted one-on-one, semi-structured interviews to better understand the individual views and conceptions of these students. We asked the following questions (in German) based on findings from the focus groups:

1. What do you generally think about statistics?
2. In your studies, you are required to take statistics. What do you think about that?
3. Do you think statistics is useful? Why? Why not?
4. Do you think statistics is difficult compared to other courses you take? Why? Why not?
5. What do you think, how good will you be at statistics?
6. What would you say is statistics? How would you describe its nature? Or could you define it?
7. If you compare statistics to other courses you take, what is special about statistics?
8. When someone does statistical work, what do they do?
9. How do you think you can recognize a statistical expert? What distinguishes him or her as a person and what competencies does he or she have?

In total, we interviewed 39 students at the beginning of a social science program (Beginning Student 1 (BSt01) to BSt39) and 17 students enrolled in the advanced courses (Advanced Student 1 (ASt01) to ASt17). These interviews were completed in batches, with a review of the transcripts and a basic coding for themes completed at the end of each phase (described in more detail in the next paragraph). Once no substantively new developments on the grounded theory could be made from these interviews, this marked the end of data collection, analysis, and theory development. Thus, data collection included a total of 75 students -- 19 across two focus groups and 56 through individual interviews in five interview blocks until theoretical saturation was reached.

Data analysis was conducted at each step of this research in three phases of coding (Charmaz, 2014; Corbin & Strauss, 2008). First, initial coding of the text was performed. For this purpose, the text was either paraphrased sentence by sentence in a sense-preserving manner or directly adopted as a literal in vivo code. In the second phase, we performed focus coding. The basic principle of permanent comparison of grounded theory was implemented here in a way that always sorted codes that were as similar as possible to each other and separated substantively different codes. Similar to the quantitative procedure of hierarchical cluster analysis, codes were repeatedly compared with codes, codes with categories, and categories with categories, forming ever larger groups of codes into consolidating categories. In the third phase, we completed theoretical coding. The categories formed are compared to determine whether there is an inherent structure in them that constitutes a grounded theory.

In the present research, this revealed that the descriptions of statistics by the students interviewed run along three meaning-bearing dimensions: Codes either a) describe an abstract characteristic about the nature and/or purpose of statistics, b) make statements about the character or course of the statistical process, or c) describe ideas about the (typical) statistical expert. These three dimensions of a statistical conception -- the nature of statistics, the statistical process, and the statistical expert -- were partially laid out in the interview questions, but emerged from the focus groups and increasingly became apparent in the interviews as the pillars of a conception. We will therefore return to these three dimensions in the development of the measurement instrument for research objective 3.

For research objective 1, however, the core result of the grounded theory was identification of four conceptions of statistics that could be traced across these dimensions. In the results section that follows, we outline each of these conceptions, supported by student statements.

2.3.2 Result: Four Conceptions of Statistics

2.3.2.1 The Rules-based Conception

This conception describes statistics as a collection of rules, formulas, and procedures that involve data or probability. Students and practitioners follow these rules, formulas, and procedures, and statisticians are responsible for developing new ones. "Statistics for me is to put something into formulas ... and scientific progress in statistics is to develop a new formula" (FGP05). Statistics is tightly linked with mathematics in this conception. "For me, it's really just math ... because it's simply formulas." (BSt10)

The nature of statistics derived from this is strict, clear, and precise, representing a rather objectivist epistemology on the discipline. For example, BSt08 describes: "There are clear rules or clear formulas for how to calculate certain things." Not only is statistical practice described as clear, but the statistical result is also clear and objective: "you just use several calculation methods to get a clear statement" (BSt05). Statistics is non-debatable, objective, and incorruptible. FGP09, for example, states: "At the beginning, you have these numbers that you have calculated, and they are either right or wrong; and depending on how you interpret them, you then go further" (FGP09).

These links to epistemological beliefs of an absolute Truth are also evident in the ideas about the statistical process, where mathematical operation plays a central role. In statistics you "answer questions based on the data by calculating with these data...You just use several calculation methods to get a clear statement" (BSt05). FGP14 describes: "In statistics, you have to perform many small intermediate steps. You must not make any careless mistakes. Otherwise everything is messed up." Besides this pure execution of predefined calculation steps, the complexity in the statistical process arises from the need to know the right formulas and rules and to combine them appropriately for more complex problems. BSt14 describes: "First you apply formulas to facts, and sometimes you have to combine formulas, I think."

Like the statistical process, the description of an expert in statistics in the rules-based conception relies heavily on the link between statistics and mathematics. "A statistical expert is good with numbers and mathematics" (BSt02) and "he or she can calculate mathematically and has fun with it." (BSt10) are two of many quotes emphasizing that. The statistician's personality has to be detail-oriented, as pointed out by BSt23: "A statistician must above all be careful and detail-oriented so as to not making mistakes." In addition, statistical experts are expected to be perseverant. "In statistics, you have to have ambition. Don't give up right away if you don't get the right result the first time" (FGP11).

The rules-based conception ascribes low importance to the application of statistics, or the context in which it is used. When asked about statistics in this conception, students often associate a course, a learning subject, or a duty within their education. FGP13, for example, admits: "When I think of statistics, the first thing that comes to mind is that I had to take a course called `Statistics´." Overall, it is also noticeable that students rarely speak about the purpose of statistics within the rules-based conception. While the presentation of the other conceptions will also contain an idea about the purpose of statistics, no clear statement can be quoted here for the rules-based conception.

2.3.2.2 The Confirmation Conception

The second cluster of statements describes statistics as a procedure aiming to verify or falsify real-world hypotheses or theories using data. FGP02 explains: "Statistics is able to show whether my hypothesis is reflected in the data." The nature of statistics derived from this is described as systematic and straightforward. For example, FGP04 describes: "Statistics is ... something hard, where you can test and make a statement as: With such and such a probability the hypothesis will be true. ... Statistics helps of course argumentatively quite a lot with theories to make assumptions and to be able to prove them with a high probability. Yes, you can say that in statistics, it is about minimizing errors."

Similarly to the rules-based conception, the confirmation conception thus views statistics as a hard science that is strict and based on given procedures one has to follow. Consequently, students with a confirmation conception often hold epistemological beliefs similar to BSt06: "with statistics, there is probably not a wide range of answer options, but there is right or wrong. In other subjects, especially in the social sciences, there is a lot of leeway. But in statistics, you make hypotheses and then prove them or not." While the rules-based conception and the confirmation conception share this epistemological belief of an absolute Truth, the confirmation conception has a clear idea about the purpose of statistics: "Statistics links the theories of our studies with reality in order to better understand both of them" (FGP02). With this comes a high relevance of application of statistics for the confirmation conception: "For us sociologists, statistics is like a tool ... you can use it to build theories, to support them, or to refute them. And that gives it a very reality-based foundation" (FGP08).

This is also reflected in the statistical process. In the confirmation conception, the statistical work starts within some field of application where one theoretically develops hypotheses or theoretical models. This first part of the process is considered to be open. After formulating concrete hypotheses or theoretical models, this openness has to be dropped to pick and execute the correct statistical test or procedure to test the hypotheses and models. A limited amount of openness returns when one interprets these results within the original context and draws conclusions for the field or for future research. This way, statistics is a powerful tool for uncovering truth and informing real-world decisions using empirical evidence: "Statistics helps incredibly to question theories. ... you can just reflect your own theory much better, or make criticisms very, very quickly unsubstantiated" (FGP03).

A statistical expert is someone who is able to perform all steps of the statistical process: "This person can make statistics, can calculate them, and can hypothesize beforehand, can prove or disprove them, and can probably apply them to reality" (BSt06). Most importantly, however, an expert in statistics can be distinguished by knowing the right statistical test or procedure for a wide variety of situations. Researchers and practitioners in any field can be a statistical expert with the right training. They combine statistical expertise about many tests and procedures with domain knowledge from their field to perform the full statistical process on their own. On the other hand, statisticians can come into the process as consultants to help researchers lacking that kind of statistical knowledge. On top of that, statisticians may sometimes develop their own methods if they find one missing to test a specific hypothesis.

Students with a confirmation conception see purpose and relevance for statistics in a wide range of fields, especially in research and science. For example, BSt20 reports: "I encounter statistics when I watch MaiLab [a science channel] on Youtube. There are always statistics they show to prove what they are saying. That is quite useful."

2.3.2.3 The Descriptive Conception

The third cluster of statements describes statistics as a discipline aiming to communicate complex reality using displays, visualizations, and meaningful measures. FGP05 and FGP03 describe this idea well: “Statistics is about … describing empirical reality” (FGP05); “Statistics makes the complexity of the world communicateable through numbers or graphs” (FGP03). This directly leads to the purpose of statistics according to the descriptive conception-- "Statistics is the attempt to find a system to simplify reality in order to present it in a summarized way" (ASt09). "Statistics, for me, is to put the abstract of the world into concrete numbers" (BSt22). BSt17 extends this idea with more specificity, describing statistics "as the recording of social phenomena with numbers, for example, to see how certain occupational fields are distributed in the population, in general to represent something statistically, to know to what extent this exists." Many of these statements represent epistemological beliefs that assume the existence of an absolute Truth that one is trying to represent in the statistics. As FGP04 states "what I like about statistics is that you deal with facts." Flexibility and room for decision-making are seen in the selection of the data to be presented and, to some degree, in their interpretation.

The statistical process is described as some sequence of "collecting data, processing data, [and] displaying data in tables" (BSt01), or as BSt07 puts it: “You look at a lot of different data, enter them into Excel spreadsheets and then calculate something that you need for a scientific paper or assist some scientist who wants to calculate some characteristics in the population and see how many people are Muslim or something like that. ...Collecting data and presenting it.”

A statistical expert is someone who is experienced with all steps of handling the data, knows many different forms of visualizations, statistical measures, and indices, and is good at choosing the appropriate representation of the data for their audience. For some students, this implies certain quality criteria for the graph, table or measure produced. For example, BSt15 argues one should "display the data clearly and concisely, even for people who don't know much about it." Other students see an obligation to explain what the statistic means on the side of the producer: "Statistics never stands alone. The point is to have reality, to package it in statistics, and to use these statistics to create connections to reality" (FGP03).

Students with a descriptive conception see application for statistics in many everyday life situations. For example, BSt25 and BSt19 discussed news and politics: "When you read the news, you always read about some statistics or survey that showed something" (BSt25). and "I encounter statistics in political election predictions and in facts and numbers, which are always referred to in political discussions" (BSt19). Others mentioned sports: "When I watch soccer on TV, the commentator always reports these statistics, the team has so-and-so much ball possession and played so-and-so many passes, and stuff like that" (FGP18). But statistics may still serve a role in the sciences: "I also think that statistics are very important in the sense of research, because you try to give a depiction...I think people are also more convinced when you say that this percentage of the population or this target group has done this and that" (BSt07).

2.3.2.4 The Investigative Conception

The fourth and last cluster of statements describes statistics as a flexible process of exploring and making sense of data. FGP01 summarizes: "for me, statistics is the handling of data, and in particular, the filtering out of existing relationships...Learning statistics is learning what you can extract from data and how to do it." Regarding the data investigation process: “Statistical thinking abstracts on the one hand...but it also simplifies. Otherwise, when you think about something, you have many different points you can think about. In statistical thinking, you often break it down to a few things...this structured simplification, I think that characterizes statistical thinking” (FGP01). This back-and-forth between simplification and abstraction gives room for curiosity and creativity: “I think statisticians need curiosity and creativity to get to things purposefully” (BSt21). The overarching goal is to gain new insights based on data: “Statistics is used for the in-depth scientific work, to find out something new with data” (BSt16). The idea of an absolute Truth is still possible, but unnecessary in this conception. In particular, the back-and-forth implies the construction of theories and insights rather than objective processes to determine Truth.

The statistical process involves a continual pivot between theoretical ideas and investigation of data--asking new questions and finding new data-based insights. Creativity and innovative strength are important drivers: "In statistics, you need the ability to think outside the box" (FGP06). This thinking outside the box is embedded in a process to which, for example, ASt04 assigns critical reflection on the data source, numerous calculations, their interpretation, and multivariate thinking about the research object: “You can calculate a lot, but you should not forget that the calculations only result in numbers, and you have to explain the relationships and backgrounds yourself. It also always depends on the source of the data or what you want to compare. You still have to interpret the numbers and explain how these relationships occur or what plays into them” (ASt04).

The idea of an absolute Truth is still possible, but unnecessary in this conception. In particular, the back-and-forth between theory and data implies the construction of arguments and insights rather than objective processes to determine Truth. As FGP04 argues: “To be good at statistics, you have to be able to question critically. For example, why exactly this data is used. Or to draw other conclusions from the data that might be better… "Statisticians do not make absolute statements, but rather `It is more likely than not that...´”.

A statistical expert is described not in terms of specific skills, but in terms of a mindset and way of thinking. “I guess he or she is quite flexible and is not very focused on one way, so is not so narrow-minded, because in statistics, there are a thousand different ways to evaluate and interpret things, and with this interpretation, no one says that they are right or wrong, or whether this is really right 100% or not, or whether there is a 100% suitable method at all. And I think that's why a good statistician has to be very flexible and think outside the box sometimes” (BSt03).

As FGP04 highlights: "What makes a good statistician is asking the right questions to the data." ASt03 underlines that the critical thinking of the statistical expert accompanies the whole process from data collection to the last analysis: "[The statistical expert is characterized by the fact] that he chooses the right approaches to collect and analyze data and that he critically questions them himself." This critical attitude is combined in the statistical expert with the ability to recognize problems in the analysis independently and to develop their own solutions. For BSt09, this reads as: "I think experts in such a field can be recognized by the fact that they have ideas on how to solve their problems and that they may not have an answer to all questions. But they have an idea how to deal with the problem."

The applications for statistics within the investigation conception is all-encompassing, but students with this perspective still expressed certain contexts more clearly than others. For example, BSt24 said: "I always associate statistics with research, so I have trouble thinking of it separately. So I always automatically associate it with some other sub-field in which research is then done with the help of statistics." Overall, the investigative conception stands out from the other three in that it does not require an absolute Truth, even if it does not exclude it. It understands the statistical process not as linear, but as dynamic in the constant linkage of theoretical considerations to empirical evidence.

2.3.2.5 Concluding Remarks on the Typology

It is important to recognize that the statements of one person were sometimes sorted into different conceptions. Not all students can always be exclusively assigned to one conception. Instead, among the students studied in this sample, just over half of the students could be clearly assigned to one conception across all statements. However, it often happened that a student shared beliefs belonging to two conceptions. In addition, FGP04 was the only person who made statements in three different conceptions; persons with statements in all four conceptions do not appear in our data.

If we look at which conceptions occasionally appear together in a student, the most frequent combination is a mixture of the rules-based conception and the confirmatory conception. Almost as often, the rules-based conception and descriptive conception are mixed. The investigative conception occasionally appears together with the confirmation conception and just as occasionally appears together with the descriptive conception. The descriptive conception and confirmation conception rarely appeared together, and the rules-based conception and investigative conception never occurred together. Research Objective 2 will elaborate on this and present an overall model of the conceptions.

2.4 Research Objective 2: Modeling Students´ Conceptions

We next consider how these various conceptions could be theoretically linked. We want to represent each conception using common imagery to highlight exactly how each imagines the process of data analysis. We draw from imagery in two key works in constructing our framework.

2.4.1 Framing Data Analysis in Theory

Bailyn (1977) draws on her experiences to describe different ways one might engage with data in research. In her paper, she advocates for conducting data analysis as a cognitive process--a continuous back-and-forth between data and theory that aims to reconcile empirical findings with conceptual understanding. Unfortunately, according to Bailyn (1977), data analysis is often undertaken as a confirmation process--using data as a checker of hypotheses, rather than a source that can be explored for insights and mingled with theory in an ongoing evolution of meaning-making.

To distinguish these two conceptions of research, Bailyn (1977) describes research as the connection of two planes: a conceptual plane and an empirical plane. Research as a confirmation process begins in the conceptual plane, visits the empirical plane to affirm or dispute a theory, and then returns to the conceptual plane to shape the theory accordingly. Research as a cognitive process is continuously moving between the two planes and only ends when the conclusions are considered sufficiently settled by the researcher. We see tremendous value to conceptualizing statistics using this analogy.

Grolemund and Wickham (2014) detail an expert approach to data analysis that adds further nuance to the cognitive process outlined by Bailyn (1977). They argue: "Data analysis is the investigative process used to extract knowledge, information, and insights about reality by examining data…data analysts focus less on the properties of a method and more on the connections between the data, the method, its results, and reality" (p. 185). The authors describe these activities as taking place within a larger sensemaking process, where analytical work "revolves around noticing discrepancies between schemas and reality" (p. 186).

Grolemund and Wickham (2014) also break down this cyclical process into two hemispheres--exploratory and confirmatory analyses (Figure 3). In confirmatory analysis, the researcher starts from a theory and then moves to data to test that theory. On the exploratory side, analysts start with information (e.g., data) and activate related schema (e.g., theories and knowledge structures) as a way to understand and make sense of this information in light of current understanding. The analyst may go through several cycles of schema reorganization as a result of the information found, which may ultimately lead to revisiting the data for more perspectives, or perhaps, finding new data.

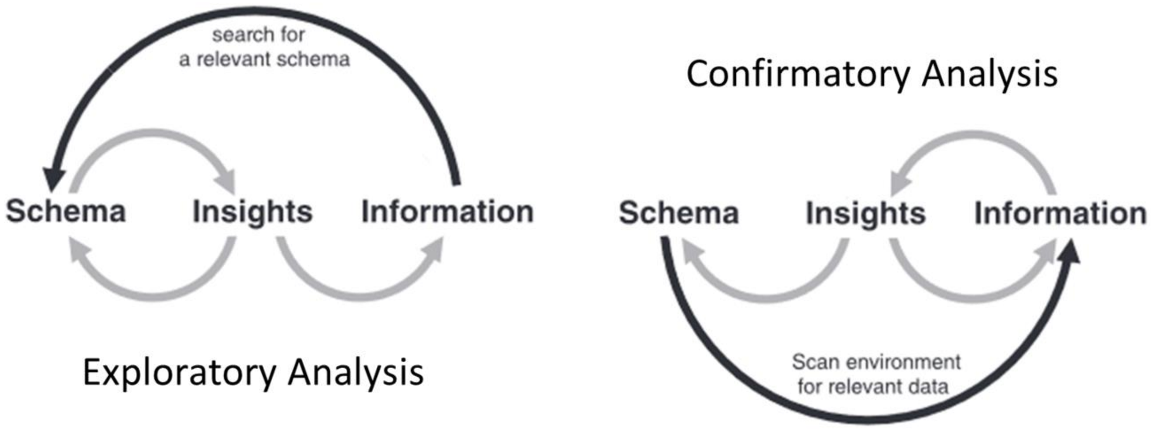

Figure 3: A Cognitive Interpretation of Data Analysis (Grolemund and Wickham, 2014)

2.4.2 Representing the Four Statistical Conceptions

We find the imagery from both Bailyn (1977) and Grolemund and Wickham (2014) to relate closely to the conceptions we identified. We represent our four conceptions using the imagery of a conceptual and empirical plane. We also analogize each of these conceptions by viewing "theory" and "data" as two actors in relation.

Confirmation conception: Theory speaking. This conception closely matches what Bailyn (1977) described as a confirmation process of data analysis. In this conception, the user begins with a theory, then uses data as a way to check or tests that theory. In this way, it is theory speaking to data, and data acknowledging and answering theory's questions. The conversation is one-sided, and data only responds with what is asked, but does not inspire new topics or insights beyond what theory is posing. We represent this approach to analysis as starting from theory and checking in with data at points (Figure 4).

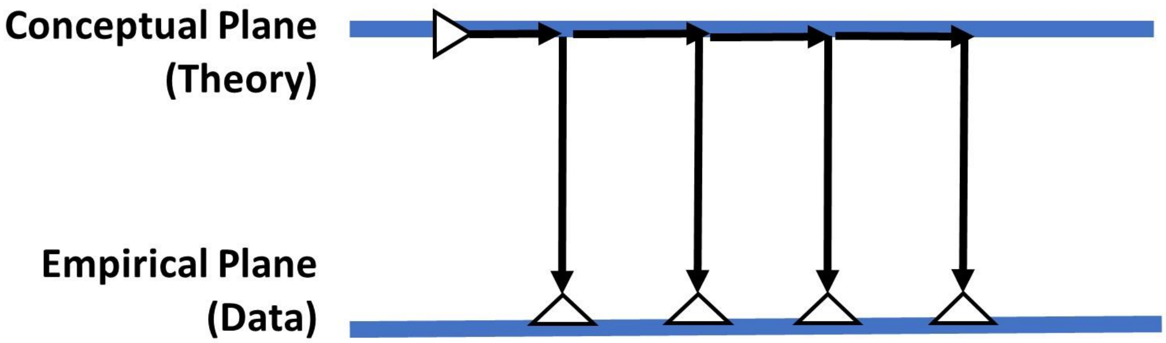

Figure 4: Theory Speaking

Descriptive conception: Data speaking. In other scenarios, analysts want to use data to communicate information. In these and other cases, analysts may let the data speak themselves (Figure 5). In response, theory listens and collects this information, either for immediate insights, or as tidbits for someone else to think about. For an analyst proceeding in this fashion, it may be seen as outside the scope of statistics to be pondering the theoretical meaning of data. The analyst may instead see their role as describing and communicating the data well, and it is for someone else to ascribe meaning to these insights or to argue about consequenses for the context.

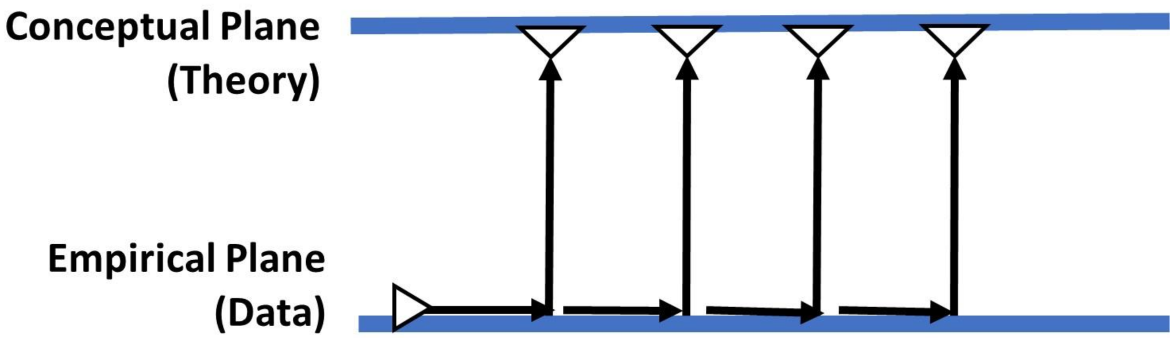

Figure 5: Data Speaking

Investigative conception: Theory and data in conversation. According to Bailyn (1977) a truly cognitive process of data analysis should value both theory and data as having valuable things to say in conversation (Figure 6). The process may start from either the empirical or conceptual plane, and the analyst may choose to linger in either plane at any point in the analysis. By allowing for a dynamic continual conversation between the two planes, where both planes can speak and add to the conversation, the analyst may allow insights to continually develop and build on one another toward richer conclusions and understandings.

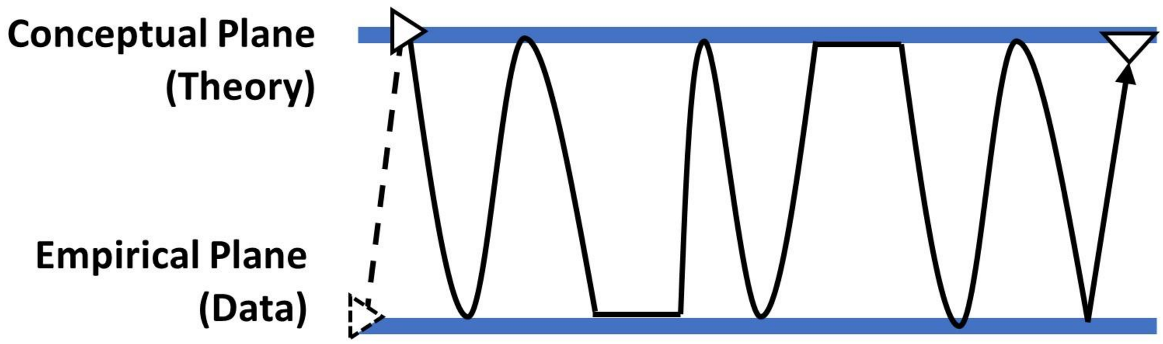

Figure 6: Theory and data in conversation

Rules-based conception: Theory and data disconnected. While the previous three conceptions reflect authentic components of data analysis outlined by Bailyn (1977) and Grolemund and Wickham (2014), the fourth conception we identified from our student data appears more parched. Rather than seeing data and theory talking together, the rules-based conception finds the two disconnected (Figure 7). Instead, statistical work is viewed as set of isolated static procedures and ideas. Some of these elements may be more related to things done with data (e.g., creating a histogram), some may be more related to theory (e.g., completing a hypothesis test). But central to this conception is a lack of cohesion around what these components have in common and how they are combined as part of a larger process of drawing insights from data that connect to real-world understandings.

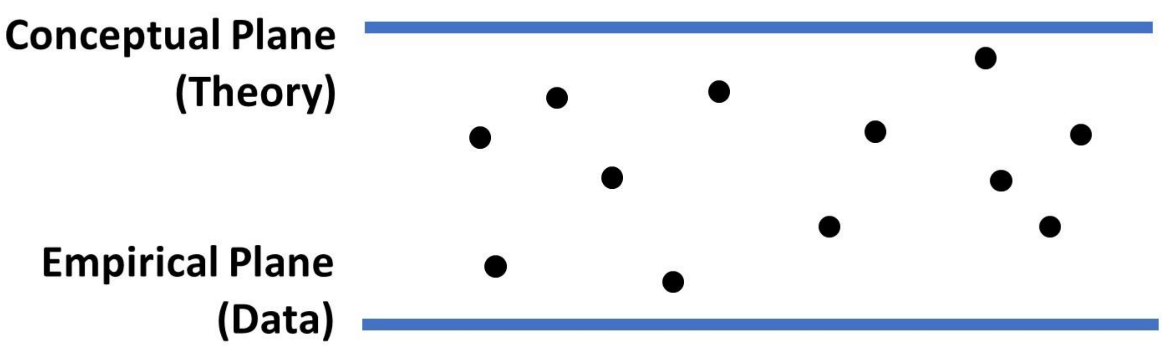

Figure 7: Theory and data disconnected

2.4.3 The Diamond Model of Statistics

Now that these four conceptions have been fleshed out, we share a model that positions these conceptions in intentional relation with one another (Figure 8). To do this, we decompose the dialogue between the conceptual plane and the empirical plane into two dimensions. The first dimension describes how dynamic, connected and interactive the dialog between conceptual plane and empirical plane is. The spectrum here ranges from a static, disconnected, basically non-existent dialogue in the rules-based conception to the very dynamic and connected exchange in the investigative conception. The second dimension describes from which of the two planes the dialogue originates or which plane shapes the dialogue and is thus central to the conduct of the conversation. Here the spectrum ranges from an empirically driven dialog in the descriptive conception to a theoretically driven dialog in the confirmation conception.

We place the rules-based conception in the first dimension at the very bottom, static, disconnected end of the spectrum. Since there is hardly any dialogue in this conception, we can neither speak of a centrality of the data nor of a centrality of the theory. In the second dimension, we therefore place the rules-based conception in the middle. The investigative conception lies at the other, dynamic, connected end of the spectrum in the first dimension. With respect to the second dimension, data and theory are equally influential, which is why we also locate this conception centrally in the second dimension. The confirmation conception and the descriptive conception form the two endpoints of the second dimension, but are on the same level with respect to their dynamics and connectedness. They are more dynamic and connected than the rules-based conception and less dynamic and connected than the investigative conception. We argue that both are closer to the dynamics and connectedness of the investigative conception than to the disconnectedness of the rules-based conception, provided that they are well implemented by the analyst, and therefore place them above the middle of the first dimension. Thus, we portray this model as a diamond, with the investigative conception only slightly elevated above descriptive and confirmation. The rules-based conception sits opposite to investigative and represents a disconnected perspective on statistics.

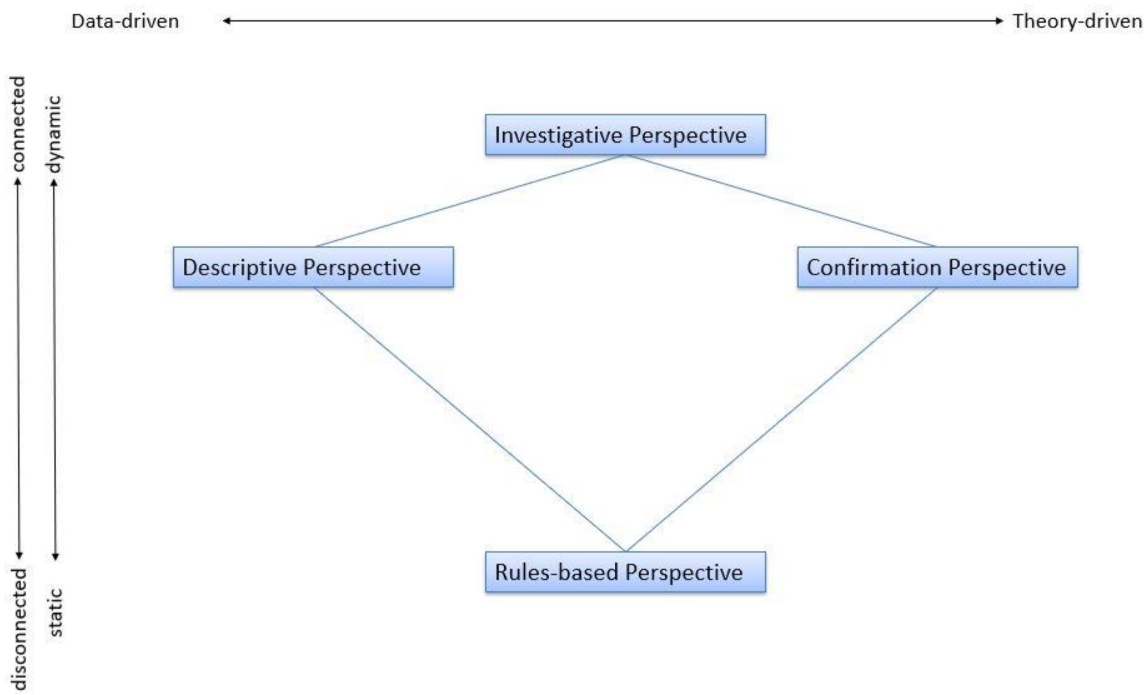

Figure 8: The Diamond Model of Statistics

2.4.4 Triangulation through an Independent Model

Justice et al. (2020) completed a related research study concurrently, yet independently, to our own. Their study surveyed 44 university students across a series of open-ended questions. These questions asked students to write their thoughts about the tasks statisticians complete, the skills and character traits they need, what makes statistical methods useful, what distinguishes statistics as a discipline, and what statistics is centrally about. The research team identified several belief themes and organized these orientations into four general conception quadrants (Figure 9). “Paint by Number” reflected a view of statistics as procedurally focused and oriented toward correct answers. The “Step-by-Step Class” views statistics as a strict process for analyzing data and drawing conclusions. The “Realist” conception views statistics as a way to describe the real world and provide results, but without much need for interpretation or an acknowledgement of uncertainty around results. The “Picasso” conception presents a picture of statistics that is highly focused on meaning and with a clear acknowledgement of abstraction and uncertainty in making claims through data.

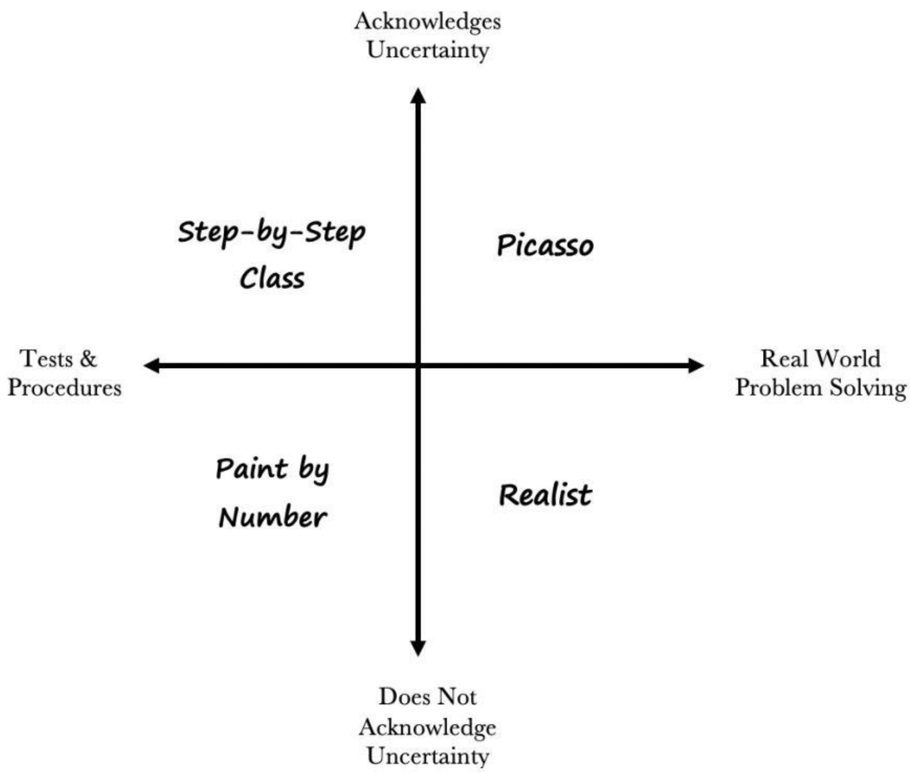

Figure 9: Conceptions of Statistics Framework (Justice et al., 2020)

Even though this work was completed independently, we found several clear connections between the conceptions of Justice et al. (2020) and our own. The “Paint by Number” category maps closely to our “Rules-based” conception, “Step-by-Step Class” to our “Confirmation” conception, “Realist” to our “Descriptive” conception, and “Picasso” to our “Investigative” conception. Even with different data sources, our findings resulted in fundamentally similar perspectives.

The diamond model we proposed differs slightly from the framework of Justice et al. (2020) in how we understand the relations between each conception. We see our investigative conception as connecting two secondary tier conceptions--opening a conversation between theory and data. We also chose a diamond rather than a square to emphasize how the rules-based conception is markedly different than the other three. It lacks a statistically-valid perspective, whereas the descriptive and confirmation conceptions represent valid views on data analysis. We believe our findings both reinforce the conceptions outlined by Justice et al. (2020) and extend these findings by redefining the theoretical relationships between these conceptions. At the same time, we feel strengthened by Justice et al. (2020)'s analysis, as it provides additional qualitative empirical support for the four conceptions we want to measure in Research Objective 3.

2.5 Research Objective 3: Measuring Students' Conceptions

2.5.1 Structure of the developed instrument

As described earlier, the grounded theory analysis yielded four clusters of conceptions. There were also three dimensions of statistics that emerged from focus groups and interviews: The nature of statistics, the statistical process, and what it means to be a statistical expert. Within each dimension, we also identified more fine-grained themes that pervaded students' discussions. In total, this resulted in nine topics on which each of the four conceptions can comment in terms of how statistics is to be described (or described by its absence) in each theme. This resulted in a total of 36 items result, grouped into nine groups of four. In terms of choosing an order, our empirical investigations suggested that topics on statistical process appeared to be the most accessible, followed by the items on statistical experts. Therefore, we start the instrument with these items, leaving the more abstract items on the nature of statistics to follow at the end. The order of topics is as follows:

1. The process of statistical decision-making
2. Steps of statistical work
3. The role of application in the statistical process
4. The focus of good statistical work
5. Skills of a statistical expert
6. Developing statistical expertise
7. Attempts to define statistics
8. The nature of statistical results
9. The purpose of statistics

The four statements of the four different conceptions of statistics are kept together in the measurement instrument in order to make it easier for respondents to work through the questionnaire and to mark the similarity of the statements in terms of content while at the same time making the contrasts visible. The order of the four statements within the groups is changed from group to group. Following this overall structure, we developed a first draft of the measurement instrument, which incorporates terms and formulations from the qualitative data and at the same time takes into account the described conceptions and the theoretical model of the diamond model of statistics as a whole. As a response mode, we collected this data using a seven-point Likert scale.

2.5.2 Cycles of evaluation and improvement of the instrument

The first draft of the instrument was refined in a cyclical process of twelve rounds, evaluating and improving the items with each round. To ensure the usability of the instrument across countries and disciplines, the evaluation cycles were conducted with data from Germany and the United States. Data from Germany come from students enrolled in a social science program at a large public university. The data for the United States come from students entering a statistics for life sciences course at a large public university. The experience levels of the students interviewed were varied to also make the instrument variable with respect to different experience levels. An overview on the twelve cycles is presented in Figure 10.

Cycle	1	2	3	4	5	6	7	8	9	10	11	12
Method	Think aloud interviews				Surveys	Assign items to a category description		Free clustering of items			Surveys	Experts cluster freely
N (Ger.)	1	2	1	1	471	3	0	3	3	6	286	2
N (US)	2	2	2	2	206	0	4	2	2	0	171	2

Figure 10: Cycles of evaluation and improvement of the instrument

Following Lenzner, Neuert, and Otto (2016), we used cognitive pretesting for the first four cycles of work on the measurement instrument. During qualitative interviews, respondents were asked to complete the questionnaire in a think-aloud setting. The interviewers also specifically probed issues of comprehension related to word and phrasing options. After every round of interviews, an analysis of the collected data and a revision of the items of the measurement instrument followed. Thus, after a total of thirteen interviews, the fourth revision of the measurement instrument was available.

For cycle five, we gave our latest revised version of the instrument in a quantitative form for the first time. In Germany, 471 students participated in this survey, and in the United States 206 students took part using a Likert scale. The collected data were analyzed using Cronbach's alpha, Principal Component Analysis, and Confirmatory Factor Analysis both at the level of the overall instrument and at the level of the four conceptions, as well as at the level of the individual items. After evaluation of the quantitative data, the fifth revision of the measurement instrument was carried out.

Cycles six through ten were characterized by more qualitative analyses. For the sixth cycle, we conducted three interviews with German students, where each were presented all 36 items on separate sheets of paper. They were then asked to sort each item to one of four descriptive columns, where each description represented one of the four conceptions. The students were also asked for general feedback at the end. In the seventh cycle, the same procedure was followed, with slightly revised items, to four students from the United States. Cycles eight through ten were similar to the previous cycles, but instead, students were merely asked to sort these items together into clusters without cluster descriptions provided. Following this task, the students were asked to explain their clusters. If this did not result in four clusters, the students were asked in a second round to divide or combine their clusters in such a way that four clusters were created. This way, insights could be gained both about the individual items and their formulations, as well as about the relationship of the items to each other. Thus, these qualitative interviews also helped to further revise the items.

In the eleventh cycle, a quantitative survey was again conducted. In Germany, 286 students participated, in the United States 171 took part. In order to vary the response mode once at this point and possibly gain new insights from this, only in Germany were students asked to respond on a Likert scale in these surveys. In the United States, on the other hand, students were asked to choose between two items of the same group the one that was more suitable for them. From these data, rankings of the items within the groups emerged, providing a different view on the items. The German data were analysed using Cronbach's alpha, Principal Component Analysis, and Confirmatory Factor Analysis. On the basis of all results, an eleventh revision of the items was carried out.

As the twelfth and final cycle of evaluation, we conducted four expert interviews. In both Germany and the United States, one interview was conducted with a professional statistician and one with a statistics education researcher. The procedure of the interviews followed the same scheme as had been implemented with students in cycles eight to ten. The analysis of these interviews was followed by the final revision of the measurement instrument. Since this and also the previous revisions meant only minimal changes to the measurement instrument, the development process was terminated after twelve steps in order to finally evaluate the instrument. We report this final evaluation in the section below.

2.5.3 *Final Assessment of the Quality of the Instrument*

In order to extensively evaluate the developed measurement instrument, both qualitative and quantitative methods were used for the final assessment of the instrument. For the psychometric evaluation of the instrument, another 369 social science students from a large public university in Germany were surveyed with the instrument. At the same time, 186 students from life science majors at a large public university in the United States were surveyed. In both cases, students were provided with seven-point Likert scales for their responses.

For the evaluation of the psychometric properties, Cronbach's alpha was determined as a first step in order to describe the average degree of interrelatedness of the items within a conception. In this sense, alpha > 0.7 is commonly considered a fair threshold for instrument reliability. That said, Taber (2018) aptly describes that such a criterion can only describe the quality of a measurement instrument to a very limited extent. Nevertheless, it is used as a first step in reliability testing. The data from the United States and Germany show that all values are above 0.75 and thus indicate acceptable reliability within the respective concepts.

In the second step of the quantitative evaluation, all four conceptions are examined together in one model in a confirmatory factor analysis (CFA). Following Hooper, Coughlan, and Mullen (2008), the CFA is evaluated in two stages: first, the absolute fit of the model is determined in order to check the fit of the model to the data. A root mean square error of approximation (RMSEA) of less than 0.08 is considered a good fit by MacCallum, Browne, and Sugawara (1996) and Hu and Bentler (1999), while Steiger (2007) promotes a stricter 0.07. Hu and Bentler (1999) also suggest a cutoff value of 0.08 for the standardized root mean square residual (SRMR). In our data, the RMSEA is 0.072 for the German data and 0.073 for the United States. The SRMR is 0.070 for the German data and 0.080 for the United States data. Thus, all four values suggest a good, but not excellent fit.

In the second stage of the evaluation of the CFA, incremental fit indices are evaluated to compare the given model with a null model. This can be used to determine whether the correlations implied in the model are (substantially) higher than other correlations in the data. For the evaluation of the incremental fit, we chose the comparative fit index (CFI). For CFI, Hu and Bentler (1999) favor a cut-off value of 0.95, and analysis of our data falls below that. For the data from Germany, the CFI is 0.752, for the United States a CFI of 0.760 is obtained. At first glance, we can therefore not speak of a good fit in the sense that the given model is not significantly better than other possible models.

However, various authors, including Marsh, Hau, and Wen (2004), argue that fit indices, especially incremental fit indices, need to be evaluated more contextually. According to there argumentation, especially in the case of non-orthogonality of several constructs measured in the model, weaker scores on the incremental fit indices can be accepted if RMSEA attests acceptable model fit. Furthermore, McNeish and Wolf (2023) add that incremental fit indices systematically decrease with model size and should therefore be interpreted more liberally when measuring more than one factor.

Besides the fact that in our test a multi-factorial model is tested, we see two reasons why we might expect a comparatively low CFI in context to this instrument. First, it should be taken into account that the items used, as described above, involve different aspects of students' conceptions about statistics; in particular, items about the nature of statistics should be distinguished from items about the statistical process and from items about the statistical expert. In this context, it seems plausible that items relating to the nature of statistics or to the statistical process or to the statistical expert are more highly correlated within these themes than between these different topics. If we test this assumption in the data, we find that the CFIs then calculated are considerably higher. For the data from Germany, taking only the twelve items on the nature of statistics results in a CFI of 0.843. For the United States, this CFI is 0.891. If one selects only the items on the statistical process, the result is a CFI of 0.856 for Germany and a CFI of 0.843 for the United States. Thus, the CFI of these partial data are already much closer to at least the more liberal threshold for the assessment of CFI. The second reason pertains to the fact that our four conceptions are not disjoint, but instead have built in relations. CFI assumes uncorrelated factors. The data reveals that neighboring conceptions are indeed correlated. Thus, the null model used to calculate the CFI contains theoretically intended correlations that cannot be accounted for in the final calculation. Taking this reasoning into account, we no longer assess the CFIs as a severe cause for concern.

In addition to these quantitative measures, we complement the evaluation of the measurement instrument with a qualitative component. The qualitative evaluation of the measurement instrument was carried out by conducting cognitive interviews, as was the case in the first cycles of its development. A total of 27 students of social sciences in Germany were interviewed under the umbrella of another ongoing interview study. Each of these students was asked to process a randomly selected third of the items while thinking aloud about his or her response. This process was followed by open feedback on the items, after which critical terms and phrases were explicitly probed.

The data generated this way were subsequently analyzed mainly to determine whether the students understood the meaning of the item in terms of the intended conception of statistics in each case. In addition, the terms and formulations queried in the probes were specifically checked for comprehension. The analyses show that only a small proportion of these interviewees were experiencing misunderstandings or problems in comprehension. These problems were also not concentrated with particular items, but were instead distributed over several different items. We conclude from this that the students surveyed understood the items of our measurement instrument in accordance with our idea about the conceptions they represented, further supporting a satisfactory level of validation. While not all quality criteria could be fulfilled to the desirable extent, we believe it valuable to publish the instrument.

2.6 Discussion

We first described four domain-specific conceptions of statistics using a grounded theory approach. We then combined these into a holistic diamond model of statistics based in accordance with previous statistical frameworks and based on additional theoretical considerations from both Bailyn (1977) and Grolemund and Wickham (2014). In the third step, we then developed and evaluated an instrument for the quantitative measurement of the described conceptions, which we publish with this paper. Since the development of the instrument is based on data from Germany and the United States and at the same time comes from students of social sciences and from students of life sciences, a certain range of applicability of the instrument was ensured by the development process. At the same time, however, it cannot be clarified to what extent the developed instrument is equally suitable for all levels of learning or disciplines. Based on our experience with the qualitative interviews, however, we do feel confident that the instrument works well for those with much experience in statistics or data analysis. A limitation of the instrument is that it we are doubtful in its evaluation for those below a college level, or who have very little experience or exposure to statistics to draw on. Studies evaluating the extent of applicability of the instrument in the future are likely to be needed.

At the same time, we see great potential in future studies that use the developed instrument to examine potential implications that the surveyed conceptions have on the behavior of the individuals who hold them. Studies on the sources of personal conceptions are also of interest. We look forward to an exchange on ideas like these, as well as on any other ideas. The developed measurement instrument may be used freely for any study. However, we would be happy to be contacted about this.

Finally, however, we would like to comment on a question that we were often asked in the course of the development process of the instrument, but which we have so far only partially answered in this paper. That question would be which of the four conceptions is the "correct" or at least the "desirable" conception. Since we have not explicitly investigated this question, we only present our thoughts on this topic here, which are partly theoretically underpinned, but partly just thoughts and suggestions for further discussion.

As suggested in Research Objective 2, we do believe that a rules-based conception would be the least desirable of the four. The rules-based conception may not be "wrong" in a strict sense, but by definition, it represents the absence of a unique vision for the purpose and identity of statistics. It views statistical work as static and disconnected, lacks a clear integration to application, and ignores critical thinking in context as an essential part of statistical work.

Looking at the other three conceptions, no such clear judgments about their correctness or desirability can be made. Thus, for all three conceptions, we can think of application scenarios in which such a perspective on statistics seems appropriate for dealing with the task at hand. For example, a statistician who has been tasked in a clinical trial to determine whether a particular new drug is more effective in than the previous state of the art may feel very comfortable with a confirmation conception of statistics and may make a reasonable decision about the drug using a process that is consistent with the confirmatory view. On the other hand, in a statistical office one may (rightly) see his or her task as collecting data, preparing them, and then communicating them to the public or specific decision-makers, in line with the descriptive conception. The investigative conception, much like the imagery used by Bailyn (1977), is one we ourselves have already followed in our scientific work in order to better understand learning and draw conclusions for theory and practice. In this sense, we assume that all three of these conceptions can be useful in their own contexts.

That said, from an educational perspective, we would hope that our students might find room for an investigative conception at some point in their studies. This conception contains elements of both a descriptive and confirmatory conception, but molded together with a philosophy that values a complex pursuit of knowledge. In any case, successful statistics education should at least introduce learners to the spirit of such an investigative conception. Though even a simple awareness of the particular characteristics of the confirmation or the descriptive conception may still be helpful and important for new learners starting their journey into the world of statistics.

2.7 Appendix to Study 1

2.7.1 Instrument (English version)

Group 1: (the process of statistical decision-making)

- In statistics, decision-making is a descriptive process that aims to represent the data accurately. (D)
- In statistics, decision-making is a cyclical process that aims to report insights from the data and propose explanations. (I)
- In statistics, decision-making is a planned process that aims to verify or falsify a claim systematically. (C)
- In statistics, decision-making is a strict process that aims to apply the appropriate formula and arrive at the right statistical answer. (R)

Group 2: (steps of statistical work)

- Analyzing data implies making calculations accurately to generate the result precisely. (R)
- Analyzing data implies creating and choosing graphs and models that communicate the data plainly to others. (D)
- Analyzing data implies many rounds of exploring data, evaluating ideas, and gathering further questions. (I)
- Analyzing data implies first identifying a suitable research question, then using a carefully chosen method to come to an appropriate conclusion. (C)

Group 3: (the role of application in the statistical process)

- When analysing data, the context both inspires questions for the researcher and shapes the analytical decisions made. (I)
- When analysing data, context is not needed and could even endanger the objectivity of the analysis. (R)
- When analysing data, context informs the research question and method, but the analysis can be completed separate from context. (C)
- When analysing data, the researcher creates graphs and data summaries, and readers interpret these results in context. (D)

Group 4: (the focus of good statistical work)

- The most important part of statistical work is to ensure all of the math is calculated correctly and that the statistical rules have been followed. (R)
- The most important part of statistical work is to select an appropriate method to address the research question. (C)
- The most important part of statistical work is to explore the data and ask good questions. (I)
- The most important part of statistical work is to find representations that are informative and easy for readers to understand. (D)

Group 5: (skills of a statistical expert)

- A good statistical analyst is explanation-oriented and draws on their curiosity and experience to identify insights and patterns. (I)
- A good statistical analyst is conclusion-oriented and draws on their familiarity with various statistical methods to confirm or disconfirm a hypothesis. (C)
- A good statistical analyst is answer-oriented and draws on their knowledge of mathematics to carefully executive statistical methods. (R)
- A good statistical analyst is description-oriented and draws on their communicative abilities to provide a clear overview of the data. (D)

Group 6: (developing statistical expertese)

- Expertise in statistics is best developed by learning about many different types of statistical tests and methods. (C)
- Expertise in statistics is best developed by formally learning important statistical formulas and practicing calculations. (R)
- Expertise in statistics is best developed by gaining experience with different kinds of data and learning to think in context. (I)
- Expertise in statistics is best developed by learning to create various representations to effectively communicate data. (D)

Group 7: (Attempts to define statistics)

- Statistics is a systematic approach for evaluating claims with data. (C)
- Statistics is a system of formulas and procedures for answering data-related questions. (R)
- Statistics is a process of choosing visualizations and key figures for depicting data. (D)
- Statistics is a flexible approach for exploring data and making sense of it. (I)

Group 8: (the nature of statistical results)

- Statistics usually produces calculated results that answer a question. (R)
- Statistics usually produces messy results that prompt insights, discussion, and sometimes further questions. (I)
- Statistics usually produces conclusive results that support or refute a theory. (C)

- Statistics usually produces descriptive results that provide an accurate picture of a situation. (D)

Group 9: (the purpose of statistics)

- Statistics is centrally concerned with effectively presenting information taken from data. (D)
- Statistics is centrally concerned with testing a theory or a hypothesis. (C)
- Statistics is centrally concerned with finding insights within data. (I)
- Statistics is centrally concerned with developing and properly using rules and formulas for data analysis. (R)

2.7.2 Instrument (German version)

Gruppe 1:

- In Statistik ist Entscheidungsfindung ein auswählender Prozess, der darauf abzielt, Darstellungen zur Kommunikation der Daten zu finden. (D)
- In Statistik ist Entscheidungsfindung ein zyklischer Prozess, der darauf abzielt, Erkenntnisse aus den Daten zu berichten und Erklärungen zu finden. (I)
- In Statistik ist Entscheidungsfindung ein geplanter Prozess, der darauf abzielt, eine Behauptung systematisch zu belegen oder zu widerlegen. (C)
- In Statistik ist Entscheidungsfindung ein strenger Prozess, der darauf abzielt, die geeignete Formel anzuwenden und zur richtigen statistischen Antwort zu gelangen. (R)

Gruppe 2:

- Statistik treiben bedeutet Berechnungen genau durchzuführen, um das Ergebnis präzise zu ermitteln. (R)
- Statistik treiben bedeutet Grafiken und Modelle zu erstellen und auszuwählen, die die Daten anderen klar und deutlich vermitteln. (D)
- Statistik treiben bedeutet viele Runden der Datenerhebung, der Auswertung von Ideen und der Sammlung weiterer Fragen. (I)

- Statistik treiben bedeutet, zunächst eine geeignete Forschungsfrage zu identifizieren und dann mit einer sorgfältig gewählten Methode zu einer angemessenen Schlussfolgerung zu kommen. (C)

Gruppe 3:

- Bei der Analyse von Daten regt der Kontext sowohl Fragen für den Forscher an als auch prägt er die getroffenen analytischen Entscheidungen. (I)
- Bei der Analyse von Daten ist der Kontext nicht erforderlich und könnte sogar die Objektivität der Analyse gefährden. (R)
- Bei der Analyse von Daten ist der Kontext ausschlaggebend für die Forschungsfrage und -methode, aber die Analyse kann auch unabhängig vom Kontext durchgeführt werden. (C)
- Bei der Analyse von Daten erstellt der Forscher Diagramme und Schlüsselindikatoren, und der Kontext gibt diesen Ergebnissen Bedeutung. (D)

Gruppe 4:

- Der wichtigste Teil der statistischen Arbeit ist die korrekte Berechnung der gesamten Mathematik und die Sicherstellung, dass die statistischen Regeln eingehalten werden. (R)
- Der wichtigste Teil der statistischen Arbeit besteht in der Wahl einer geeigneten Methode, um eine Entscheidung über die Forschungsfrage zu treffen. (C)
- Der wichtigste Teil der statistischen Arbeit besteht darin, die Daten zu erforschen und gute Fragen zu stellen. (I)
- Der wichtigste Teil der statistischen Arbeit besteht darin, Darstellungen zu finden, die informativ und für Leser leicht verständlich sind. (D)

Gruppe 5:

- Ein guter statistischer Analytiker ist erklärungsorientiert und stützt sich stark auf seine Neugierde und Erfahrung, um Einsichten und Muster zu erkennen. (I)
- Ein guter statistischer Analytiker ist entscheidungsorientiert und stützt sich stark auf seine Vertrautheit mit verschiedenen statistischen Methoden, um eine Hypothese zu bestätigen oder zu widerlegen. (C)

- Ein guter statistischer Analytiker ist antwortorientiert und stützt sich stark auf seine Mathematikkenntnisse, um statistische Methoden sorgfältig auszuführen. (R)
- Ein guter statistischer Analytiker ist beschreibungsorientiert und stützt sich auf seine kommunikativen Fähigkeiten, um einen klaren Überblick über die Daten zu geben. (D)

Gruppe 6:

- Expertise in Statistik lässt sich am besten durch das formale Erlernen vieler verschiedener Arten von statistischen Tests und Methoden entwickeln. (C)
- Expertise in Statistik lässt sich am besten durch das formale Erlernen wichtiger statistischer Formeln und das Üben von Berechnungen entwickeln. (R)
- Expertise in Statistik lässt sich am besten entwickeln, indem man Erfahrungen mit verschiedenen Arten von Daten sammelt und lernt, in Zusammenhängen zu denken. (I)
- Expertise in Statistik lässt sich am besten entwickeln, indem man lernt, verschiedene Darstellungen zu nutzen, um Daten effektiv zu kommunizieren. (D)

Gruppe 7:

- Statistik ist ein systematischer Ansatz zur Beurteilung von Behauptungen durch Daten. (C)
- Statistik ist eine Sammlung von Verfahrenzur Lösung datenbezogener Probleme. (R)
- Statistik ist eine Sammlung von Visualisierungen und Kennzahlen zur Beschreibung von Daten. (D)
- Statistik ist ein flexibler Ansatz, Daten zuerforschen und zu verstehen. (I)

Gruppe 8:

- Statistik liefert in der Regel berechnete Ergebnisse, die eine datenbezogene Aufgabe beantworten. (R)
- Statistik liefert in der Regel unübersichtliche Ergebnisse, die zu Einsichten, Diskussionen und manchmal zu weiteren Fragen führen. (I)
- Statistik liefert in der Regel schlüssige Ergebnisse, die eine Theorie unterstützen oder widerlegen. (C)

- Statistik liefert in der Regel beschreibende Ergebnisse, die ein klares Bild der Realität vermitteln. (D)

Gruppe 9:

- Das zentrale Anliegen der Statistik ist es, Informationen aus Daten übersichtlich darzustellen. (D)
- Das zentrale Anliegen der Statistik ist es, eine Theorie oder eine Hypothese zu testen. (C)
- Das zentrale Anliegen der Statistik ist es, in Daten Erkenntnisse zu finden. (I)
- Das zentrale Anliegen der Statistik ist es, Regeln und Formeln für die Datenanalyse zu entwickeln und korrekt anzuwenden. (R)

2.8 Concluding Remarks and Transition to Study 2

In its qualitative first section, Study 1 was able to contribute to research question 1a of this thesis by describing and conceptualizing four types of beliefs about statistics. The four types certainly add some nuances that were less present in previous publications, but are not entirely different from previous publications, especially from Justice et al. (2020). This part of the study can therefore be seen as a replication study of existing work, which has its own value, especially in the area of qualitative typology development.

The second section, in which the types are compared with existing literature on the nature of statistics, adds considerable value. In this section, the described types are embedded in a theoretical model that enables a more comprehensive conceptualization of beliefs about statistics. Section 2 of Study 1 thus provides a contribution that enriches research question 1a and thus research on the conceptualization of beliefs about statistics.

The third section of Study 1 proposes and tests a measurement instrument that fits the conceptualization. Study 1 thus also provides an answer to research question 1b. This enables further research in this thesis and beyond to investigate beliefs about statistics in quantitative research designs, which make many research questions, especially the main research question of this thesis, possible in the first place.

However, the measurement of beliefs about statistics using Study 1 is not sufficient to address the main research question. The main research question aims to relate non-cognitive factors, including the aforementioned beliefs about statistics, to learning success by viewing learning behavior as a mediator between the two. As explained in chapter 1, this learning behavior should not be described using self-reports, but as objectively as possible, for example using digital behavioral data.

However, the possibilities to use digital behavioral data for the analysis of learning were still very limited at the institution where these studies were conducted by the time work on this thesis began. The main platforms used to organize learning were and still are Stud.IP as a learning management system and ILIAS as a learning platform. Both have their beginnings in times when learning analytics was hardly a concept and when the use of digital behavioral traces for learning research was not yet part of the usage scenarios the developers had in mind. Even though both systems have since been opened up somewhat for the integration of learning analytics, the scope and granularity of the data are not yet at the highest achievable level. A learning platform that is better suited to these goals and in whose development learning analytics has been incorporated from the start is therefore a great opportunity to address the main research question of this thesis in detail.

It was possible to actually develop such a platform because the desire for a learning analytics-friendly learning platform could be combined with a pedagogical didactical idea, which, coming from a learner-centered approach, provided support for students through a chatbot-based digital tutor. The Stifterverband gratefully supported the development of this digital tutor and honored it with an award for innovative higher education teaching (https://www.stifterverband.org/lehrfellowships/2018/hobert_berens, grand number H120 5228 5008 32762).

Both the pedagogical and didactic background behind this tutor and the development and testing of this tutor are presented in Study 2.

3 Study 2: AI-based Digital Tutors as Intermediaries between Students, Teaching Assistants, and Lecturers in Large-Scale Formal Educational Settings – A Design Science Research Study

This chapter originated as a manuscript co-authored with Sebastian Hobert, University of Goettingen. The two authors share first authorship of the manuscript. A shortened form of the manuscript is published at *Educational Technology Research & Development*. It is reproduced in this dissertation with the kind permission of the publisher, Springer Nature, and the co-author. The original publication can be found under

Hobert, S., & Berens, F. (2024). Developing a digital tutor as an intermediary between students, teaching assistants, and lecturers. *Educational Technology Research & Development, 71*(5), 1-22. https://doi.org/10.1007/s11423-023-10293-2

Abstract: Individualized learning support is an essential part of formal educational learning processes. However, in typical large-scale educational settings, resource constraints result in limited interaction among students, teaching assistants, and lecturers. Due to this, learning success in those settings may suffer. Inspired by current technological advances, we transfer the concept of chatbots to formal educational settings to support not only a single task but a full lecture period. Grounded on an expert workshop and prior research, we design a natural language-based digital tutor acting as an intermediary among students, teaching assistants, and lecturers. The aim of the digital tutor is to support learners automated during the lecture period in natural language-based chat conversations. We implement a digital tutor in an iterative design process and evaluate it extensively in a large-scale field setting. The results demonstrate the applicability and beneficial support of introducing digital tutors as intermediaries in formal education. Our study proposes the concept of using digital tutors as intermediaries and documents the development and underlying principles.

Keywords: digital tutor, conversational agent, chatbot, technology-enhanced learning, e-learning, large-scale educational setting

3.1 Introduction

A common phenomenon that can be observed across many institutions in education is an increasing student-to-lecturer ratio, which is considered to be “a challenge for many schools and universities worldwide" (Maedche et al., 2019). This is especially true and problematic for introductory courses in which lecturers teach basic skills that are essential for the further learning success of students. The increasing student-to-lecturer ratio usually results in less interaction among learners and lecturers (Lehmann, Söllner, & Leimeister, 2016). At the same time, smaller ratios are expected to contribute "to a better learning environment for the students, and to improved working conditions for teachers and staff" (OECD, 2019b). However, interaction and the corresponding individualization and personal feedback are regarded as key factors for learning success (Hattie, 2015). From a didactical perspective, a learning setting would be desirable in which lecturers can (1) address the students' individual needs and (2) support them in one-to-one training situations when additional learning support is needed. In most educational institutions, such a low student-to-lecturer ratio is unimaginable. The main reasons for this are resource constraints and increasing numbers of students. For instance, according to the (OECD, 2018), the number of students is increasing in many countries. Simultaneously, we can observe a decrease in public spending on education in at least some countries (Maedche et al., 2019).

To overcome these challenges in education, e-learning approaches have been proposed for many years to provide students with additional resources even in times when no lecturer is available. This should enable students to learn even in times of informal learning settings (e.g., as preparation before or as a follow up after a lecture or seminar). The number of available e-learning tools and concepts disseminated substantial during the last two decades. When the first e-learning systems were widely approached in educational institutions, those information systems (IS) were often designed as tools that should support organizational aspects of lectures and seminars. These early learning management systems (LMS) mainly focused on delivering content and information about the course (Wolfe & Cedillos, 2015). Typical functionalities of LMS are the management of participant lists, communicating the lecture dates as well as announcements, and distributing files (e.g., slides and scripts). In the subsequent development, organizational aspects become less important and focusing on actual learning processes evolved as important targets of e-learning systems. Nowadays, providing learning contents as learning modules (e.g., Anke & Schumann, 2018), offering formative learning assessments (e.g., quizzes), or even providing recordings of lectures are state of the art. In many didactical learning scenarios, these e-learning resources are offered as an addition to in-class teaching. If a majority of the learning is moved from in-class lectures to online education, scenarios like blended learning evolve. Even though the rise of including e-learning tools in formal learning settings as additional resources, the level of interactivity and personalization that can be achieved by individual human tutoring is still not yet reached using e-learning tools. This results in the fact that in many formal learning settings, human lecturers and tutors still struggle with providing proper individualization and feedback to learners.

To address this issue of insufficient individualizing and feedback that could be solved with one-to-one lecturing when there would not be any resource constraints, conversational agents (CAs) have been proposed in recent research as a solution for improving education (see recent literature reviews on CA in education like R. Winkler and Söllner (2018) and Hobert and Meyer von Wolff (2019)). CAs are computer programs that imitate human conversations by communicating with humans using natural language (Rubin, Chen, & Thorimbert, 2010). In educational settings, conversational agents typically interact with students using text, speech, gestures, or haptics and aim to support learners individually (Hobert & Meyer von Wolff, 2019). Thus, in contrast to typical online learning tools, CAs try to communicate with learners in a human-like way. The research on CAs is grounded in the works of Weizenbaum (1966), who developed the first known chatbot. In educational settings, the research dates back at least to the 1990s (e.g., Graesser, Person, & Magliano, 1995) and was followed by many successful showcases of CA use in educational settings (e.g., Nye, Graesser, & Hu, 2014). Experimental research studies proved their effectiveness in different settings. CAs have, for instance, been used to support learners by reflecting their own learning experiences (e.g., Song, Oh, & Rice, 2017), explaining mathematical concepts (e.g., Grossman et al., 2019) and provideing feedback. Despite these successful research studies, CAs have not been adopted widely in education. One reason for this might be that CAs have most often only be used in very narrow cases. Thus, previously available CAs are, in many cases, restricted and unable to answer student questions in a similar way as human lecturers or teaching assistants could answer them. Holistic learning support offered by CAs in order to imitate the gold standard of education — one-to-one tutoring — was not yet reached.

Due to recent technological improvements and the availability of smart personal assistants for customers (e.g., Amazon's Alexa or Google's Assistant), the capabilities of CAs evolved massively. The availability of open-source frameworks or commercial cloud services enabled high-quality human-like conversations based on natural language processes (NLP) that were not available before. Thus, using CAs for educational purposes received increasingly more attention by researchers and practitioners recently (Hobert & Meyer von Wolff, 2019). By building on the prior research results of using CAs in education and the paradigm that computers can be recognized as social actors (CASA-paradigm) (Nass & Moon, 2000; Nass, Moon, Fogg, Reeves, & Dryer, 1995), the currently available technologies allow developers to strive towards reaching the goal of providing each student an artificial tutor to enable individualized learning and feedback.

There exists some design knowledge on developing CAs in information systems (IS) literature. However, transferring it from common use cases like customer support or sales to educational settings in order to develop an artificial tutor is not easy as the use in education differs significantly. Usually, CAs are only used for a short time (e.g., to get support). In contrast to that, learning systems that are integrated into formal educational settings need to support learners in long-term use (e.g., a complete lecture period). They should interact like artificial tutors, and they need to provide guidance to students during the whole learning process. Thus, a more in-depth, long-term support is needed (i.e., answering knowledge questions or questioning the learners as part of formative assessment). Based on this need for design knowledge, we focus on designing an artificial tutor for large-scale educational settings by conducting a multi-cycle design science research study. Thus, we answer the following research question:

> **RQ:** How does design chatbot-based learning systems that are suited to supporting learners through long-term learning processes in formal educational settings?

To address our research question, we have followed the three-cycle design science research method by (Hevner, 2007) and Hevner, March, Park, and Ram (2004) to specify, conceptualize, design, demonstrate, and evaluate a software artifact solution for the given problem situation. We ground our research in the computer are social actors (CASA) paradigm (Nass et al., 1995; Nass & Moon, 2000) and the interactive, constructive, active, and passive (ICAP) framework (Chi & Wylie, 2014), which act as our kernel theory. Based on this approach, we finally propose a nascent design theory for digital tutors based on CA technology for long-term learning processes that accompany lecturers in formal educational settings.

This paper contributes in multiple ways to theory and practice: First, we conceptualize the research problem from a theoretical view of learning. Second, we provide a design science level 1 contribution by demonstrating a system architecture and a corresponding software artifact as a “situated implementation of [an] artifact” (Gregor & Hevner, 2013). Third, we contribute with a nascent design theory based on the comprehensively evaluated artifact in multiple settings that is transferable to further problem situations. This represents a design science level 2 contribution (Gregor & Hevner, 2013).

The remainder of the paper is structured as follows: In Section 2, we outline related work on chatbot-based learning systems and describe the theoretical background of this paper by introducing the CASA paradigm and ICAP framework. In Section 3, we describe our three-cycle design science research method. In Section 4, we present our design process, including the developed software artifact and the results of multiple evaluations. In Section 5, we discuss the results, the contributions for research, and the limitations of this paper before we conclude the findings in Section 6.

3.2 Theoretical Background

In the following subsections, we present the theoretical background, including related research of chatbot-based learning systems that base on CAs, as well as the ICAP framework and the CASA paradigm. In doing so, we show how our research builds on this existing knowledge base to achieve our research objectives.

3.2.1 Conversational Agents as Digital Tutors

The term conversational agent refers to software-based information systems designed to understand and interact with human-beings using natural language. Nowadays, both voice-based and text-based systems are available. Voice-based systems are mainly used in commercial assistance systems like Apple’s Siri, Amazon’s Alexa, Google’s Assistant, and others. They are often either integrated into smartphone operating systems or embedded into smart speakers. Text-based conversational agents are usually known as chatbots or alternatively as chatterbots, talkbots or digital assistants. In educational settings, researchers also refer to the term pedagogical conversational agents.

From a technical perspective, chatbots apply natural language processing based on artificial intelligence and machine learning. Most chatbots currently act reactively (i.e., respond to users' questions with suited answers). However, proactive communication started by a chatbot is also possible. One main feature distinguishing chatbots from other information systems is the processing of user input and the generation of the output. Whereas classical user interfaces provide graphical interfaces controlled with a mouse, touchscreen, or keyboard, chatbots must process natural language input and output. Thus, natural language processing components represent a major part of the typical technical architecture. The task of the natural language understanding component is to preprocess the users' input and to redirect it to a dialog manager where answers are searched in connected knowledge bases. The knowledge bases usually consist of a pre-structured database of information. However, unstructured web searches or the computation and processing of data can also be used for generating suitable answers. To respond to the users in natural language, so-called natural language generation components generate messages that are sent to the user as output.

Due to this human-like conversation between chatbots and users, researchers assume that users see chatbots as social actors (Liebrecht & van Hooijdonk, 2020). Users often communicate with chatbots in a similar way they would interact with other humans. This phenomenon is known as the computers are social actors (CASA) paradigm, which was described, for instance, by Nass and Moon (2000) and Nass et al. (1995). Many studies were conducted in which different aspects of this phenomenon were identified. For instance, it was identified that users associate chatbots with a gender (Feine, Gnewuch, Morana, & Maedche, 2020), judge the chatbot's tone and politeness, and predict the chatbot's personality. Chatbots that communicate in a socially expected and natural conversation are more likely to be accepted. Feine et al. (2020) developed a taxonomy that structured social cues that can be used when designing a conversational agent. This research on conversational agents, in general, can be transferred to educational settings as well. Since chatbot-based learning systems, which are focused on in this paper, aim to provide artificial teaching assistants who carry out tasks that are similar to human teaching assistants, it becomes important that users acknowledge the systems as eligible actors. The chatbots in educational settings should encourage learners to converse with them about the learning content in order to improve learning success. Based on the CASA paradigm and the taxonomy of social clues, this can be achieved with chatbots.

Chatbots have been developed and analyzed in research studies since the 1960s when Weizenbaum (1966) invented the first widely known chatbot, ELIZA. ELIZA focused on natural-language communication and is acknowledged as the starting point of chatbot research. Many prior, well-known chatbots, such as A.L.I.C.E (Wallace, 2009), applied rule-based natural-language-processing tools (e.g., using the artificial intelligence markup language AIML). Nowadays, state-of-the-art chatbots reply mostly using more sophisticated, natural-language-processing approaches that are based on machine learning that is embedded into publicly available frameworks or cloud-based services. Due to technological improvements in the last few years, chatbots have started to spread across many domains in research and practice.

Recent literature reviews targeting chatbots in educational settings indicate opportunities for improving learning processes (e.g., Hobert & Meyer von Wolff, 2019; R. Winkler & Söllner, 2018). Characteristic prior research papers that use a chatbot in educational settings include that of Mikic, Burguillo, Llamas, Rodriguez, and Rodriguez (2009), who developed Charlie, a chatbot that is based on AIML. Charlie provides learners an alternative user interface for an existing e-learning system. Another example is the MentorChat software, extensively researched, by Tegos, Demetriadis, and Karakostas (2011), which is a conversational agent that supports learners in collaborative learning tasks in chat discussions. Further, Graesser, Cai, Morgan, and Wang (2017) developed the AutoTutor prototype that focuses on complex problems and reasoning. Besides these selected chatbot-based learning systems, further design- and evaluation-oriented software prototypes have been developed and researched (see, e.g., Hobert and Meyer von Wolff (2019) for further examples).

However, integrating chatbots in long-term formal learning settings has only been researched in a few cases. Most existing chatbots focus on supporting very specific use cases. Thus, students use them only for a specific purpose in a very short time frame. In contrast to that, the learning setting that we investigate in this study address a long-term learning scenario. The chatbot that we call digital tutor in this project should support students during the whole learning process of a lecture period. Thus, focusing on a single use case is not sufficient, instead, providing guidance through the learning process is required (see Section "Specifying the Problem Statement" below for an in-depth description of the scenario). Thus, the scope of a digital tutor needs to be much broader compared to typical chatbot use cases.

3.2.2 *ICAP Framework as Theoretical Grounding*

In the subsection above, we outlined that computers and particularly chatbots can act human-like by applying socially accepted and expected behaviors. Whereas, on the one hand, this is expected to improve users' adoption of chatbots, it needs to be questioned whether this is also beneficial in learning settings for improving learning success. Learning success is the most important factor of (chatbot-based) e-learning systems and thus needs to be focused on. The human-computer interaction with a chatbot-based system can be characterized by increased user engagement with the system. Due to the conversational discourse using natural language, the users do not interact in a passive way but need to interact with the system in an intellectual way. Based on the perception of chatbots as social actors, this can even be strengthened.

Prior research, particularly the ICAP framework by Chi and Wylie (2014), shows that engagement is a key factor for increasing the learning outcome. According to them, learners can engage with learning materials in different ways. Starting "from passive to active to constructive to interactive", the learners' engagement will increase in learning settings. Alongside an increase in engagement, an increase in the learning outcome is expected.

To bridge the gap between the ICAP framework and the chatbot technology, an in-depth look into the different engagement types and corresponding learning activities is needed: Passive engagement encompasses activities in which learners only consume learning content. For instance, in lectures, students often only passively engage with the learning materials when they just listen to the lecturer without taking notes or making other efforts to interact with or rethink the contents. If students start to manipulate the presentation of the content, the engagement type increases. For instance, if students highlight important text passages in printed lecture notes, this is classified as active engagement, as the learners' interaction with the learning content increases. Thus, outcome of the learning activity is expected to improve. In contrast to the first two engagement types, constructive engagement requires that learners generate new content on their own when interacting with learning content. Chi and Wylie (2014) state that the generated contents of the learners "should contain new ideas that go beyond the information given" (Chi & Wylie, 2014). To reach constructive engagement, learners need to be able to successfully generate at least some output that was not given directly. To reach the final stage of the engagement taxonomy called interactive engagement, interaction between a learner and a third party in some form of dialogue is needed. The third-party might be, for instance, another human, a learning environment, or an information system. However, Chi and Wylie (2014) make additional demands towards dialogues to reach interactive engagement. Specifically, two criteria are defined: First, both participants of the interaction need to engage in a constructive way. If one or both participants do not reach this criterion, the whole conversation cannot be labeled as interactive. Second, Chi and Wylie (2014) specify that "a sufficient degree of turn-taking must occur". This specification of interactive engagement can be reached in learning settings that allow dialoguing. As this is the dominant interaction form of chatbot-based systems, it can be adapted to this technology, if and only if the two criteria are reached. Thus, chatbot should not only be a restricted participant of the discourse but should also able to engage in it in a constructive way as required by Chi and Wylie's (2014) first criterion.

The applicability of interactive engagement to chatbot design has been demonstrated in several recent scientific projects. For instance, Ruan et al. (2019) developed the QuizBot prototype that focused on teaching factual knowledge. The authors showed that the conversational interface outperformed a classical graphical user interface. Hobert (2019) conceptualized and designed a chatbot to support students learning to program. His Coding Tutor prototype was evaluated by novice programming students as being a substantial improvement as it was able to provide individualized feedback independently of a specific time or place. Thus, it could be available when no human teaching assistant is available. R. Winkler and Söllner (2020) designed and evaluated a method to support educators in creating digital assistants using existing tools and services. They based their research on the ICAP framework as their kernel theory for the design process and "show[ed] that educators can create IT-based systems that are able to bring students into an interactive learning behavior" (R. Winkler & Söllner, 2020). In this paper, we also rely on the ICAP framework as it provides a theoretical foundation for bridging learning processes to the technology of chatbots and has been shown as suitable in prior research.

3.2.3 Research Objective

Considering the theoretical background and the related work on chatbot-based learning systems, we observe that some research on using CAs for fostering students to learn already exists. However, to the best of our knowledge, there are only limited insights on designing chatbot-based learning systems that target long-term learning processes in formal educational settings. As the existing research already provides some insights into the design of CAs and CA communication, we build our research on this existing design knowledge. Nevertheless, we argue that the long-term use of CAs in formal educational settings differs substantially from ordinary use cases of CAs. Thus, research is needed to derive theoretical design knowledge that is based on rigorous design process studies. Thus, we target the following research objectives in this paper:

1. Instantiating a theoretically grounded artifact of a CA that targets long-term learning processes in a large-scale formal educational setting, which is systematically evaluated in a real-world setting.
2. Proposing a nascent design theory derived from generalized design principles that can be applied to the class of CAs that target long-term learning processes in formal educational settings.

3.3 Research Design

We followed the three-cycle design science research framework by Hevner (2007) and Hevner et al. (2004) in this project to address the question of how to design a digital tutor based on a conversational interface that aims at supporting learners in a formal learning setting. As part of the design and evaluation process, we implemented the developed artifact in the lectures of a large university as a case-based demonstration and tested it in this real-world setting during a four-month period. The university is a full-scale university with 13 faculties and a total of approx. 30,000 students. It has a focus on in-class lectures and seminars and uses e-learning mainly as an addition to in-class courses. In total, approx. 700 students engaged with the developed artifact during the rigor cycle. By grounding our project on an established theory base and implementing the artifact in the described real-world setting, we balanced “the synergy between relevance and rigor” (Hevner, 2007).

We applied nine consecutive research steps in the relevance, rigor, and design cycles as displayed in Figure 11. In the first step, we specified the research problem by deriving it from the educational practice of university lectures with a large-scale audience. To get a deeper and broader insight into the educational practice in university courses with large-scale audiences, we conducted an expert workshop with eight participants (lecturers, educational researchers, e-learning experts, and instructional designers) in which we discussed how CAs could support learners in large-scale educational settings. By integrating the expert workshop into our design process, we managed to include the practitioners’ view in our research as part of the relevance cycle. From a theoretical perspective, we also grounded this research step on the ICAP framework and the CASA paradigm, as described in the previous section. Thus, we also incorporated the rigor cycle.

Using the rigor and relevance cycle as a basis, we derived requirements and corresponding design principles. The derived requirements and design principles are the foundation for the first design cycle. We used those as a basis to deduce the technical concept with a focus on both the graphical user interface for the students’ view and the natural language processing. Subsequently, we implemented a first software artifact as a digital tutor who acts as an intermediary between students, teaching assistants, and lecturers. To allow a deep integration into the lectures and to provide the students with a fully-functional CA that is capable of answering questions to all contents of the lecture, we cooperated with several teaching assistants who created an accessible knowledge base covering learning-related and organizational lecture topics. In total, the teaching assistants created over 500 different entries in our knowledge base for each lecture. Additionally, formative assessments, downloadable files, and lecture recordings were included.

Starting in April 2019, we introduced the chatbot-based learning system to the students of a course in the first week of the lecture term. During the following four months of the field study, approximately 700 learners interacted with the system. After a first qualitative evaluation at the end of the onboard phase in a field setting, we revised the artifact and resulted in a fully-functional chatbot-based learning system. To evaluate the developed software artifact and the underlying design principles after the end of the lecture period, we accompanied the field study by a mixed-method evaluation study (quantitative survey, qualitative interviews, and analyses of log data and chat messages). Finally, we documented the design knowledge based on core components of design theories as proposed by Gregor and Jones (2007) after the final evaluation.

In this paper, we present the full design process, including rigor and relevance cycles, and discuss the evaluation results of both quantitative and qualitative studies in detail. In the subsequent subsections, we also provide more details about the corresponding evaluation approaches.

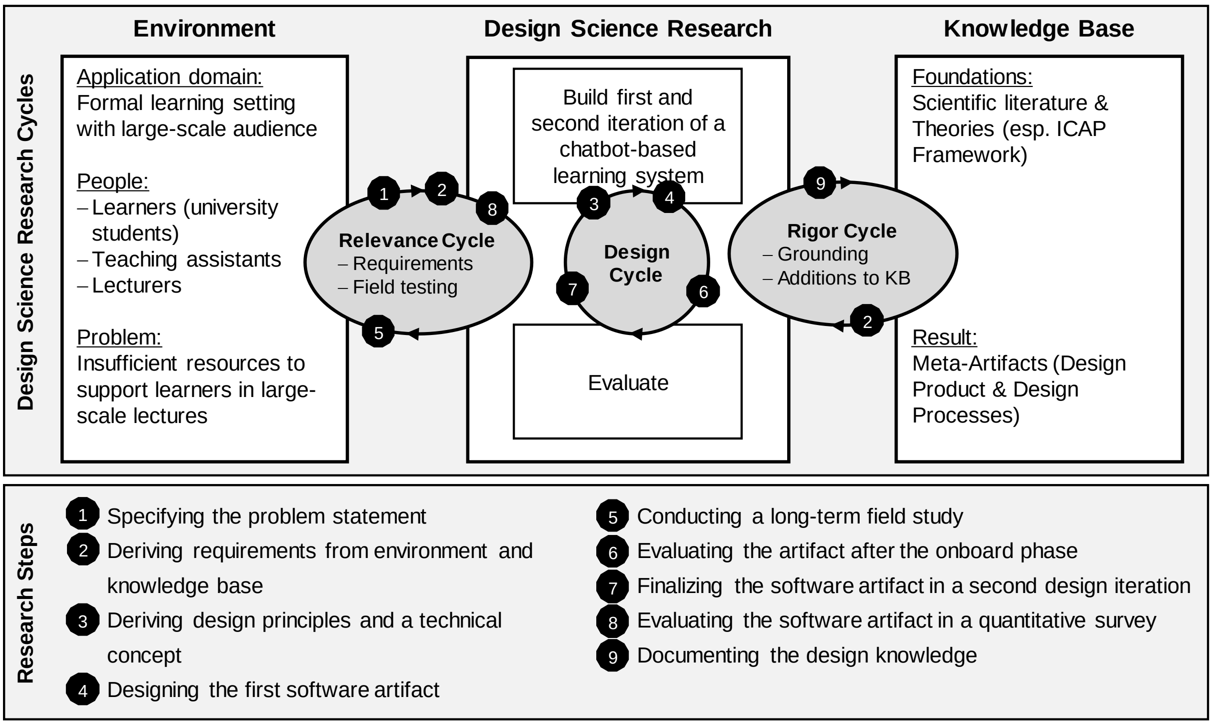

Figure 11: Adapted design science research approach (based on Hevner et al., 2004; Hevner, 2007)

3.4 Designing and Evaluating the Online Learning Environment

In the following section, we outline the results of our iterative design science research approach. As a result, we present the developed artifact (design science level 1 contribution) as well as a proposal for a nascent design theory (design science level 2 contribution).

3.4.1 Specifying the Problem Statement

In this paper, we focus on addressing the issue of insufficient individualization and feedback in large-scale formal learning settings. As noted in the introduction, this issue could possibly be solved by one-to-one lecturing. However, due to resource constraints (Hien, Cuong, Le Nam, Nhung, Ho Le Thi Kim, & Le Thang, 2018), particularly the lack of lecturers and teaching assistants, one-to-one lecturing or at least lecturing in small groups cannot be achieved. This is particularly true for times that are intended for self-study (e.g., at home). During these times, students usually do not have the option of getting fast and instant support on the questions that may arise. Thus, the learning process cannot be supported in a suitable way.

As illustrated in Figure 12 (left), the classical settings in lectures consist of three main actors: students, lecturers, and teaching assistants. Whereas lecturers give lectures that are often content-oriented and focus on imparting new learning concepts to a large audience (up to several hundreds of students in large-scale settings), teaching assistants usually focus on deepening and practicing the learning content in tutorial sessions (usually for approximately 30 students or fewer). Lectures usually allow only limited interaction between students and lecturers, and thus, individualization and feedback are rather impossible. In tutorial settings, the student-to-teaching assistant ratio is better. Nevertheless, the common group sizes often only allow group-based individualization and feedback. An ideally desired one-to-one tutoring can still not be achieved. Finally, lecturers and teaching assistants often communicate (e.g., in weekly meetings) to exchange problems and possibilities for improvement. However, this only allows macro-level feedback without focusing on the students' individual needs. Thus, we identify three major communication flaws among students, teaching assistants, and lecturers in large-scale lecture settings:

Communication flaw #1:

Only limited interaction among lecturers and students in large-scale lectures possible due to a high student-to-lecturer ratio.

Communication flaw #2:

Only group-based individualization is possible in mid-size tutorial sessions. Individual one-to-one tutoring impossible due to still too high student-to-teaching assistant ratio.

Communication flaw #3:

Exchange between lecturers and teaching assistants only in an aggregated way possible without focusing on the students' individual needs.

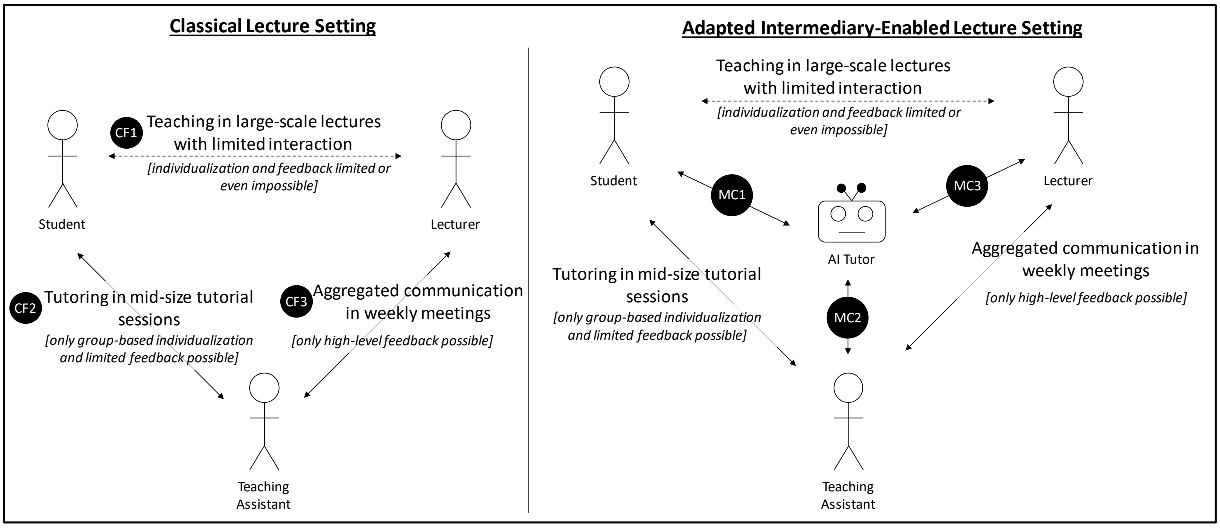

Figure 12: Lecture settings: classical settings vs. intermediary enabeld setting

To address the identified communication flaws, we suggest to include a digital tutor that acts as an intermediary between the three actors. This allows the students, teaching assistants and lecturers to interact in three supplementary mediated communications and thus, improves the possibilities to provide individualization and feedback. Each of the three types of mediated communication enables one actor to receive input from the two others.

Mediated communication #1:

Students interact with the digital tutor in simulated one-to-one tutoring settings based on the conversational agent principles as described in Section 2.

Mediated communication #2:

Teaching assistants get analyses about their students' progress during the tutorial sessions.

Mediated communication #3:

Lecturers get access to analysis of the students’ learning progress and state of knowledge. Additionally, individual feedback can be retrieved.

In the following subsection, we address the three proposed mediated communication types in detail by deriving requirements from the environment and knowledge base.

3.4.2 Deriving Requirements from the Environment and Knowledge Base

As the basis for the design process, we deduce requirements for the specified problem of integrating a digital assistant as an intermediary into the large-scale formal educational setting. We consider both rigor and relevance by analyzing the environmental perspective as well as the knowledge base.

As described in Section 3.2, we base our design science research study on the ICAP framework (Chi & Wylie, 2014) as our kernel theory. Additionally, we follow the principle of computers as social actors (Nass & Moon, 2000) in designing the interaction between students and the digital tutor (mediated communication #1). To incorporate the environment in the design process, we conducted an expert workshop with lecturers, educational researchers, e-learning experts, and instructional designers. Following the first steps of a design thinking process, we discussed possible solutions for addressing the problem statement of this paper. Based on an open idea creation phase, we discussed possible problem solutions, needed functionalities and user stories of all actors. An overview of the whole derivation process is shown in Figure 13.

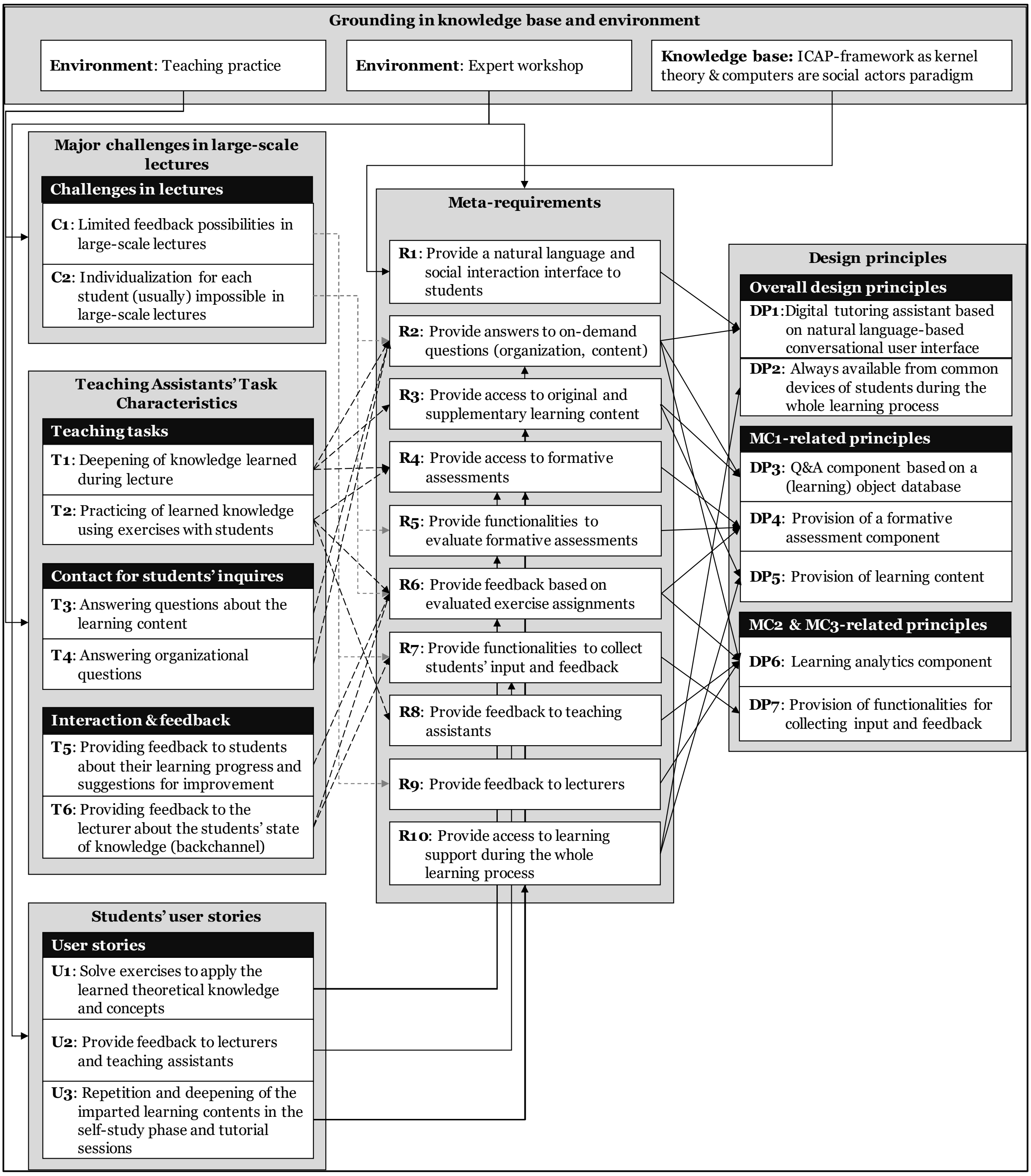

Figure 13: Overview of requirements (step 3) and design principles (step 4)

3.4.2.1 Major challenges

First, we identified two major challenges in large-scale lectures derived from the teaching practice and the expert workshop. As already outlined in the introduction as well as in the problem identification phase, limited feedback possibilities (Challenge C1) and insufficient possibilities to allow individualized learning (C2) in large-scale lectures are major challenges in large-scale formal educational settings (see Section 3.1 and Section 3.4.1 for detailed information).

3.4.2.2 Task characteristics of teaching assistants

Second, we derived the task characteristics of teaching assistants from the teaching practice. In common tutorial session settings, the overall goal of teaching assistants is to supplement the lecture by supporting the students during their learning progress. Tutorial sessions are usually characterized by midsized groups in which about 30 students participate. This allows a more interactive teaching setting, compared to large-scale lectures. In those settings, we derived three main categories of teaching assistants' tasks that we consider: First, teaching assistants have to fulfill teaching tasks. Particularly important is that teaching assistants focus on deepening the knowledge that students learned during the lecture (Task T1). To this aim, teaching assistants often repeat the major topics briefly and extend them with supplementary explanations. The second task is to practice the learned knowledge and concepts with the students by solving exercises or discussing solutions of homework assignments (T2).

As the second major group of tasks, teaching assistants act as the primary contact person for students' inquiries and questions. As teaching assistants usually supervise a mid-sized group of students, a more communicative relationship exists. Thus, teaching assistants often act as the primary contact person for answering questions about the learning content (T3) and answering organizational questions (T4).

Finally, teaching assistants take over tasks related to interaction and feedback. Most important is that teaching assistants have the possibility to provide at least a group-based feedback to the students about their learning progress and suggestions for improving their learning process (T5). Finally, they often also have the task to give the lecturer an overview of the students' state of knowledge and questions and topics that remain unclear during the lecture times and tutorial session (T6).

3.4.2.3 Students user stories

Third, based on input from the expert workshop, we identified three user stories describing how students are expected to interact with the intermediary-enabled digital tutor to be designed. We focused on the students' actions in the expert workshop, as the whole intermediary-enabled learning process targets improving the students' learning process. The first user story involves solving exercises and homework assignments (user story U1). In this story, students apply theoretical knowledge and concepts by solving exercises thar particularly correspond to teaching tasks T2 and T5. In the second major user story, students get and provide feedback to lecturers and teaching assistants (U2). This is particularly important in formal educational settings, as it enables students to influence the way lecturers and teaching assistants impart knowledge and concepts. In small- to mid-sized learning settings, this can usually be done orally in student-to-lecturer communication. However, in large-scale learning settings, this is not possible (see challenge C1). Finally, repetition of the imparted learning content is identified as the third user story (U3). As lectures are often designed for students to also learn in self-study phases (e.g., as a repetition for the final exam), students need to repeat the learning content individually and/or during tutorial sessions. During the tutorial sessions, they can also ask the teaching assistants for help. Thus, the user story corresponds to the teaching assistants' tasks T1 and T3.

The remaining two teaching tasks that are not covered in the students' user stories are T4 and T6. They are both particularly important for teaching assistants and lecturers but less important for students. When teaching assistants provide students with all needed information about organizational issues (4), this reduces the lecturers' workload. In T6, the interaction between teaching assistants and lecturers is described. Thus, students are not directly involved. Consequently, we did not incorporate both tasks in the students' user stories.

3.4.2.4 Meta requirements

The first and overarching meta-requirement that we derived directly from the knowledge base and the problem setting is to provide a natural language and social interaction interface to students (requirement R1). We derive the benefit for natural language interfaces from the

ICAP framework (Chi & Wylie, 2014), as it is expected that human-like communication between students and the system can create an interactive engagement. This engagement form might be beneficial for the learning outcome (see Section 3.2). The social interaction component of the interface can be derived from the CASA paradigm, as outlined in Section 2. Social interaction that uses social cues (see, e.g., Feine, Gnewuch, Morana, & Maedche, 2019) might be beneficial, as students might recognize the system as a social actor. We expect that this might even strengthen interactive engagement and the learning outcome.

The second requirement is closely realted to R1, as it focuses on the basic contents of the natural-language interaction. In order to allow a conversation about learning-related topics, the system needs to be able to answer on-demand questions concerning the organization of the lecture, and tutorial sessions, and, most importantly, concerning the learning content (R2). This requirement can be derived directly from the teaching assistants’ task characteristics, and it addresses the challenge of limited possibilities for individualization in lectures. By implementing this requirement, a system can respond to students’ questions individually, which will address this challange.

In order to provide effective answers to the students’ questions and to target the teaching assistants’ task to deepen the knowledge of the lectures, students need to be able to access the original learning contents. As in typical tutoring sessions, the system also should be able to provide supplementary learning content like a human tutor would if students need further information to understand a given topic. We summarize both as requirement R3.

During tutorial sessions, teaching assistants usually focus on practicing the knowledge and concepts that were imparted during the lecture (T2). To achieve this with an intermediary-enabled digital tutor, the system needs to provide access to formative assessments (R4). Furthermore, it should be able to automatically evaluate the assessment (R5) and give students feedback based on the results (R6). These requirements, particularly automatic evaluation and feedback provision, directly address challenge C1 of limited feedback possibilities in large-scale lectures. As an automated digital tutor is capable of evaluating the results of all students independent of the number of students or submitted assignments, this challenge no longer depends on resource constraints (e.g., the number of available teaching assistants or class-sizes).

However, the limited feedback possibilities in large-scale lectures not only address the missing capabilities of lecturers to provide individualized feedback to every student, but it also describes the challenge of students giving the lecturer feedback. To address this, the digital tutor should act as an intermediary to collect input and feedback from students (R7) and provide it to the teaching assistants (R8) and lecturers (R9).

As the final and overarching requirement, learning support should be provided by the digital tutor during the whole learning process, even when no human teaching assistant or lecturer is available (R10). In order to give students the possibility to learn whenever it suits them best, the system should be available 24 hours a day during the whole lecture period.

Figure 13 depicts the derivation process of the requirements based on the major challenges, teaching assistants' task characteristics, and students' user stories in detail.

3.4.3 Deriving Design Principles

Based on the ten requirements for intermediary-enabled digital tutoring systems for large-scale lecture settings, we derived seven design principles that we grouped into three categories.

First, we identified two design principles that target the overall system design of the digital tutors and not only specific mediated communications (MC1 or MC2 and MC3). Grounded on the requested natural language interface with social interaction (R1) and the questioning and answering of students' requests, we define as the first design principle that the digital tutor should be based on a natural language-based conversational user interface (DP1). To achieve this, the digital tutor must include a natural language processing component to enable communication with the users. As the digital tutor should be available during all times of the learning process, independently of the students' current location (e.g., in the university lecture room or at home), the digital tutor must be available all the time and be accessible using common devices that are used by the students (DP2). These overarching design principles define the technical framework of this paper.

Specifically related to the digital tutors' goal to improve learning are the remaining principles. The category MC1-related principles encompass design aspects that address the communication between students and the digital tutor. Grounded in R2 and R3, a questioning and answering component is needed, which is able to respond to the students' questions (DP3). In order to enable this, the system needs access to databases or a knowledge base containing the needed information. Particularly important is the availability of learning objects that cover the lecture's content as it is likely that students will request related information. Additionally, organizational issues should be encompassed in the database as related questions are to be expected as well. From a technical perspective, a Q&A component also requires functionalities that enable the digital tutor to understand the students' natural language input (known as natural language understanding) and generate appropriate responses (known as natural language generation).

The second design principle in the MC1-related principles is the provision of a formative assessment component (DP4). From a learning-oriented perspective, this is particularly important as practicing the knowledge and concepts imparted in the lectures is one of the key tasks of a teaching assistant (T2). This design principle builds on top of R4, R5, and R6, and encompasses all related functionalities of providing, evaluating, and responding to students' practicing tasks. Closely related to the training tasks of students is the provision of learning content (DP5), which is required when students want to deepen their knowledge. This is particularly important from a learning perspective when students failed to solve the provided exercises. This component should enable them to close the remaining knowledge gaps.

Finally, we derived two design principles that are related to mediated communications 2 and 3. We combined both communications to one category as both types of communication — those between the digital tutor and teaching assistants and between the digital tutor and the lecturer — are similar from a function-oriented perspective. Nevertheless, the subsequent implementation might provide them different data. To be able to provide teaching assistants and lecturers the possibility to get feedback about the students' learning processes and progress, learning-related data can be analyzed. This so-called learning analytics enables the analysis of learning processes and progress (Siemens & Long, 2011) and can be used, for instance, to identify "at-risk learners" (Siemens & Long, 2011), problematic topics that should be deepened in the lectures or derive recommendations for improving "learning habits" (Siemens & Long, 2011). The design principles can analyze data from formative assessments (R6), the Q&A component, or the learning content (R2) that were requested by students. Thus, a comprehensive overview of the learning processes can be gathered. Even though learning analytics enable it to get deep insights into the students' learning activities, direct feedback cannot be collected. However, as this was a major challenge (C1) as discussed in the problem identification step, we propose that functionalities are needed to collect direct feedback (DP7). This is also grounded in R7 and T6.

3.4.4 First Iteration: Designing the Software Artifact

Based on the derived design principles, we started the first design iteration. To this aim, we elaborated on the underlying software architecture depicted in Figure 14 that allows us to address the three mediated communications by considering the seven design principles.

The overall goal of this design science research study is to improve the learning support of students in large-scale lectures. Thus, enhancing this learning support represents the key element of our conceptual design and system architecture. As shown in Figure 14, the students interact with the digital tutor student interface that is designed as a chat-based user interface as requested by DP1. The chat-based user interface represents the main element of human-tutor interaction and is controlled using written natural language communication (DP3). Thus, students can write messages to the digital tutor that are automatically processed. In addition to the natural language interface, we also decided to include content overlays for easier access to some information and a sidebar with clickable menu entries that allow easy access to core functionalities (i.e., request a formative quiz by clicking on the menu entry). Even if some information, such as quiz questions, are opened in content overlays, the questions and the students' selected quiz solutions are also included in the chat history. Therefore, the main interaction remains in the chat-based interaction.

The automatic processing of the students' written messages is done in the natural language processing component using intent and entity recognition. The messages' intents and recognized entities are then forwarded to the dialog messenger, which is responsible for gathering information to generate a proper response in the natural language generation step. To this end, the dialog manager can request information from the learning object data storage component, which stores all required learning materials such as lecture notes, work materials, and lecture recordings. Additionally, a database of micro contents is included, which stores all relevant definitions, formulas, and explanations for key terms mentioned in the lecture as well as organizational information concerning the tutorial sessions, the lecture, and the final exam. These small text components that were specifically written for this purpose are the main basis for generating natural language answers to the students' questions. The other materials, such as lecture slides or recordings, are also used to respond to the students' inquiries. However, the digital tutor is only capable of delivering them to the students as they are encapsulated in files (e.g., PDF files or video files). The learning object data storage is thus based on DP3 and DP5.

To allow students to train their knowledge practically, we implemented DP4 and the so-called formative quiz question storage which contains closed- as well as open-ended questions. The component is capable of delivering students new quiz questions and can analyze them, as well, to provide feedback about the students' success in solving them.

To collect feedback and input from the students and deliver it to the teaching assistants and lecturer as requested by DP7, we implemented an audience response system as part of the digital tutor. If teaching assistants or lecturers request input from the students, the digital tutor is able to notify each student for this request. As in common audience response systems, the students are able to respond anonymously, and the teaching assistants and lecturer receive the collected input in an aggregated visualization.

To provide further feedback (DP6), all messages sent by the students and the digital tutor are stored in the discourse storage component, and other learning actions (e.g., quiz results, click behavior, etc.) are stored in the learning action log. This data is used for analyzing the students' learning actions to provide them, in an aggregated form, to the lecturer and teaching assistant as a learning analytics component.

Finally, teaching assistants and lecturers can access their interface to manage all data storages (e.g., learning objects, quiz questions, and audience response questions) and assess the resulting visualization of analysis (e.g., learning analytics, students' quiz performance, etc.). While lecturers get full access to all data (mediated communication #3), teaching assistants only can view a subset of information that is necessary for improving their tutorial sessions as described in mediated communication #2.

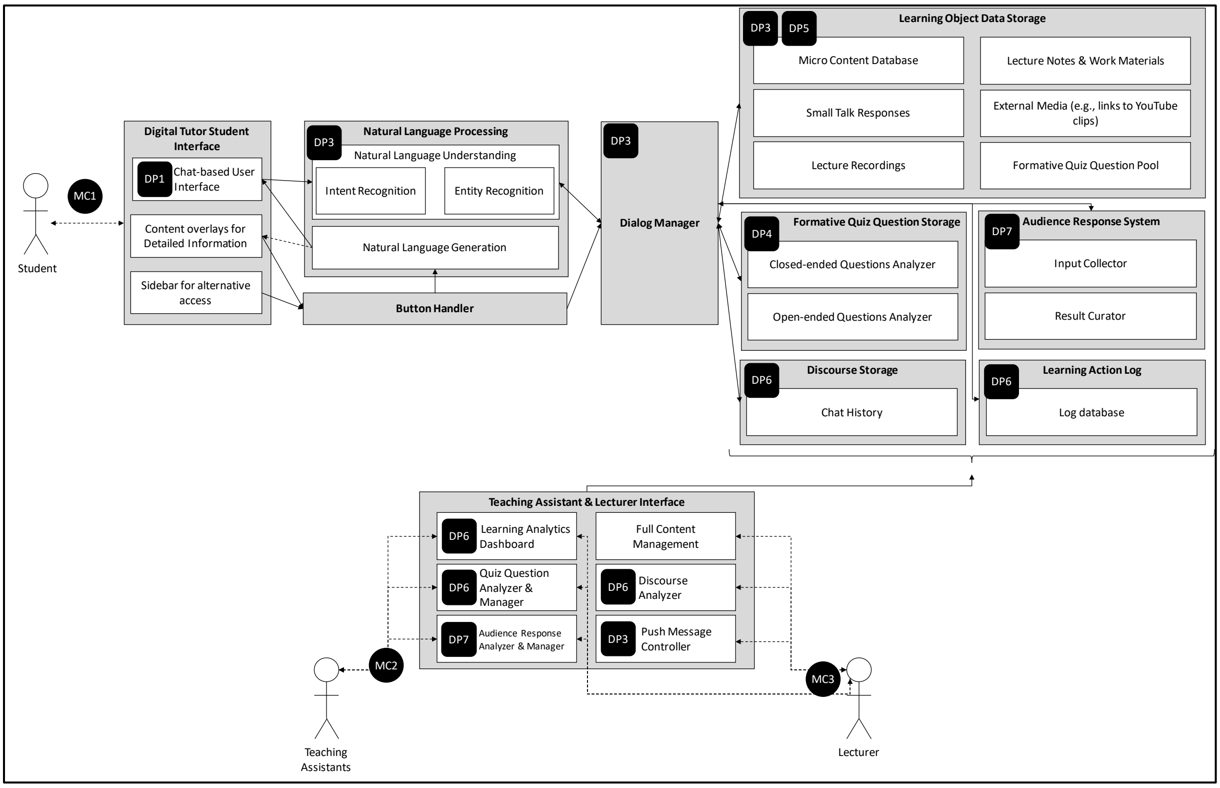

Figure 14: Overview of the system architecture

To implement the system architecture in our software artifact by enabling a ubiquitous availability, we developed a progressive web app that is based on state-of-the-art user interface frameworks and techniques. For instance, we used the Bootstrap framework for designing the students' interface and used AJAX to call the needed functionalities via RESTful web services from our backend infrastructure. Using this approach, we implemented the user interface and interaction logic. Figure 15 shows screenshots of our digital tutoring system.

Figure 15: Overview of digital tutor (students' view)

The natural language understanding is based on the open-source library NLP.js by AXA Shared Services Spain S.A. (2019). Finally, the teaching assistants and lecturer interface is implemented in a separate technical component based on the AdminLTE template (Github, 2020) (see Figure 16).

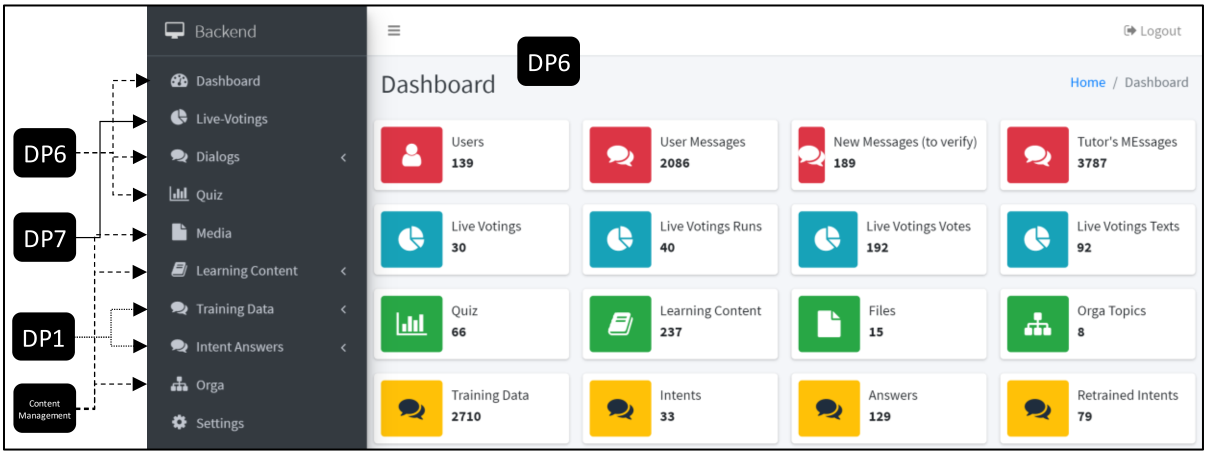

Figure 16: Overview of teaching assistants and lecturer backend

3.4.5 Demonstrating the Artifact in a Field Setting

To evaluate the developed software artifact and the underlying design principles, we introduced it to the students of an introductory lecture on statistics for social sciences in spring 2019. It gave us the possibility to demonstrate the digital tutor to approximately 700 students. The participants were distributed among more than ten different study programs (and many different combinations). Due to the selection of a course not in the information systems or computer science discipline but in social sciences, we aimed to not get a bias on technology affinity.

The demonstration phase was divided into multiple phases. In the first phase, the lecturer of the course introduced the chatbot-based learning system to the students and provided them access codes. Thus, it could be ensured that only students participating in the lecture logged in into the system. The students had some time to make themselves familiar with the capabilities of the system. However, as the lecture period just started, only few learning contents were made available. Thus, this phase can be defined as the onboarding phase. At the end of this phase, a random selection of the students was invited to participate in a qualitative interview. The students were questioned about their first experience with the chatbot-based learning system, their expectations, perceived flaws in the systems' design, and suggestions for improvements. Approximately 90 students were contacted, out of which approximately 16 percent agreed to participate in the qualitative interviews. The results of this qualitative interview study are outlined in the following Subsection, "Evaluating the Artifact after the Onboarding Phase".

Based on the results, we revised the artifact in the second iteration before we continued with the second phase of the demonstration. The students interacted with the revised artifact for the remainder of the lecture period. To evaluate the revised artifact, we conducted a mixed-method study (quantitative survey as well as qualitative interviews) among the students participating in the last lecture before the final exam. The results of this evaluation are described subsequently in the Subsection “Final Evaluation of the Software Artifact”.

3.4.6 Evaluating the Artifact after the Onboarding Phase

After introducing the chatbot-based learning system in the first lecture, we interviewed 14 students to get early feedback. Overall, the students rated the idea of using our chatbot-based learning as an interesting new possibility to support their learning efforts, for example, “I really like the system and I'm glad to be able to try it out [during this lecture period]. […] I like using it” (Student_8). The students’ expectations varied from very positive to skeptical, for instance, Student_6 “hope[s] that this program can help me in the same way as a real person would do”, and Student_5 “tried out the system and asked some question. I hope to receive information faster and this is sometimes certainly the case. If I need a certain definition [of term], I do not need to look it up manually in the lecture slides”. In contrast, some students, such as Student_5, were more skeptical, as their first experiences were slightly negative: “I actually don’t expect too much from it, because my first experiences are that the system answers a lot of nonsense or doesn’t understand many questions”. Even though some minor issues arose, such as Student_3’s complaint that “It is still a little bit buggy, a little bit new and a bit unusual to learn this way”, the students’ expectations were positive: “I hope that the system can answer my urgent questions if I am really desperate while learning and I cannot find a solution in a book” (Student_3).

Implementation flaws:

Based on the students’ feedback, we derived three issues. Two of them addressed our specific implementation of the design principles, and one addressed a didactical consideration of formative quiz results.

First, some students complained about the system’s failure to understand their questions and to respond with useful answers (implementation flaw 1): “The system gives some stupid answers" (Student_11). This issue was mentioned by students who interacted with the system during the onboarding phase. It needs to be mentioned that some students that we interviewed also acknowledged that they tried to trick the system. For instance, Student1 reported that "it's quite funny to write the system a few messages and see what it replies or to provoke it a little bit and see if it can react". Despite some complaints about the system's capabilities and the experience that it "answers a lot of nonsense" (Student_1), the student acknowledged that "if it gets really smart, then I will certainly continue to use it". We received similar feedback from other students that improvements in the natural language generation and understanding would improve the overall experience. For instance, Student7 suggested revising the answers by providing more details: "The system should give somehow clearer answers to certain questions. For example, […] [if I] request certain information about [organizational aspects like] the lecture or the tutorial session" (Student_7). All remarks by the students participating in the interviews concerning the natural language capabilities did not question the usefulness of a chat-based system, but some did complain about its capabilities. Thus, we addressed this issue in the second design iteration (see next subsection).

The second issue raised in the interviews addressed the usability of the chat-based user interface (implementation flaw 2). As all actions executed by the students in our chatbot-based learning system result in a message in the chat history, the number of messages can be quite high. For instance, Student14 "request[s] a cleaner interface. At the moment, I find it rather confusing if you want to find certain things […] you have to scroll up [in the chat history]". In a similar way, Student8 suggested improving the visualization of available formative quiz questions. In the first version, the system responded with a list of all available quiz questions if a student asked. However, the list was quite long — even after the first lectures — because up to 20 quizzes were provided per lecture. Thus, Student8 suggested that the system should only show a list of quizzes "which are missing or show a table or a diagram, which quizzes you have already done" (Student_8). Based on this feedback, we revised the implementation of the chatbot-based user interface (see the next subsection).

Finally, in addition to the two issues that focused on the implementation, we also received feedback concerning one didactical aspect. During the conceptual design, the lecturer of the course in which we conducted the field study decided neither to reveal solutions to formative quiz questions nor to show in detail why a student's submitted solution was not correct. In contrast to that, the lecturer decided that students should be encouraged to try solving the quizzes again until they solved it on their own. This decision was based on didactical considerations as the lecturer worried that otherwise, students could just submit any random first solution to a quiz without trying to solve it correctly just to receive the correct solution. Nevertheless, some students such as Student_8 complained about it: "The only thing I would like to improve would be […] that you get the solution or solution path". However, this student also acknowledged that "every teacher […] can argue whether this makes sense". Others also supported this aspect: "I find it a bit difficult with the quizzes that you just don't have any solutions or something like that. Because I think it's up to everyone to decide whether they want to look at the solution or not and that's why I think it's kind of stupid that you can't see it" (Student_12). Even though the suggestion to show correct solutions to formative quiz questions was mentioned by multiple students, the lecturer maintained his didactically justified design decision. Thus, we did not revise the response behavior of the chatbot-based learning system in this case. However, we explained the reason behind this decision to the students.

3.4.7 Second Iteration: Revising the Implemented Software Artifact

As outlined above, particularly two issues needed to be addressed in the second design iteration. The overall idea and general concept were rated as suited for supporting the students in the large-scale lecture scenario. Particularly critical were the mentioned issues regarding the chatbot's capabilities to understand the students' questions and to respond adequately using high-quality learning contents (implementation flaw 1).

To address this first implementation flaw by improving the quality of the natural language processing, we (1) manually reviewed and reclassified the students' input and (2) set up a quality assurance process. In the beginning, the digital tutors' natural language understanding capabilities relied on the pre-trained NLU model (see first design iteration). Initially, we trained this model using manually created training data. To this end, approximately 3,000 training entries were used to create the first NLU model. The first experiences of the students relied on this data set. As the students reported that they rated the NLP's performance too low, we manually reviewed all messages sent by the students and reclassified them if needed. In doing so, we could extend the training set massively. By integrating the dialog review component, the lecturer and the teaching assistants had the possibility to access the dialogs between students and the digital tutor (step S1 in Figure 17). For each student, the messages could be seen (S2 in Figure 17) and reclassified (S3 in Figure 17). By integrating this dialog reviewing component into the digital tutor's backend, no technical skills are required to improve the NLU.

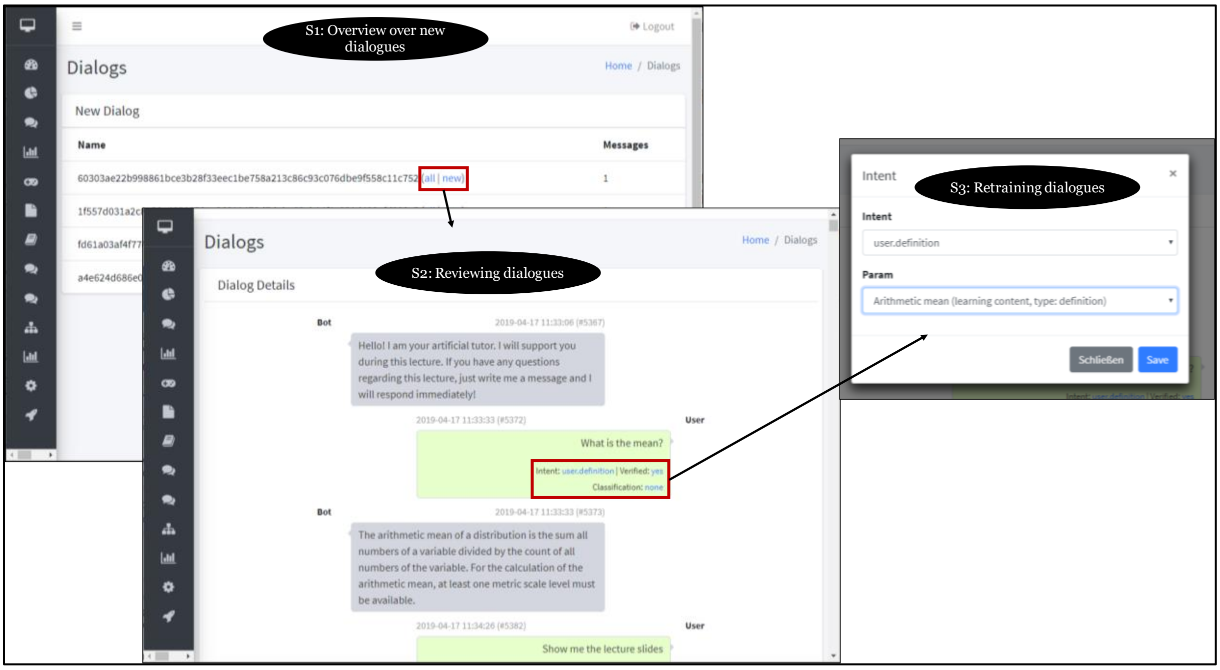

Figure 17: Overview over dialog review and retraining procedure

Based on the manual review, we also initiated an enhanced NLU process. Figure 18 presents and overview of this process. Instead of relying only on the trained NLU model, we also incorporated the newly available data of the manually reviewed data. If a student's message was received by the digital tutor, it was processed by the NLU component. The NLU component now relies on all available data: (1) The initial training data, (2) the manually reviewed messages from the discourse storage, and (3) the trained NLU model. If the student's message is identical or similar to a message (i.e., only contains minor typing errors) already stored in the initial training data or in the reviewed discourses storage, the NLU model will be skipped, and the stored intent will be used as the result. If the message is unavailable in the first two data storages, it will be classified using the trained NLU model. The NLU model calculates the most probable intent, which will be used as the result. If the NLU model has a low certainty whether the result is correct (e.g., lower than 80 percent), the intent will be sent to manual reviewing. In this case, a teaching assistant will review it and, thus, improve the NLU capabilities instantly. The effectiveness and applicability of this procedure become apparent by analyzing the messages that needed to be classified during the whole study. Whereas in the first week, almost 1,000 messages needed to be classified manually, the number decreased rapidly to less than 200 in the third week — an easily manageable number.

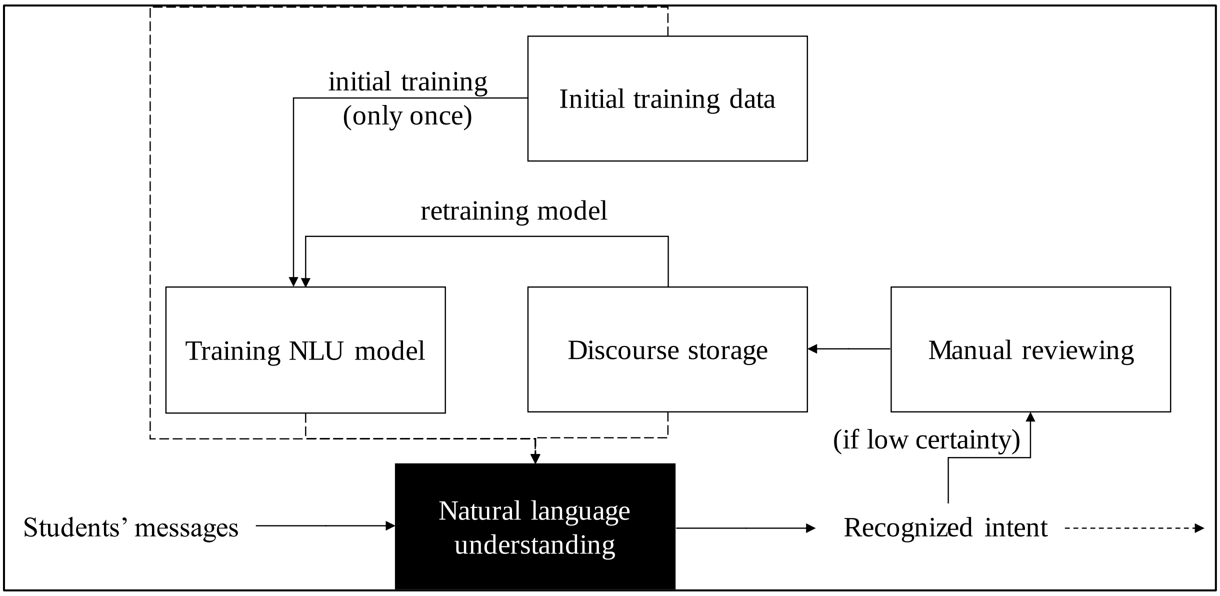

Figure 18: Overview of the enhanced NLU process

In addition to the enhanced NLU process, we also adjusted the digital tutors' knowledge bases by inserting new information concerning organizational aspects (e.g., details about the organization of tutorial sessions), requested small talk, and additional learning contents. Thus, the digital tutor could respond to the students' messages in a more satisfying way.

To address the second implementation flaw — the usability of the chat-based user interface — we carefully analyzed exemplary dialogs from students to identify problems. According to the students, there was too much information displayed in the message list in some cases. For instance, if students asked the digital tutor for formative quiz questions, the digital tutor responded with a list of all available questions. This list was quite long and could result in an information overload. To reduce the list and improve simplicity, the digital tutors' answer was shorted, and students could select to only see quiz questions related to a specific topic or lecture. Additionally, the digital tutor automatically asked the student to continue with a new, yet unsolved quiz after the student completed one quiz successfully. If a student failed to solve a quiz, the digital tutor proposed that the student retry the quiz. Compared to the first implementation, this adapted behavior of the digital tutor (1) reduces the displayed information substantially, as only selected quiz questions are shown (i.e., only the quiz questions that are not yet solved by the student), and (2) simplifies the usage as the digital tutor now recommends further actions (e.g., retrying a quiz or working on the next unsolved quiz).

As a result of fixing both implementation flaws identified in the qualitative evaluation during the onboarding phase, we finished the second design cycle. Whereas the second implementation flaw only addressed minor changes in the user interface and chat behavior, enhancing the natural language understanding process, including a quality-assurance process, is a substantial modification. Thus, we deduce an additional design principle DP1b (providing an easily manageable quality-assurance procedure for effective natural language processing) based on this as an addition to DP1. Finally, we deployed the revised software artifact to the field settings directly after the onboarding phase.

After the deployment of the revised software artifact, it was instantly available to all students participating in the field setting. As the software is implemented as a web-based application, all students directly received the update immediately.

3.4.8 Final Mixed-Method Evaluation of the Software Artifact

In the following, we outline the final evaluation results of the software artifact within the field setting. First, we give an overview of the students' usage of the chatbot-based system. Next, we focus on the evaluation of the design principles based on a quantitative survey distributed to the students during the last lecture setting. Finally, we assess the students' written qualitative feedback concerning suggestions for improvements and discuss the results of 21 interviews with students who used the system during the field study.

All of these parts of our final evaluation of the software artifact are the foundation for the design knowledge, which is documented in the last subsection.

3.4.8.1 Overview of the students' interaction with the digital tutor

During the field study, approximately 700 students used the system and initiated over 37,000 user sessions. Thus, each student initiated more than 50 sessions on average during the field study — which means on average the students used the system at least every second day. During the sessions, the students solved more than 80,000 formative questions (more than 110 per student), and the chat history log contained more than 200,000 messages initiated by the students. Despite these interactions that are related to the mediated communication #1 between the students and the digital tutor, the mediated communications #2 and #3 were addressed as well. The digital tutor collected more than 9,100 responses using the audience response system in in-class settings (i.e., lectures and tutorial sessions). Additionally, approximately 1,100 audience response feedbacks were collected in self-study times.

Overall, the students' actions within the system resulted in approximately 1,800,000 log entries in the learning action log.

3.4.8.2 Quantitative evaluation of the design principles

To evaluate the design principles, we asked all students participating in the last lecture to evaluate our chatbot-based learning system in a quantitative survey. Approximately 270 students attended the last lecture. Out of them, approximately 170 filled out our questionnaire, which we consider a good participation rate.

Besides some typical evaluation metrics about the overall lecture experience and the lecture contents, we focused particularly on the evaluation of our digital tutor, which we outline in the following. For each design principle, we created at least two questions concerning the implemented design principles, with an exception to DP6. As DP6 is not available for students, it cannot be evaluated by them. Table 1 gives an overview of the design principles, the corresponding items, and its evaluations.

Table 2: Overview of the evaluation results

Item	N	ø	σ	Visualization of the average rating
DP1: Digital tutoring assistant based on a natural language-based conversational user interface (including the enhancements of DP1b)				
Personal tutor	161	5.38	1.48	
Natural language communication	162	5.42	1.80	
DP2: Always available from common devices of students during the whole learning process				
Location independence	161	4.99	1.85	
Time independence	163	5.99	1.13	
Device independence	163	5.29	1.81	
DP3: Q&A component based on a (learning) object database				
Explaining learning objects	159	5.93	1.45	
Proactive suggestions	158	5.87	1.51	
Providing organizational aspects	159	5.22	1.87	

DP4: Provision of formative a assessment component				
Providing quiz questions	159	6.13	1.36	
Automatic evaluation and feedback	163	6.26	1.26	
DP5: Provision of learning content				
Provision of files	161	6.20	1.22	
Provision of lecture recordings	156	6.08	1.43	
DP6: Learning analytics component				
Not visible to students. Thus, no evaluation by students possible.				
DP7: Provision of functionalities for collecting input and feedback				
Audience response system in lectures	162	5.73	1.52	
Audience response system in tutorials	159	5.49	1.72	
Scale: +1 (strongly disagree) to +7 (strongly agree)				

Overall, the quantitative results show that the students rated all items corresponding to design principles on average positively (i.e., larger than 4.0 on a Likert scale ranging from 1 to 7).

The students acknowledged the idea of providing a digital tutor (DP1) as an intermediary between them and the teaching assistants and lecturer (ø 5.38), including the natural language communication in the provided chatbot-based learning system (ø 5.42). Also, the accessibility of the digital tutor independently of a specific time or place was acknowledged (DP2). However, the location independence was rated substantially less important than the time independence (4.99 vs. 5.99). This is in line with the general idea of providing a digital tutor: The tutor is available to the students in self-learning phases when no human teaching assistant or lecturer is available (see Subsection Specifying the Problem Statement). The results concerning the automatic and immediate answering of questions (DP3) are interesting. Whereas learning-related questions and automatic suggestions were rated as very positively, answering organizational questions was not considered as equally important (5.22 vs. 5.93 and 5.87). The students rated the usefulness of the design principles 4 and 5, which are related to the formative quizzes and the provision of learning content, with an average rating of larger than 6 as very high. Both design principles reach the highest values compared to all other principles in our survey. Finally, the collection of feedback (DP7) was also positively evaluated by the students.

Summarizing the descriptive evaluation results, we conclude that the students perceive the outlined design principles DP1 to DP5 and DP 7 as useful and thus acknowledge the importance of them. To further analyze potential flaws in the implementation of the design principles as well as to identify potentially missing design principles, we further asked the students to give us written feedback in the survey questionnaire. We describe the results in the following subsection.

3.4.8.3 Written feedback evaluation

As a result of asking the students to give us suggestions for improving our digital tutoring system, we received more than 150 suggestions. To analyze the suggestions, we coded them using open coding, grouped them into 23 categories, and rated them based on the number of mentions.

15 categories encompass suggestions from students that were only named up to three times. Thus, those categories are suggestions that only represent the opinion of a few individuals. For instance, one student stated that the digital tutor should also contain a forum for enabling discussions between students, another stated that bilingualism would be nice (even though the lecture and the whole study program are only monolingual), and three

other students stated that the digital tutor should be able to tell more jokes and talk about small talk topics. We consider all suggestions in these categories as less important as they only attribute specific issues that are not considered by a larger number of students or are not directly related to the actual learning process (i.e., telling jokes).

Four categories of suggestions were named at least five and up to 15 times. Some students wish that the digital tutor could serve the lecture recordings faster, and others complained that they would like to access the uncut version of the lecture. These suggestions are not directly related to the actual software implementation but rather address the underlying knowledge base. The videos were only available some days after the lecture because the lecturer wanted to post-process each lecture recording to slice it into short pieces (i.e., microlearning videos) that cover a complete subsection. This suggestion does not affect the design principles but the organizational aspects of the lecture. Finally, eight students suggested implementing a print or export functionality of the quiz questions. As our digital tutoring system is designed as a mobile-first application that is intended mainly for use on smartphones and tablets to learn interactively, we did not consider this suggestion as critical. Twelve students suggested ways to improve the overall simplicity and clarity of the user interface. Of particular important to this issue is that the chat history gets long after solving a large number of quiz questions, and the chat history becomes a bit unmanageable if someone needs to access results of prior quizzes or search for a specific message in the log. To fix this issue, the students proposed (1) implementing a search functionality, (2) enabling them to save messages as favorites for easier access, and (3) reducing the lists of quizzes in the chat history and only show the answered questions. We consider these suggestions to be suitable for improving our software implementation. They can be related to DP1 and address our specific instantiation of a digital tutor that acts as an intermediary between students, teaching assistants and lecturers.

Finally, students made two specific suggestions more than 15 times, making them important and critical. First, they asked for additional learning content via our digital tutor. Even though the tutor covers all learning content taught in the lecture, 16 students desired additional supplementary content. The students proposed several types of desirable content, ranging from hints on how to use mathematical commands in spreadsheets to additional examples for each formula presented by the digital tutor. Strongly related to this is the suggestion to provide solutions for formative quiz questions and to show the full calculation method if students fail to solve a quiz on their own. This suggestion was named by 52 students in our evaluation, showing the importance of a complaint that the first evaluation in the onboarding phase had already discovered. Overall, the most critical suggestions from the students in our survey were related to didactical or content-related aspects. This highlights the importance of the actual learning content and the underlying didactical decisions. Both aspects are unrelated to the actual design principles that we derived in this design science research study.

3.4.8.4 Qualitative evaluation

In addition to the data retrieved from the survey questionnaire (quantitative data and written feedback), we relied on an analysis of 23 conducted qualitative interviews with students. The interviews were conducted with the intent to allow us to further investigate specific issues and aspects in detail. As part of this design study, we used qualitative interviews as an additional data source to supplement the data retrieved from the survey questionnaire. Particularly interesting for this paper are, (1) whether our digital tutor was able to support the students' learning process and success and (2) to explore the didactical issue of not showing solutions to formative quiz questions in more detail, as students criticized this in the survey questionnaire.

To investigate whether the digital tutor successfully supported the students in their learning process, the interviewer first asked the students in an open manner if they could get support from the digital tutor. To further deepen the insights, we particularly focused on the natural language chat conversation, the audience response system for collecting feedback, and the formative quizzes.

Overall, the qualitative interviews with the students support the quantitative results. Students stated, for instance, that the digital tutor "supports my learning, as the digital tutor offers me good guidance" ($\text{Student}_{\text{final}}$23), and the digital tutor "certainly supports me. I use the formative assessments often, and if I do not know how to proceed or if I cannot find a specific term on the lecture slides or if I do not understand a specific term, I use the digital tutor [to get an answer] ($\text{Student}_{\text{final}}$7). Other students particularly highlighted the availability of the digital tutor: "It's great that the digital tutor is available 24 hours a day and that you can access it at any time. That's pretty good, and I think the digital tutor can help me with learning!" (Student$_{final}$6). Even though the responses were quite positive, some students also were a bit more skeptical, such as Student$_{final}$2, who said that not everything is perfect, but "what is really useful is that you can ask for specific terms and that you can practice exercises". A few other students claimed that they did not participate a lot in the course at all. Thus, they did not communicate much with the digital tutor. Nevertheless, the overall impression was positive, as one student said: "I cannot really say [whether the digital tutor supports me] because I have not invested a lot of time lately, but when I learned something, the digital tutor supported me".

Besides the natural language communication, which allowed the students to ask the digital tutor content-related questions and the formative quizzes, the students also evaluated the audience response functionality as useful because it allowed them to provide the lecturer and the teaching assistants instant feedback during in-class lectures and tutorial sessions. Student$_{final}$23 described this as follows:

> *"In the lecture, the digital tutor uses live votings [..], and I think that's really good, […] because you're forced to participate in the lecture somehow. Yes, and the feedback is also supposed to identify what we [the students] want to do in the lecture. I think that's really good. And yes, the lecturer can go into more detail about what we actually want to do; I think that's really, really good"* (Student$_{final}$23).

This statement was also supported in similar ways by most of the students interviewed.

Based on the qualitative insights we briefly summarized above, we conclude that the qualitative interviews supported the quantitative survey results. The statements of the interviewed students further emphasize the usefulness of the digital tutors' capabilities to answer topical questions, provide learning content and formative quizzes for practicing the exercises, and enable the provision of feedback to the lecturer and teaching assistants. Thus, the corresponding design principles can also be supported. Nevertheless, it needs to be acknowledged that not every student benefits from the digital tutor. For instance, as quoted above, if a student is unwilling to invest time in participating in the lecture and does not learn in self-learning phases, even a digital tutor may not help in this case.

Finally, we investigated the didactical issue in more detail, even though it does not affect the digital tutors' design. The didactical decisions are, however, important to consider for future digital tutoring systems.

In the interviews, we recognized mixed results concerning the fact that the lecturer decided not to reveal sample solutions for the formative quizzes. Some students did not complain at all and rated the quizzes as "extremely good. They are really well done, also with the evaluation results. And I find it very helpful that the digital tutor is able to explain definitions, formulas, and examples" (Student$_{final}$6). The student combined the digital tutor's functionalities to work on the provided quizzes. This approach can be classified as a best practice approach that was desired by the lecturer. The lecturer wanted to achieve students engaging with the different learning contents intensively in order to solve the quizzes.

This approach was, however, not applied by all students. Multiple students who we interviewed rather complained about the fact that the digital tutor is not providing sample solutions. For instance, Student$_{final}$12 stated that "it is sometimes a bit stressful because if I got a quiz wrong, then I need to work on it for a long time until I can solve it. At a certain point, I just start to guess [the correct solution] until I solve it". Similarly, Student$_{final}$19 added, "if I make a mistake in a quiz, the digital tutor just says that it's wrong, but sometimes I still can't find a solution. That's the main thing that bothers me".

These exemplary quotes on the didactical issue of not providing sample solutions show the diversity of the approaches. On the one hand, some students keep on trying to solve the quizzes by utilizing other available features of the digital tutor. On the other hand, some students just resign and try to guess the correct solution. Whereas the first approach is desirable from a didactical point of view, the second one might be disadvantageous for the learning process.

Thus, the qualitative interview results also emphasize the importance of weighing the advantages and disadvantages of an appropriate instantiation of the design principles.

3.4.8.5 Summary of the final mixed-method evaluation study

By considering the mixed-methods insights of our final evaluation — the overview of the usage data, the quantitative evaluation, the written feedback, and the qualitative interview

— we conclude that the students supported the design principles. The students used the digital tutor extensively during the field study and evaluated the functionalities positively. Nevertheless, the suggestions for improvements highlight that the specific implementation can still be improved and is highly related to the learning content provided by lecturers. In some cases — such as formative quizzes — the lecturers' didactically grounded decisions (i.e., to show correct solutions of quizzes in order to encourage the students to retry solving them) are questioned. The conducted interviews further showed that students have different opinions on this topic. The specific didactical instantiation of the design principles is, however, not in the focus of this design science research study. Nevertheless, the didactical aspects should be a topic of future research.

3.5 Discussion and Documentation of the Design Knowledge

The overall goal of this design science research study was to support students during their learning processes in large-scale educational settings. The main issues in supporting learners in large-scale lectures are caused due to resource constraints. Even though the lecturers would like to support each student individually, this is usually not possible if several hundreds of students participate in a university lecture. Thus, lecturers set up tutorial sessions that are led by teaching assistants. These teaching assistants act as intermediaries between the students and the lecturer and are able to support the students in smaller groups. However, the group sizes of tutorial sessions often exceed 20 to 30 students and are therefore not small enough to provide individual learning support. Based on this concept of instantiating intermediaries, we propose the concept of digital tutors who act as intermediaries between all existing actors: students, teaching assistants and lecturers.

By grounding our research on the ICAP framework (Chi & Wylie, 2014) and treating the computers as social actors (Nass et al., 1995; Nass & Moon, 2000), we transfer the technical concept of conversational agents to large-scale educational settings. Whereas the overall technique is similar to conversational agents in other settings, the long-term settings differ in formal education. In typical use cases of chatbots and other conversational agents, the users interact with an agent for specific, time-limited purposes. For instance, users need a specific solution for a problem (e.g., in customer support) or need process guidance (e.g., to book a flight) (Meyer von Wolff, Hobert, & Schumann, 2019). In contrast to that, (digital) tutors should not only be available to solve a particular problem of students once but also support the students during the whole learning process, which usually lasts for several months. Common educational learning processes in academic teaching are intended to last for approximately three to four months (i.e., one lecture term).

Based on this novel use case of digital tutors acting as intermediaries between students, teaching assistants, and lecturers, we derived major challenges, task characteristics, user stories and meta-requirements from both practice and theory. Using the meta-requirements as a basis, we deduced seven design principles, which we implemented in our software artifact in two design iterations. During our evaluation studies in a real field setting, several hundreds of students interacted with our digital tutor. As a result of the first evaluation round in the onboarding face, we (1) fixed some minor issues to improve the overall user experience, and (2) deduced the new design principle DP1b as an extension of DP1. The students' responses outlined the need to instantiate an enhanced quality-assurance process to improve natural language processing. As even large-scale educational settings are comparably small in contrast to the business context in which thousands of customers are potential users of conversational agents, training a digital tutor is more difficult as the number of interactions is more limited. Thus, we introduced a multistage, partially automated quality-assurance process.

In the final evaluation, after a trial period of several months, the students were asked to participate in a quantitative survey in which we evaluate the design principles as well as the students' overall experience with the digital tutor. The results indicate the usefulness of the digital tutor as an intermediary concept between students, teaching assistants, and lecturer. The students approved the design principles with which they interacted.

3.5.1 Proposing a nascent design theory

To document the results of this design process of the digital tutor concept, we use the core components of a design theory by Gregor and Jones (2007). In doing so, we summarize our practical design knowledge to create a theoretical contribution in the form of a “design and action” theory (Gregor & Jones, 2007). Our contribution is based on a rigorous design process, which comprises multiple design and evaluation steps. Table 3 summarizes our systematically derived design knowledge, including rigor and relevance using the following components: purpose and scope, constructs, principles of form and function, artifact mutability, testable proposition, and justificatory knowledge (Gregor & Jones, 2007).

Table 3: Documentation of the design knowledge based on Gregor and Jones (2007)

Component	Description
Purpose and scope	The purpose of the proposed concept of digital tutors as intermediaries between students and teaching assistants and lecturers is to support students in their learning processes in formal educational settings. By addressing existing communication flaws in large-scale settings, particularly the individualization of learning processes and individualized on-demand feedback should be enabled.
Constructs	Chatbot-based learning system as a digital tutor consisting of several constructs: automated natural-language communication for students, knowledge bases, Q&A component, formative assessment component, audience response system, learning analytics component
Principles of form and function	DP1: Digital tutoring assistant based on a natural language-based conversational user interface DP1b: Providing an easily manageable quality assurance procedure for effective natural language processing DP2: Always available from common devices of students during the whole learning process DP3: Q&A component based on a (learning) object database DP4: Provision of a formative assessment component DP5: Provision of learning content DP6: Learning analytics component DP7: Provision of functionalities for collecting input and feedback

Artifact mutability	The artifact can be applied in large-scale lectures that are accompanied by tutorial sessions. Prerequisites for the usage are (1) the availability of a micro content database that contains small explanations of the contents to be learned in the lecture, (2) the availability of a pool of formative quiz questions that are suited for an automated evaluation, (3) an adoption of the concept of the lecture to introduce the digital tutor as an intermediary.
Testable propositions	To test the design principles and implementation, each aspect as mentioned above needs to be surveyed. To evaluate the effects on students' learning performance, the following propositions should be considered: (1) Digital tutors are accepted as intermediaries between students and teaching assistants and lecturers in large-scale educational settings. (2) Digital tutors improve the individualization of learning processes in large-scale settings by responding to the students' individual needs comparable to a one-to-one tutoring setting. (3) Digital tutors improve the possibility to get individualized feedback in large-scale settings.
Justificatory knowledge	Scientific literature, particularly the ICAP framework (Chi & Wylie, 2014) and the computes are social actors paradigm (Nass & Moon, 2000), as well as empirical knowledge, particularly results from a workshop, teaching practice, task characteristics of teaching assistants, and two evaluation studies in a real field setting.

3.5.2 Implications for the Kernel Theory

The aim of this research study was to generate insights and theoretical contributions to the design of digital tutors. We grounded our research on the ICAP framework (Chi & Wylie, 2014) as our main kernel theory. Our aim was not to test the ICAP framework. Nevertheless, we can derive some implications based on our large-scale field study.

The main idea behind creating a conversational digital tutor is to improve the students' engagement with the digital tutor and the provided learning content. The ICAP framework proposes that an interactive learning setting will result in increased engagement and increased learning (Chi & Wylie, 2014). In our field study, we recorded millions of log data entries, thousands of exchanged chat messages, and hundreds of thousands of solved quiz questions by 700 students participating in our field study. The students' interaction with the digital tutor exceeded our expectations by far. Based on prior experiences from multiple other technology-enhanced university courses, we conclude that the usage of the digital tutor successfully fostered the students' engagement. Based on the evaluation results and particularly the qualitative interviews, we have evidence that the digital tutor improved the learning of the students who interacted with them as proposed by the ICAP framework. To further investigate this assumption, future research should be conducted. For instance, comparing the learning success of learning sessions enhanced by digital tutors with classical learning settings in field studies seems promising for this case.

3.5.3 Limitations

As with every design science research study, limitations of the individual research steps need to be discussed. We ground our research on rigor as well as relevance. Particularly, the relevance cycles need to be discussed.

First, we base our problem statement on experiences from the educational practice and particularly, a workshop with experts with lecturers, educational researchers, e-learning experts, and instructional designers. Even though we expect that the problem statement is valid for most universities with large-scale settings, it needs to be analyzed in the future to which settings it can be transferred.

Based on this empirical basis and the theoretical foundation, we derived requirements and design principles. The design principles act as the basis for our developed software artifact in the first design iterations. We conducted a qualitative evaluation study to assess the suitability of the first design iteration. Even though we tried to select a random sample for our interview study and asked the participating students to give us feedback, we cannot ensure that we identified all issues of the first software prototype in this study. Nevertheless, we improved our software artifact in the second design iteration. Our final evaluation took place after approximately 700 students interacted with our digital tutor in a field setting of an introductory course on statistic education. We assume that this setting is typical for large-scale educational settings. However, reproducing the final evaluation in a similar setting would strengthen the design contribution.

3.6 Conclusion

As a result of the outlined design science research project, we propose the concept of digital tutors as intermediaries between students, teaching assistants and lecturers. Over multiple rigorous design and evaluation cycles, we developed a software artifact and proposed a nascent design theory.

We contribute to the IS community in two ways. First, we provide a novel, rigorously developed IT artifact of an IS education system that uses AI technologies to support learners in large-scale courses. According to Baskerville, Baiyere, Gregor, Hevner, and Rossi (2018), novel IT artifacts that are built and evaluated already represent a valid contribution. Gregor and Hevner (2013) rated this as level 1 contributions that are situated in the problem statement. Second, we generalize and document our design knowledge based on the two cycles of artifact development, using core components proposed by Gregor and Jones (2007). Thus, we propose a nascent design theory as a level 2 contribution (Gregor & Hevner, 2013). With these two contributions, our results extend the existing CA and IS Education knowledge base. Additionally, we will publish our intermediary digital tutor as open-source in conjunction with the publication of this paper. Thus, besides the theoretical contribution of our proposed nascent design theory, practitioners and lecturers can benefit from our software artifact.

3.7 Concluding Remarks and Transition to Study 3

The prototype developed and the underlying design theory primarily made it possible to develop a digital tutor that helps students learn, for example statistics. For scientific progress in the domain of information systems, in which the cooperation partner Sebastian Hobert is active, and also for ethical issues, it is important that the system used brings real added value for the learners. Some evidence that this is the case has already been provided and discussed in Study 2. In addition, the fact that the system continued to be used until spring 2023 in the teaching context in which it was developed can speak for the added value of the system. Furthermore, three colleagues, two of them at other institutes, also used the system for their courses.

The added value of the system is also supported by its extensive embedding in teaching in the context in which it was developed. In order to document this embedding and to inspire other lecturers in their digital teaching, a detailed, practically oriented guide for the development and use of such digital tutors was developed (Hobert & Berens, 2021a). Secondly, concepts were tested and evaluated that integrate the digital tutor into digital teaching with further ideas and concepts (Berens, Reusch, & Hobert, 2021; Hobert & Berens, 2021b, 2023). Thirdly, further studies show that various design elements of the digital tutor have positive effects (Hobert & Berens, 2019, 2020, 2023).

All of this shows that the development of the digital tutor focused on developing the best possible learning aid and embedding it in the overall learning scenario. The software developed in this process can be requested from Sebastian Hobert.

At the same time, the compatibility of the software with projects using digital behavioral data through learning analytics was always taken into account when designing the system (Hobert & Berens, 2021b). In particular, very comprehensive offers are available that invite learners to use the software regularly and intensively and thus enable the generation of comprehensive digital behavioral traces. The processing of this digital behavioral data is helped by the fact that the system combines all digital teaching offerings in one system, making data linking easy. The digital system developed in Study 2 thus provides almost ideal conditions for using digital behavioral data of a course for learning analytics.

However, using these prerequisites for actual comprehensive learning analytics projects is not trivial. As the look into the literature has shown, this has not been done very often in the context of non-cognitive factors. And although there is now a great deal of literature on learning analytics, measuring learning behavior is still a complex topic. In particular, identifying complex learning strategies in digital behavioral data is still in its infancy (for work on this, see e.g. Gijsen, Catrysse, Maeyer, & Gijbels, 2024; Jovanović, Gašević, Dawson, Pardo, & Mirriahi, 2017). And there is also surprisingly little literature using learning analytics on topics that have long been researched with other methods. One example of this is spaced learning, for which it has long been known from experimental studies that this promotes learning (Cepeda, Pashler, Vul, Wixted, & Rohrer, 2006). Since digital behavioral data is stored with a timestamp as standard, it is actually well suited for investigating dimensions of learning behavior related to the distribution of learning. Nevertheless, only a few authors have so far take on such work (for exceptions, see e.g. Miyamoto et al., 2015; Yeckehzaare, Mulligan, Ramstad, & Resnick, 2022).

Study 3 of this paper addresses this phenomenon and investigates the effects of spaced learning and learning engagement using digital behavioral data. This first of all fills a research gap that is presented within the study but is not part of this thesis. Nevertheless, Study 3 fulfills an important dual function for this thesis. On the one hand, it serves as an experimental space in which data processing, operationalization of variables and data analysis of the digital behavioral data are tested in the context in which the other studies were conducted. On the other hand, the study intends to demonstrate that and when distributed learning actually constitutes effective learning behavior. Because only if this is the case it would make sense to investigate whether an increase in distributed learning can be a mediator between non-cognitive factors and learning success. In this sense, Study 3 directly prepares the work on the main research question of this thesis.

4 Study 3: Spacing Can Be More Effective Than Scaling in Video-based Learning: A Learning Analytics Field Study

This chapter originated as a manuscript co-authored with Sebastian Hobert, University of Goettingen. The two authors share first authorship of the manuscript. The manuscript is currently under review at a peer-reviewed journal. It is used and pre-printed in this dissertation with the kind permission of the publisher and the co-author.

Berens, F., & Hobert, S. (Under reviewc). Spacing Can Be More Effective Than Scaling in Inverted Classrooms: A Learning Analytics Field Study.

Abstract: Multiple experimental studies have demonstrated that spaced learning is more effective than the same amount of massed learning. Recent studies also provide information about the most favorable distributions of learning. While on an experimentally controlled level, a good basis of knowledge about spaced learning is given, the investigation of the same effects in the field is much rarer and more difficult. In particular, there is a lack of a multidimensional view of spaced learning that goes beyond the distribution over an entire course. This study addresses this gap by using a field study to describe the distribution of learning through videos in many dimensions and then examining the relationship to learning outcomes. It is found that spaced learning correlates with learning success but not always positively. Some dimensions of spaced learning across the course even correlate more strongly with success than quantity. The investigated course shows that in video-based learning, spaced learning can even be more effective than increasing the quantity of learning. Spacing, then, can be more effective than scaling. In contrast, spacing within a week or a day is shown to have a negative effect. The relationship found between distribution and quantity of learning emphasizes the importance of spacing in receptive learning through video. Teachers and learners should therefore design learning processes in such a way that the spacing of learning is ensured to a sufficient degree of multidimensionality. Further research on spaced learning benefits from the introduced multidimensional view of spaced learning in field studies.

Keywords: Distance education and online learning, Media in education, Mobile learning, Post-secondary education, Teaching/learning strategies

4.1 Introduction

The advantages of spaced learning over massed learning have long been known (Bjork, 1966). Various experiments have not only demonstrated this for different forms of learning (Nakata & Suzuki, 2019; Wegener et al., 2022), but have also investigated a wide variety of rhythms into which learning can be spaced. Knowledge on spaced learning from experimentally controlled settings has therefore been extensive for some time (Cepeda et al., 2006; Yeckehzaare et al., 2022). In contrast, researching spaced learning in real-world learning situations has been less common (Rodriguez et al., 2021). In particular, investigation through non-invasive methods, such as analysis of digital behavioral data through learning analytics, is rare and yields less clear results (Miyamoto et al., 2015; Yeckehzaare et al., 2022). In particular, it is important to note that, unlike experimental designs, in real-world learning situations, the amount of learning does not have to be constant but can be freely chosen to some extent by instructors and learners. This creates a complex learning situation in which stakeholders decide on both the quantity and the spacing of learning. For educational research, this means that both aspects of learning behavior have to be considered simultaneously, and the effects on learning success should be investigated together.

The exploratory field study presented here, therefore, investigates a compulsory introductory course in statistics at a large public German university using learning analytics, which took place at the beginning of the COVID-related restrictions in the spring of 2020. By this time, almost every university worldwide had implemented emergency remote teaching (ERT) as a digital reply to the lockdowns caused by the pandemic (Bozkurt & Sharma, 2020). The fully digitized teaching due to ERT at this point in time makes it possible to see the learning behavior of the students, particularly comprehensively mapped in the digital behavioral data. While students were already free in their work on exercises beforehand and could be observed digitally to some extent, students' work with videos as material for instruction and for presenting content is new. For the first time, students have the opportunity, instead of attending a 90-minute lecture per week, to make free decisions about the distribution of their learning and also to change the quantity of their learning, for example, by watching a video several times. This study addresses this new situation in learning by first describing student behavior and then showing connections to learning success. Specifically, the following two research questions will be answered:

1. How do students space their learning in an online video-based learning setting and make use of the learning flexibility during a time of emergency remote teaching?
2. How is the students' learning behavior during video-based learning related to learning success?

To investigate both research questions, we apply learning analytics methods to digital behavioral data from a large-scale field setting with more than 700 participants. The course was taught exclusively digitally in 2020 during the COVID-19 pandemic. In our field setting, lectures were provided to the students as videos on demand in the lecture's online learning app. The resulting video-based learning setting is comparable to other video-based digital courses in normal times (i.e., non-emergency remote teaching). Due to this, we argue that the results from this study are expected to hold not only for times of emergency remote teaching but also for other fully-online video-based learning settings. Consequently, this study aims to contribute to the body of research on video-based learning. In the remainder of this article, we first briefly outline related research on spaced learning and video-based learning before focusing on measuring time-based learning activities using learning analytics. After outlining our case and research details in more detail, we present and discuss the results. Finally, we present the contributions of our study in the implications section.

4.2 Background

4.2.1 Spaced Learning

It has been known since the 1960s that distributed learning leads to better long-term learning performance than concentrated learning. Bjork (1966) was one of the first in this field to show experimentally that learning which is divided into small chunks and then performed in a regular rhythm, leads to better long-term retention of the learned knowledge. As early as 1970, he also presented in a review article that numerous psychological studies in comparable contexts and with comparable designs came to similar effects and used the term spaced learning to refer to that effect (Bjork, 1970). In the same year, Bjork and Allen (1970) already analyzed not only the favorable scheduling of learning intervals but also the cognitive reasons for better retention in spaced learning. In a meta-analysis of 317 experiments, Cepeda et al. (2006) confirm the effects of spaced learning and the relevance of accurately distributing learning, emphasizing, in particular, the success of spacing learning in as granular a way as possible, but also identifying the duration of retention as an important determinant. Despite this already old body of knowledge about spaced learning, research continues to investigate the optimal timing of learning, the effects of spacing on different types of learning, or the effects of spacing on long-term retention in more detail. For example, in a recent study, Murphy, Bjork, and Bjork (2022) indicate that not only is the distribution of learning across multiple sessions favorable for memory, but even the distribution of learning of a content within a single session leads to better memory. Lyle, Bego, Ralston, and Immekus (2022) give evidence that although spaced learning can lead to more negative outcomes during the learning phase, it leads to significantly better outcomes, especially when retention is supposed to be very long.

Other authors show that the effect of spaced learning is beneficial not only for memorizing knowledge but also for different types of learning. Gluckman, Vlach, and Sandhofer (2014), for example, show that not only recalling facts but also learning to generalize experiences higher retention in a spaced format. Nakata and Suzuki (2019) indicate that spaced learning works for vocabulary learning both for related words and for linguistically very different words. Wegener et al. (2022) find evidence that spaced learning is successful in both oral and written learning situations.

The positive effect of spaced learning has thus been supported by many experiments and for many types of learning. However, its observation in real learning situations is much rarer. Thereby, these real learning situations usually differ in that the quantitative amount of learning does not necessarily have to be kept constant as in an experiment. Rather, in real learning situations, teachers and learners usually have the possibility to change the distribution of learning as well as the quantity of learning in total and per content, resulting in circumstances where real contexts deviate considerably from the experimental setting, making the observation of spaced learning in real settings particularly necessary. One approach to this has been made by (Rodriguez et al., 2021). Based on students' self-reports of their learning strategies and behaviors and linking these surveys with learning success in the course under study, they find no effect of spaced learning on success. The reason for this lack of confirmation of the established principle in the real learning situation could be problems with self-reports. Effects such as satisficing could ensure that existing positive effects of spaced learning are masked. Studies that can observe learning more directly than via self-reports have advantages here.

The digitization of learning processes and the availability of process-produced data from such digital learning environments have created new opportunities in recent years. Through learning analytics, digital behavioral data about student learning can be studied and linked to other forms of data. For example, Miyamoto et al. (2015) find for a MOOC that the number of learner-initiated sessions during the course, while controlling for the quantity of learning, has a positive impact on course success. Yeckehzaare et al. (2022) find a positive relationship between the number of active days with the eBook and course success in their course. Both analyses have in common that they measure spaced learning in a single dimension, looking at fragmentation across the semester. However, the experiments discussed earlier show that spaced learning is a multidimensional concept in which the exact fragmentation and the distances between learning units play a role. The distribution of learning within individual sessions was also shown to be relevant, which cannot be represented in the studies mentioned.

Thus, spaced learning should not only be understood as a fragmentation of learning within the entire semester but should also be considered with regard to the distribution over the different course weeks, but also with regard to the distribution within the weeks and days. Based on these considerations, this study aims to analyze multiple dimensions of spaced learning and examine their relationship to learning success.

4.2.2 Video-Based Learning

Video-based learning has been studied widely even before the pandemic and emergency remote teaching since it became a typical learning format in many emerging learning settings. The widespread use of video-based teaching (e.g., in MOOCs or blended learning) demonstrates that the concept is not novel but has become a de facto standard in online learning. Multiple literature reviews focused on MOOCs (e.g., Sanchez-Gordon & Luján-Mora, 2018; Wong et al., 2019), further studies particularly surveyed video-based learning and indicate the growing importance of video-based learning (Giannakos, 2013; Yousef, Chatti, & Schroeder, 2014).

With a focus on video-based education, many different studies surveyed this topic. For instance, the effects of different strategies on how to design videos have been analyzed. For instance, Pi, Hong, and Yang (2017) revealed the effects of the size of an included instructor’s image into a video, Cojean and Jamet (2018) researched the inclusion of a table of contents and timeline markers, and Rainer Winkler, Hobert, Salovaara, Söllner, and Leimeister (2020) tested the effects of combining video-based teaching with a scaffolding-based conversational agent. C.-M. Chen and Wu (2015) analyzed the effects of different video types. They found, for instance, that videos implementing picture-in-picture recordings of the speaker result in better learning performance compared to voice-only recordings.

In flipped classroom settings, the effects on learning success compared to traditional settings have been studied. For instance, G. S. Mason, Shuman, and Cook (2013) compared the effectiveness using multiple approaches and found that the students’ performance was equally good or even improved. Ojennus (2016) also analyzed learning gains in a flipped classroom setting and revealed that learning gains could not be significantly identified but the satisfaction improved. Peterson (2016) showed that the flipped classroom model is superior in terms of learning outcomes and satisfaction. Summarizing existing studies on video-based teaching in flipped classroom settings, it can be concluded that the topic has been studied in many cases – we only showed a subset of them in this related research. A more complete overview on video-based learning is shown in existing literature review studies (see above for examples). Even though not all studies reveal that video-based teaching in flipped classroom settings is superior compared to traditional settings, it can be concluded that positive effects on the overall performance or on the students’ perception are to be expected in many settings. These results inform this study that video-based teaching formats are in principle capable of providing suited learning settings and are often capable of fostering learning success.

4.2.3 Measuring Time-Based Learning Activities

In prior research, different types of summative measures have been used to capture learning. One of the simplest measures is to count the number of learning interactions during a specific timeframe like the number of clicks within an e-learning application (e.g., used in Andergassen, Mödritscher, & Neumann, 2014 and Mödritscher, Andergassen, & Neumann, 2013) or the number of learning sessions (e.g., Macfadyen & Dawson, 2010; Mödritscher et al., 2013 or Miyamoto et al., 2015). Transferred to video-based learning, counting the number of watched videos could be used as an appropriate measurement for the overall count of learning interactions.

One common approach to measure the students overall learning activity on a timely basis is determining how often they learned during specific time intervals. To this aim, for instance, the number of days each student used a learning system can be calculated (like in Andergassen et al., 2014; Mödritscher et al., 2013 or DeBoer, Ho, Stump, & Breslow, 2014). Similarly, other time units like the number of weeks (like DeBoer et al., 2014) in which a student learned are also possible. Also, determining the exact time spent in a learning application could be used (like in Miyamoto et al., 2015). In video-based settings, this seems, however, to be difficult to calculate, as students often do not interact with the application while watching a video.

In prior research, further time-based measures have been used to identify whether students started to learn early during the semester or only shortly before the exam. For instance, time indexes can be used to describe whether students started to learn early during the lecture term or only shortly before the exam (Riel, Lawless, & Brown, 2018). Other studies calculated the range between the first interaction and the last interaction of a student (like Andergassen et al., 2014). This gives an indicator whether students learned during the whole lecture term or just in a short part of it.

Finally, to measure the success of learning activities, a score representing the students' success is needed. For this purpose, often, the score or final grade of final examinations or formative assessments or certifications are used (see, e.g., Andergassen et al., 2014; DeBoer et al., 2014; Macfadyen & Dawson, 2010; Miyamoto et al., 2015). Even if final grades or examination scores are often used, they seem not to be an optimal solution as it can be questioned if final grades actually represent the students' learning success. Nevertheless, using grades or examination scores is usually the best measure available for learning success, and it, therefore, is an established measure in learning analytics.

4.3 Methods

4.3.1 Field Setting

The study's field setting was an introductory statistics course at a large German public university, which addresses undergraduate students majoring in a social science field. Before the COVID-19 situation, the instructor gave a voluntary in-class lecture of 90 minutes every week during the lecture period, followed by one week to prepare for the final examination. As further voluntary teaching offers, students could attend a tutorial session per week and use a provided e-learning app for accessing the course materials. Besides the exam, mandatory teaching offerings are not allowed for courses in these programs. As the COVID-19 situation resulted in a general lockdown in Germany (Merkel, 2020; Wikipedia Contributors, 2020), the university decided to move all teaching to digital learning environments – only examinations could take place in person at the end of the semester. To prevent technical problems and give students more flexibility in learning, the university recommended offering asynchronous teaching (e.g., uploading lecture recordings) instead of synchronous teaching (e.g., video conferencing) for large-scale courses. Due to this recommendation, lecture recordings were post-processed and cut into short micro content videos with a typical length of 10-20 minutes to make the videos easier accessible and useable. The lecture recordings were uploaded the night to Tuesday in the e-learning app on a weekly basis following the lecture's typical schedule during the lecture period from April to July 2020. The students could watch the uploaded videos via the e-learning app at any time until the examination date.

4.3.2 Data Collection

The data needed for this study was collected from the e-learning app used in the course, which provides students access to all relevant materials and integrates a web-based video player. Overall, N1 = 640 students (out of 723 students registered in the course) entered the video player at least once. They gave informed consent and agreed to participate in the research activities. For these students, interaction with the video player was captured and stored in a pseudonymized form. The log data encompasses (1) control interaction events like starting, pausing, seeking, or continuing, (2) configuration events like changing the playback speed or volume, and (3) a timestamp. After an embargo period, examination scores of the N2 = 266 students who took part in the exam were matched. All other students had dropped out of the course during the lecture period or opted to take the exam at a later point in time. The number of these cases does not differ from times before ERT.

4.3.3 Operationalization

4.3.3.1 Aggregation of learning activities per time

To enable the analysis of the first research question, we aggregated data that describes the students' overall learning activities. To this aim, we did not aggregate the students' learning behavior on a per-student basis but calculated it over all students per timeframe. Thus, the data describes the summative video-related activities of all students within the field study.

First, the number of videos played per week is calculated. This variable describes the number of videos played for each of the 15 weeks calculated by summing up all the events where students pressed the play button. As students may repeatedly watch videos, the total number of videos played per week may exceed the number of available videos. Second, the number of videos played per day of the week describes how many videos were initialized and started for each day of the week (Monday to Sunday). Third, the number of videos played per hour describes how many videos were started for each hour of the day. Finally, we calculated how many videos were started for each week of the semester and for each hour of the day. These data provide a more detailed insight as the aggregation timeframes are reduced.

4.3.3.2 Aggregation of configuration actions per time

Similarly, we analyzed the configuration actions per time. Those configuration actions are specific for video-based learning, as typical video players enable specific configurations that can be changed during the playback (like playback speed or volume changes).

We focused on the rate change events on a timely basis that describe when students increase or decrease the playback speed. To this end, we calculated the sum of all rate changes per week, which describes how often students changed the playback speed. As the number of rate changes per week is dependent on the number of videos watched, we also derived the number of rate changes per week and videos watched.

As an examination of changing the volume did not reveal any interesting insights at all, we do not consider these events any further. One reason for this might be that the e-learning app can only gather data from the video player. If students adjust the volume using the operating system's built-in volume setting, this cannot be captured.

4.3.3.3 Aggregation of learning activities and configuration actions per student

To analyze the relationship between the students' different learning strategies and learning success, we also applied a per-student aggregation. In this part, we focus on the subset of students that took the exam directly after the lecture period (N_2 = 266).

First, for each student of the subset, the score achieved in the examination is calculated as the percentage of points earned. The limit for passing the exam was 0.4, which is identical to previous semesters in the course.

For each student, the total number of videos and the total number of distinct videos the students started are determined. Whereas the first variable contains all videos a student watched, the second variable excludes repetitions of videos. To also capture the time a student learned by watching videos, the total time of all watched videos is determined for each student. These three variables describe the overall engagement in video-based learning for each student.

Additionally, time-related variables are calculated that describe the distribution of the learning activities across a day, a week, or the semester: To measure if students watched videos regularly, the time leak is calculated as the difference between the upload of a video and the first play for each video and student. The resulting total time leak of the first play is the sum of all time leaks and includes all videos a student watched. The greater the time leak, the later videos were watched compared to their upload. The number of days resp. weeks counts on how many days resp. weeks a student watched at least one video. The timeframe in which videos have been watched is the difference in days between the first and last day on which a student watched a video. The first day of watching videos in the semester is also captured.

Finally, we calculated variables that measure the students' regularity of learning and the average time of the day when they watched videos. The first of both calculated the amount of videos played on the most intense weekday. This item is 1.0 if a student watches videos only on one specific day of the week, i.e., the more considerable the amount, the more regularly the learning. If the student distributes the watching of videos equally among all weekdays, the value would be 1/7 (approx. 0.143). To also measure if students watched videos during typical times during the day, we calculated the average absolute time difference of all watched videos from noon and normalized it. This variable would be 0 if a student watched all videos precisely at noon and 1.0 if all watches occurred at midnight. Small values indicate that videos were watched in the daytime; large values indicate that videos were watched at night.

4.3.4 Analysis

The examination of activities and configuration actions is done based on the items described above. For analyzing the timely data, timelines are plotted and visualized as a heatmap.

To analyze correlations between the students' learning and configuration actions and learning success, we used Spearman correlation coefficients using standardized values. To examine the score's explainability based on the gathered video-based data, a multiple rank regression was used to account for the large skewness and frequent occurrence of outliers in count variables (T. Chen, Tang, Lu, & Tu, 2014). Similar to Spearman's correlation or rank sum tests, in rank regressions not the realizations of the variables themselves are included in the regression, but their ranks. For example, a regression weight of 0.2 can thus be interpreted as "If a case rises by five ranks in the independent variable, an increase by one rank in the dependent variable is to be expected on average". Moreover, in rank regressions, the coefficients are equal to their standardized coefficients beta. We calculated the variance inflation factors to check for multicollinearity and removed two factors (number of days and number of weeks) to reduce multicollinearity. The result is a model with variance inflation factors that are all less than 3.5. Calculating this model seems acceptable, even though the interpretation should be made carefully.

4.4 Results

4.4.1 Learning Activity During the Semester

The students watched a total of 55,216 videos within the timeframe of 15 weeks. Per week, approx. 3,680 videos were watched. As visualized in Figure 19 (a), two peaks can be identified: (1) At the beginning of the semester, the number of watched videos reached a local maximum of approx. 5,600 videos watched in week 3. In the subsequent weeks of the lecture period until week 13, the number of watched videos decreased to approx. 3,100 on average. Only at the end of the semester, during the period of examination preparation, the number of watched videos increased instantly and reached a global maximum of approx. 6,850. The chronological development shows that students engaged in video-based learning to an increased extent at the beginning and at the end of the lecture period. This is in line with the experiences of instructors in this course in previous semesters, for example, regarding lecture or tutorial participation.

A closer examination of the usage times of the video player is given in Figure 19 (b) and (c). The plots indicate that students watched videos mainly on Tuesdays, which was the release date of the weekly lecturer recordings. On Tuesdays, approx. 26% of all video watches were observed. On the remaining days of the week, the percentage varies between approx. 9 % to 18 %. This indicates that students used the offered flexibility to watch videos on different days, while at the same time, there is a slight tendency towards the publication day for which the lecture was also scheduled.

The analysis of the times of day in which the students watched videos reveals that starting from 8 am, the number of videos watched exceeds the limit of 1,000 videos watched per hour. This level is held until 10 pm. Only from 11 pm to 7 am, the number decreases below this limit. During the typical lecture times at the university, between 8 am and 6 pm, approx. 83 % of the total number of videos watched occurred. This also means that students take the freedom to study outside of typical lecture times (approx. 17 % of the video watches).

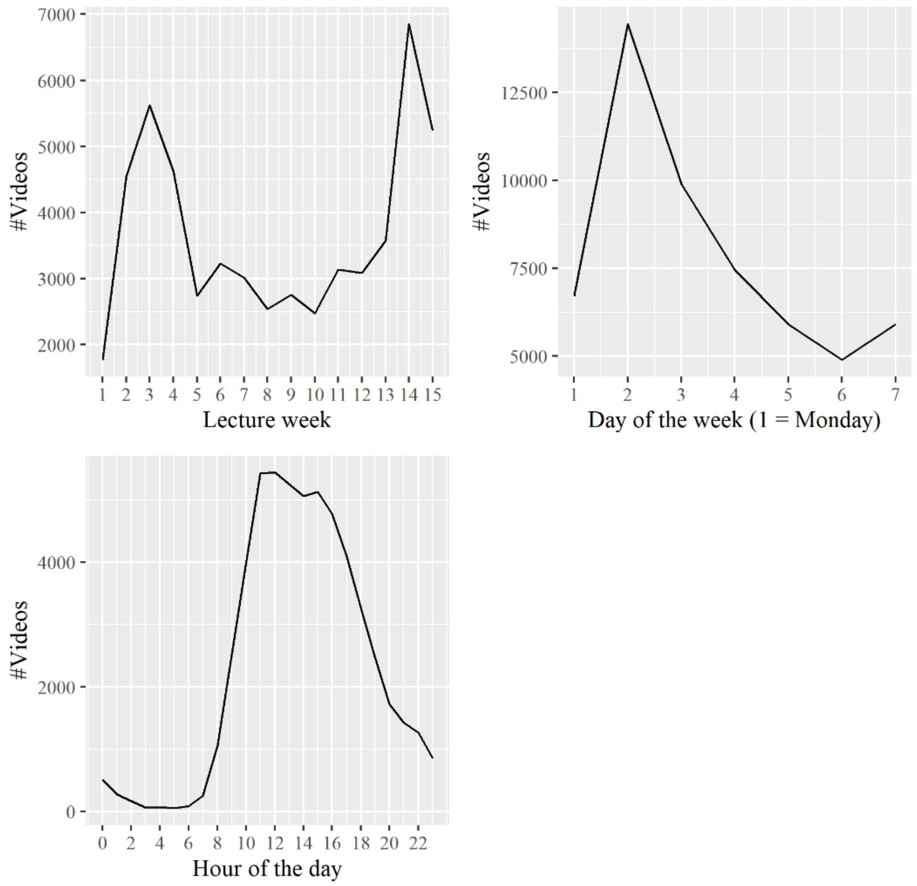

Figure 19: Dynamic of video watching aggregated per week (a), weekday (b), time of day (c)

A closer analysis of the time and day distribution reveals that the number of watched videos' maximum value can be observed on Tuesdays around 4 pm, as visualized in Figure 20. This is not that surprising, as this was roughly the time scheduled for the lecture if it would have been given in the lecture hall. It can be assumed that some of the students used the old semester planning as a structuring aid for self-organization and therefore watched the videos at the previously scheduled lecture time. Some students this way seem to have built up a simple strategy against the destructuring of their daily learning routine. The heatmap further shows that other students, in contrast, watched videos all day. Only during the night, the number decreases.

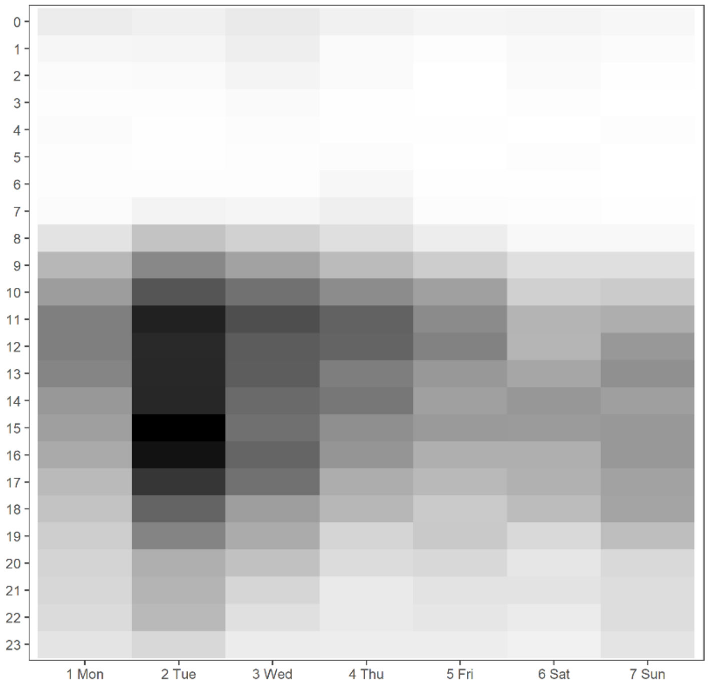

Figure 20: Heatmap of aggregated distribution of learning activities (weekdays and hours). Darker colors indicate higher activity.

4.4.2 Configuration Action

As outlined in the operationalization section, we focused our analysis of configuration events on rate changes of the playback speed. More than 90 % of these rate changes were an increase in the playback speed. Figure 21 (a) shows that the number of rate changes increases over time. Whereas in the first week, the number of rate changes was relatively small, it increased dramatically shortly before the examination.

These temporal insights are particularly interesting when matching them with the number of watched videos. As shown in Figure 21 (b), at the beginning of the semester, the rate changes per watched video were lower than 20 %. With the start of week 5, it increased to more than 25 % and finally reached approx. 65 % shortly before the exam. This shows that students watched videos during the lecture period mostly at a regular playback speed and increased the playback speed in more than 50 % of the cases before the examination date.

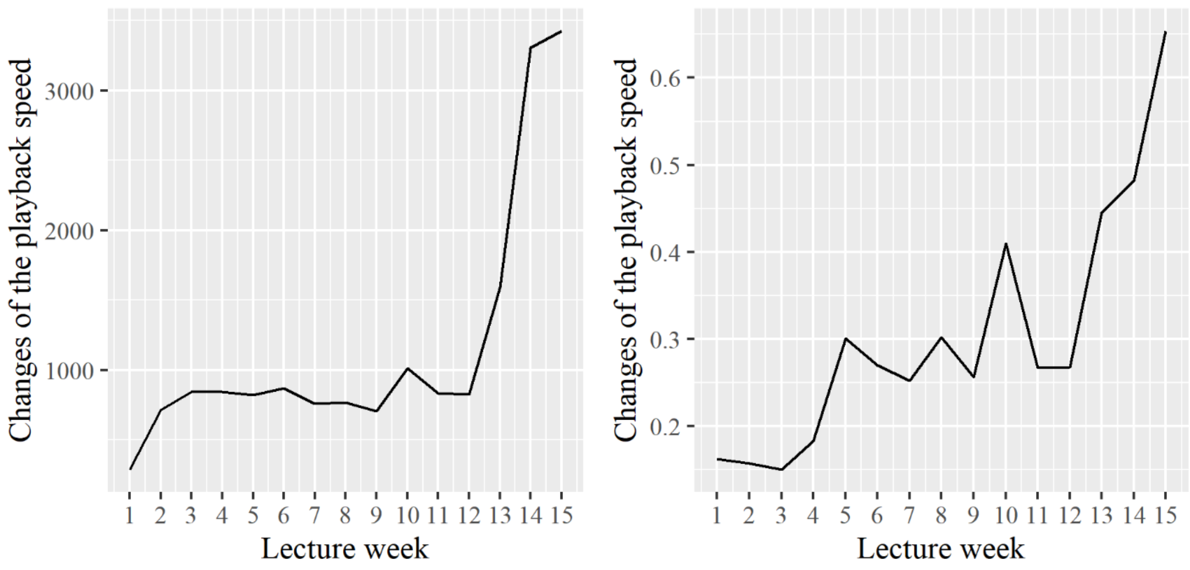

Figure 21: Rate changing configuration actions (a) total (b) total divided by number of videos watched that week.

4.4.3 Relationship Between Watching Videos and Learning Success

To determine whether the observed learning behaviors described above were successful, analyzing the relationship to learning success is needed. To this aim, we examined the associations between learning behavior and learning success for N2 = 266 students who watched videos and took the exam using Spearman's correlations. Out of them, 223 passed the exam successfully, 43 failed. The average score of the students was approx. 61 %. The correlations displayed in Table 4 indicate that the number of watched videos correlates with the examination success only slightly. It does not matter too much whether the total number of watched videos or the total time of watched videos is analyzed. Just the distinct number of videos performs slightly better. Higher correlations exist between most variables focusing on the spacing of learning. Interesting is that the number of days on which videos have been watched has a lower correlation with the examination score than the number of weeks. This may indicate that learning over a longer total period is relevant. The timeframe in which videos have been watched, however, performs worse as an indicator. Starting to learn early during the semester (first day of watching videos; -0.252) and watching videos regularly shortly after being uploaded (total time leak of the first play; -0.451) correlates medium to strongly with the examination success. Due to the operationalization, the correlations have a negative value. Thus, time leak is the most strongly correlated among the variables under consideration. The number of videos played on the weekday, which has the highest plays for the individual student, has a medium correlation with 0.304. This indicates that regular learning on a specific day of the week seems to be positive. Finally, the average of the absolute time difference from noon has a correlation of -0.259. This may indicate that learning during the day is better in terms of learning success compared to learning at night.

Table 4: Spearman correlations between students' video watching behavior indicators and the exam score

No. videos	No. videos distinct	Total time of all watched videos	No. days	No. weeks	Time-frame in watching videos have been watched	First day of watching videos	Total time leak of the first play	Amount of videos played on highest weekday	Average absolute time difference from noon
0.139	0.233	0.144	0.241	0.360	0.065	-0.252	-0.451	0.304	-0.259

The rank regression model in Table 5 shows that the total number of distinct watched videos, the time leak, and the average time difference from noon have significant effects on examination success. With a regression coefficient of -0.361, the time leak has the most important effect according to the computed model, followed by the number of distinct videos and the time difference from noon. This indicates that regular learning soon after the videos have been uploaded is superior compared to watching videos only shortly before the examination. Also, the time of the day on which videos have been watched and the total amount is important – even though with slightly less impact.

Table 5: Multiple rank regression explaining the exam score by students' video watching behavior indicators (R^2 = 0.266)

Indicator	Coeff.	Std.Err.	t-value	p-value
No. videos	-0.080	0.097	-0.825	0.410
No. videos distinct	0.231	0.082	2.813	0.005**
Total time of all watched videos	-0.036	0.082	-0.144	0.662
Total time leak of the first play	-0.361	0.069	-5.248	0.000***
First day of watching videos	-0.034	0.092	-0.372	0.710
Timeframe in which videos have been watched	-0.043	0.084	-0.515	0.607
Amount of videos played on highest weekday	0.082	0.089	0.926	0.355
Average absolute time difference from noon	-0.137	0.056	-2.433	0.016*
Intercept	183.957	23.833	7.719	0.000***

4.5 Discussion of the Results

4.5.1 Learning Activity and Configuration Actions in the Video-Based Online Learning Setting

Even though teaching and learning at universities changed dramatically during the time of emergency remote teaching (J. Crawford et al., 2020; Litao Sun, Tang, & Zuo, 2020; Weeden & Cornwell, 2020), the students' learning activity aggregated over the entire lecture period until the examination did not change dramatically according to our analyses and the teaching experience from previous lecturers. Our field study observed the typical change in the students' engagement: At the beginning of the semester, the students started motivated and took many learning activities. As the semester progresses, the learning activity decreases and starts increasing shortly before the examination, where it reaches its maximum. Even if this does not seem to be a remarkable finding at first, it is nevertheless noteworthy that this learning behavior did not change in the online video-based learning settings compared to typical in-class settings.

Nevertheless, some changes in the students' learning behavior can be recorded in our data from the field. Whereas in prior semesters, the time of a lecture was predefined, the students now gained more flexibility to watch the lecture recordings. As shown in the aggregation by day of the week, the students used the flexibility and watched videos on all weekdays. However, the peak on Tuesdays (the scheduled day of the lectures) also dominates the video retrievals. It seems that a part of the students followed the typical schedule. The same pattern can be observed at the time of video retrieval. Overall, the students used the gained flexibility and watched videos until late in the evening or early in the morning. Nevertheless, the most activity can be observed during the typical times of the lecture – on Tuesday afternoon.

In addition, analyzing the configuration actions while watching videos provide us insights into how students learn when they are watching a video. The analysis indicates that students are aware of the possibility of increasing the playback speed of videos. This does not seem surprising as video players from commercial video-sharing platforms also enable users to set this setting. However, the interesting result is that only a small number of videos was watched with an adjusted playback speed during the whole semester. Only at the end of the semester, this changes massively: Students adjusted the playback speed more than 0.6 times per video. This indicates that students either started too late to watch the videos in time using regular speed and now have only time to watch them faster or that they watch the videos a second time using a faster playback speed to repeat the learning contents. Whereas the first aspect seems a disadvantage, the second aspect seems positive. Principally, watching videos with a faster playback speed seems to be beneficial, according to a recent study by D. Lang, Chen, Mirzaei, and Paepcke (2020).

Based on the insights into the students' learning activities, we can answer the first research question:

RQ1: How do students space their learning in an online video-based learning setting and make use of the learning flexibility during a time of emergency remote teaching?

The analysis of the learning activities during the course as well as while watching videos indicates that students use the gained flexibility of the fully-online video-based learning setting. First, a huge number of students applied different forms of spaced learning. Some students followed the suggested timeline of watching the lecture recordings shortly after they were released, while others conducted binge-watching. Additionally, the students used the gained flexibility to adjust their individual learning process to their own needs. This is reflected in the configuration action data. The data indicate that students are aware of how to adjust the video playback and use it particularly often the closer it gets to the exam.

4.5.2 Relationship Between Watching Videos and Learning Success

The analyses of the relationship to the students' individual learning success indicate that the total amount of watched videos positively correlates with learning success. This finding is, obviously, not new. Interesting is, however, that this positive correlation is not very strong and much smaller than correlations with factors that summarize how students space their learning across the course.

According to the analysis, the correlations that describe spaced learning across the course are more critical. The number of days and weeks on which videos have been watched has stronger correlations than the total amount. Interesting is that students that started to learn early during the semester seem to perform better. Learning regularly described by the time leak of the first play also indicates that this strategy is superior. The strategy to learn only shortly before the examination is worse compared to regular learning.

To further analyze the effects of giving students the flexibility to learn whenever they want, we calculated the average absolute time difference from noon. This indicates (a lack of) spacing within a day. The correlation shows that this indicator has one of the strongest correlations in our analysis. This is counter to the hypothesis that spacing is always positive as it indicates that massed learning behavior towards the mid of the day is associated with more positive outcomes than spacing equally across the whole day (and night). Similar effects can be observed on the dimension of spacing across the week. Focusing the learning on one specific day of the week seems to be associated with a more positive outcome compared to spacing the learning activities equally across all days of the week. As our analysis of the learning activity over time indicates that students like to use the flexibility to watch videos whenever they want, this could be seen as problematic. Learning late in the evening or early in the morning seems to be a worse strategy than learning during the day. The rank regression model also confirms that there exist significant effects of the amount of learning, but also of the time leak and the time difference from noon.

Using these insights into the students learning behavior and learning outcome, we can answer our second research question:

> *RQ2: How is the students' learning behavior during video-based learning related to learning success?*

Our analysis of the effects of the students' learning behavior on the success in the examination indicates that using the given flexibility to learn and space learning whenever the students want might be a possible strategy for students that is yet not necessarily successful. A multidimensional perspective is needed. Even in times when students are completely free to define the amount of learning and the time and place of learning on their own, it seems advisable to follow a regular, spaced, and planned learning process. While indicators for spacing across the course indicate positive effects on learning success, the indicators for spacing across the week or across the day seem to have negative effects. However, it should also be noted that the data used do not necessarily reveal causalities. The correlations found do not necessarily indicate direct effects of the behavioral variable on learning success. For example, many of the variables may be influenced by self-organization skills, which in turn also affect success. Nevertheless, to draw conclusions about learning behavior and the likelihood of success, the correlations are still an interesting insight.

At first, these results seem to be contradicting to related research on flipped classroom settings. In related studies, it has been shown that transforming a course from in-class teaching to a digitally supported (video-based) flipped classroom results in equal or even superior results (G. S. Mason et al., 2013; Peterson, 2016). In our setting, a similar digital transformation from in-class teaching to video-based teaching has been conducted. However, in typical flipped classroom settings, usually, some in-class sessions exist that provide structure for the students' learning processes. In the COVID-19-induced emergency remote teaching, this structure was missing as students had all flexibility to learn at any given time. They do not need to attend specific in-class sessions. This might be a hint for designing future video-based learning settings during or after the COVID-19 situation: Provide video-based lectures on demand to give students the flexibility to learn at any given time, but combine it at the same time with an additional structure that motivates students to learn regularly.

4.6 Conclusion and Implications for Future Video-Based Online Learning Settings

Overall, it appears that the effects of spaced learning require a multidimensional perspective. Spacing the learning activities across the whole course seems to be very effective. Even more than scaling the quantity of learning. Our analysis of actual field data indicates that spacing the learning activities within the days of a week or within the day (and night) seems, however, not beneficial. These findings seem to be particularly noteworthy as we could observe in our field study that students make use of flexible learning to a large extent. Thus, they run the risk of being affected by the negative effects of spacing across the week and the day.

The implications of this study on spaced learning in online video-based learning settings should be distinguished by the two actors: From the lecturers' perspective, this study gives insights into how to structure asynchronous video-based teaching and advises students on how to use the offered flexibility. In contrast to non-digital learning environments, lecturers often get the unique possibility to actually measure the students' learning engagement. The typical pattern that learning often takes place intensively before the examination seems to be true for digital learning as well. In terms of giving students the possibility to learn independently of a specific schedule or time, lecturers should reflect if offering this flexibility is advisable, as our results indicate adverse effects. Instead of restricting this flexibility, we argue that it seems advisable to instruct students on how to use flexibility and spacing to learn. Limiting this flexibility by using synchronous online teaching offers seems not advisable to us since our data show a demand for flexibility in learning.

From a student's perspective, our results can be seen as a guideline on how to space learning processes. Particularly noteworthy is that we could highlight that regular learning seems to be a favorable strategy compared to learning only shortly before the exam.

Nevertheless, it should be noted that not all spacing approaches (esp. during the week and day) are helpful, but following a structured learning process and planning the learning sessions well ahead seems beneficial.

The results of this research paper should be reflected in the discussion of the limitations. This field study results are observed during the COVID-19 pandemic, which is characterized by a rapid digital transformation and emergency remote teaching within the university context. Additionally, it needs to be noted that this learning analytics-based approach was only able to record learning activities that took place within the e-learning app provided to the students. This does not seem to be critical, according to our assessment, as the app could be used as the only resource for the course. Nevertheless, it cannot be ensured that the students used other learning materials and that this influences the results. This limitation is induced by the method of a field study design.

Despite these limitations of the field study design, the results seem to be interesting for future work. We could outline insights into the effects of giving students more flexibility to learn, which enabled them to apply spaced learning in a more extended way. This may not only be interesting for situations similar to the COVID-19 pandemic but also for other video-based learning settings. Verifying whether the results related to spaced learning are also valid for other (online-based) learning settings seems to be a worthwhile starting point for further research.

4.7 Concluding Remarks and Transition to Study 4

Although the results discussed in Study 3 provide interesting findings for learning research and have implications for learning, they do not answer any of the research questions of this thesis. Nevertheless, the results of Study 3 are important for the continuation of this thesis. Firstly, the study was able to show that digital behavioral data can be extracted from the learning platform developed in Study 2. This data can be operationalized in a meaningful way to describe relevant dimensions of learning behavior. Secondly, study 3 shows that distributed learning and also the amount of learning, at least when learning with videos, are actually associated with learning success. Only this finding makes it meaningful to address the main research question in the intended sense.

This is due to the fact that the main research question aims to investigate whether the association between non-cognitive factors and learning success can possibly be explained by mediation through variables of learning behavior. However, such an argument only makes sense if the variables of learning behavior under consideration are themselves related to learning success. Study 3 can show here that this is the case for almost all operationalized variables about learning with videos. Study 3 thus confirms this assumption and legitimizes the continuation with the main research question.

The work directly tackling the main research question is now possible after all three important prerequisites have been established: the development of the missing measurement instrument on beliefs about statistics in Study 1, the development of the software in which the learning behavior to be observed takes place in Study 2 and the empirical justification of the assumption that there is a connection between the learning behavior that can be described via the digital behavioral data and learning success.

The main research question will be addressed in two steps. Ideally, all three levels of non-cognitive factors could be examined together in a model in terms of their connections to each other and to learning behavior and learning success. While the links between attitudes towards statistics and statistics anxiety have already been investigated quite well, this does not apply to the links between beliefs about statistics and the other two levels of non-cognitive factors, as has also been shown in the development of research question 2b of this thesis. Bringing all three levels of non-cognitive factors into one model for research question 3 would mean addressing research questions 2a, 2b, 3a, 3b and 3c all in one step. This appears to be too much for a single study and would also be difficult to implement with the expected number of cases available in the context of the study.

Since in line with research question 2a it is still unclear whether beliefs about statistics are related to learning success at all, priority is given to this important basic prerequisite for the main research question. This research question is, however, directly combined with answering the main research question for beliefs about statistics. Study 4 thus addresses research questions 2a and 3a of this thesis in one step. The main research question for attitudes towards statistics and statistics anxiety is then dealt with separately in Study 5.

5 Study 4: Do Students´ Disciplinary Beliefs about Statistics Shape Their Learning Behavior? A Learning Analytics Approach

This chapter originated as a manuscript co-authored with Sebastian Hobert, University of Goettingen and Kelly Findley, University of Illinois at Urbana-Champaign. The author of this dissertation is first author of the manuscript. The manuscript is currently under review at a peer-reviewed journal. It is used and pre-printed in this dissertation with the kind permission of the publisher and the co-authors.

Berens, F., Findley, K., & Hobert, S. (Under review). Do Students´ Disciplinary Beliefs about Statistics Shape Their Learning Behavior? A Learning Analytics Approach.

Abstract: In mathematics and science education research, it has long been known that beliefs about and conceptions of the domain are related to learning success. Little is known, however, about why such a relationship between beliefs and learning success can be found. An influence of beliefs on learning behavior can be assumed, but is difficult to prove. This study tests for an introductory statistics course at a university whether a relationship can be found between conceptions of statistics and learning success. Furthermore, building on digital behavioral data from the learning management system used in the course, the study examines the extent to which influences of beliefs on learning behavior can be found that might explain differences in learning success. It is shown that only small correlations between beliefs and success can be identified and that the learning behavior variables examined are not suitable as an explanation.

Keywords: Disciplinary conceptions, Learning Analytics, Learning Behavior, Operationalization of Learning

5.1 Introduction

Researchers in mathematics and science education have documented the significance of students’ beliefs in several aspects of the learning process. Disciplinary beliefs play a critical role in students’ motivation for learning (Kloosterman, 2002), how students view disciplinary advancement and knowledge construction (Muis, 2004; Tsai et al., 2011), and how students think about problem solving (P. Bell & Linn, 2002; Ozturk & Guven, 2016) to name a few. Even with the same curriculum and the same teacher, students see tasks quite differently depending on the disciplinary beliefs and conceptions they bring to these contexts (Presmeg, 2002). Köller (2001) and Geisler and Rolka (2021) even found evidence that students´ beliefs about a discipline have direct influence on their learning success.

Beliefs, however, are also difficult to study for many reasons. For example, a review of the literature on mathematical beliefs finds the construct to be inconsistently defined and operationalized in empirical work (Liljedahl, Oesterle, & Bernèche, 2012). This finding comes as little surprise to researchers in this field of study—beliefs are difficult to define and even more difficult to measure reliably. As Schoenfeld (2015) explains, what people do communicates more than what they say, and statements of beliefs do not always predict actions. Schoenfeld adds that beliefs are contextually-sensitive and take various grain-sizes. Ultimately, high-quality work with beliefs has to operate with specificity in construct, clarity in goal, and respect for the mystery that pervades research work with beliefs.

Research on beliefs in statistics education encompasses a sparser terrain than that of mathematics or science education and falls victim to the same pitfalls. Still, there are shared findings and frameworks that can be built on to better understand how domain-specific beliefs in statistics might be measured and understood (e.g., Gordon, 2004; Justice et al., 2020; Reid & Petocz, 2002; Rolka & Bulmer, 2005). Many of these studies define different ways to frame the conceptions or perspectives that introductory-level students might have of statistics. These frameworks also offer glimpses and theories to how different belief profiles may interact with statistical learning, but little empirical evidence has been provided to truly understand and validate these relationships.

This paper examines to what extent students' domain-specific beliefs in statistics might explain their learning behavior using learning analytics. To measure students' beliefs, we used a new instrument with links to several of the statistical conceptions frameworks cited previously. Using a large sample of survey data from an introductory-level statistics course, we looked for relationships between different belief profiles and various types of learning behavior and learning outcomes.

5.2 Background

5.2.1 Beliefs in Educational Research

Beliefs are a difficult construct to grasp, both in the everyday understanding of the term and in (educational) research. An example of this is the distinction between beliefs that are deeply anchored in the person, for example, about oneself (e.g., I believe that I can perform this task), and beliefs that can rather be described as assumptions, which are to be understood as statements under uncertainty (e.g., I believe that it will rain).

In research, too, there are different conceptions of the term beliefs, some of which are similar or complementary, and some of which are contradictory. For example, a distinction must be made between a very broad definition of beliefs and narrower conceptions. The former definitions subsume (almost) all affect-related personal dispositions into beliefs and then distinguish between attitudinal beliefs, epistemic beliefs, domain-descriptive beliefs, and some others (e.g., Pajares, 1992). Narrower definitions of beliefs draw stricter boundaries and include only those dispositions as beliefs that are very close to the cognitive level. Philipp (2007), for example, divides the world of affective dispositions into beliefs, attitudes, and emotions. McLeod (1992) provides the same distinction, describing beliefs as the most stable disposition over time. The main reason for this is that beliefs have a higher cognitive component than attitudes and emotions, whereas the affective component is smaller than for attitudes and emotions. Hence, these attempts have in common that beliefs are understood as a personal disposition with an affective component, which, however, can be distinguished from other dispositions by its comparatively low affective component.

Concrete attempts of a definition emphasize the strong cognitive component of beliefs. Bar-Tal (1990) goes so far as to define all knowledge, all subjectively perceived facts, all opinions and also faith convictions as forms of beliefs. V. Richardson (1996) goes in a similar direction and understands beliefs as a collection of personally held understandings that are believed to be true by the person himself. In contrast to Bar-Tal, however, she explicitly does not refer to objectively shared knowledge, but to the bundle of subjectively held assumptions. Some authors emphasize that while beliefs are dispositions that the individual perceives as cognitive and that at the same time have an affective component, as a third aspect they also have a behavioral component (Rokeach, 1972). From this already old idea of a link between personally held beliefs and personal behavior, more recent conceptions derive an action-guiding function of beliefs. A definition of beliefs frequently used in recent research is, for example, that of Philipp. Philipp (2007) defines beliefs as "lenses that affect one's view of some aspect of the world or as dispositions toward action" (p. 259). Additionally, Philipp defines conceptions as "a general notion or mental structure encompassing beliefs, meanings, concepts, propositions, rules, mental images, and preferences" (p. 259). He thus introduces another term that describes the totality of a complex system of interrelated beliefs.

For this paper, the latter definition is especially helpful, since it allows us to describe not only individual statements about a discipline, in this case statistics. Rather, the study presented here aims to investigate how students' image of statistics, composed of various components, condenses into a consistent conception, which in turn can have an action-guiding influence on students' learning behavior. The present work therefore follows Philipp's definitions while staying close to the ideas of Törner and Grigutsch (1994), who earlier referred to similar ideas as worldviews and thereby had a great influence on the debate on beliefs within mathematics education.

5.2.2 Disciplinary Beliefs in Statistics

While in mathematics education established conceptions about possible beliefs of learners about mathematics have been present for a long time and are also underpinned by corresponding measurement instruments, this has not been the case for statistics so far. Nevertheless, several authors have already presented analyses of students' conceptions about statistics. In particular, the work of Reid and Petocz (2002), Gordon (2004), Rolka and Bulmer (2005), and Justice et al. (2020) should be mentioned.

In these researches, different typifications of students and thus models about the possible conceptions about statistics were developed in qualitative approaches. Comparing the four studies at hand, it can be seen that while all four come to somewhat different conclusions, they do have much in common. For example, all four studies show one to three categories, which are considered by the authors as conceptions of beginners. Statistics therein is often described as something static, in which one operates with numbers, tools and techniques to master a task. On the other side of the spectrum, all four authors describe a category of conception to which they assign a high level of expertise. In this category, statistics is generally described as a much more flexible approach that enables and requires critical thinking within the application areas. In between, the authors each place middle-level categories, which are described somewhat differently. Justice et al. (2020), for example, focus one category on a picture of statistics that is strongly influenced by algorithms and another on the accurate representation of reality by statistics. Gordon (2004), for example, focuses more on the tool character and the selection of appropriate tools in the statistical process.

Overall, this provides a picture, in terms of previous research on conceptions of statistics, in which the existing studies have enough in common to send a common message that can inform further research. For example, the models have in common that they build a gradient from rather static to rather dynamic conceptions of statistics. This is usually associated with a notion that rather static beliefs are associated with novices, while experts hold more dynamic beliefs. From this, also for the present research, the hypothesis can be derived that beliefs that can rather be described as static tend to have a negative impact on learning, while dynamic beliefs carry positive effects.

5.2.3 Conceptions of Statistics Used for This Study

Even though several authors have already presented conceptions of beliefs about statistics, an instrument that makes one of these conceptions measurable is still missing. The authors of this article have therefore collected their own qualitative data in order to develop such a measurement instrument (see chapter 2 of this dissertation). From the collected data, they first developed their own conceptualization based on a grounded theory approach. Four types of conceptions about statistics were found in the data:

The first type of conception of statistics that emerges among students is the rules-based conception. In this type, statistics is viewed as a set of formulas and procedures for data-related problems. Accordingly, doing statistics means selecting the (objectively) correct formula or procedure for a situation, then strictly following it and thereby obtaining the correct answer. Statistics is perceived as mathematics, context plays no role in this conception. A statistical expert is characterized by the fact that he knows a variety of statistical formulas and procedures, selects them correctly for their application and then works mathematically accurately. Special statistical experts are also able to develop new formulas and procedures. The rules-based conception thus shows a very static and context-free picture of statistics, in which creativity, situativity and ambiguity have no place.

The second conception of statistics found is the descriptive conception. Here the statistical calculation is not in the foreground, but the statistical result, which in the descriptive conception is understood as a visualization, table or parameter. These are understood in this conception as means to represent (objective) reality clearly and summarizing. To do statistics in this context means to summarize data in visualizations or parameters and to communicate information via these summaries. Experts in statistics are familiar with as many different forms of representation, visualizations and parameters as possible and have great communicative skills in using these representations in such a way that users understand their significance in the context of the data. The descriptive conception thus already grants statistics a little more dynamism and creativity, since the representation of the data can certainly take place in different ways. However, the objective truth of the data remains unaffected. Context plays a role in the descriptive conception only in so far as the data originate from a context and the truth speaking from the data has sense in the context. The possible forms of representation themselves, however, are not seen as context-dependent.

The third conception of statistics is the confirmatory conception. It understands statistics as an approach to test hypotheses or theoretical models on the basis of data. A claim derived from the context is either supported or refuted by the data. Doing statistics here means identifying a research question that makes sense in context, deriving hypotheses from theory or literature, identifying the appropriate test for proving them, and arriving at a validation or refutation of the hypotheses. Statistical experts are familiar with a large variety of different tests and their correct application scenarios and perform these tests carefully. Similar to the descriptive conception, statistics in a confirmatory conception is not completely static, but it also has limited dynamism. Creativity and ambiguity lie in the selection of the research question and the development of the research hypotheses, but the process of testing them is undynamic and can only follow a proper procedure. Context is the space in which research question, hypotheses and interpretation have meaning, but in the actual statistical analysis context should not play a role in this conception.

The fourth conception of statistics found is the investigative conception. It conceives of statistics as a flexible approach of exploring data, finding patterns in them, and deriving meaningful conclusions about their context. Doing statistics here means a cyclical process that jumps back and forth between theoretical considerations and the exploration of data. In this process, new ideas or questions are continuously generated, tested on the data, and theoretically developed further. The statistical expert is characterized especially by his curiosity and by his experience in asking good questions, exploring data and finding patterns. This expertise is developed mainly through a wide range of experience in a variety of contexts and with a variety of problem-solving strategies. The investigative conception is thus clearly the most dynamic. It connects context, theory and data continuously and sees the whole process depending on context and data. The context in which work is done is thus permanently present and decisive for start, progress and end of the investigation.

After identifying and describing the four types of conceptions about statistics based on a grounded theory, the four types were ordered into a two-dimensional coordinate system based on theoretical considerations. Bailyn (1977) had described the cognitive process behind data analysis as a communicative process between the conceptual plane consisting of the state of research, theories, and prior experience and the empirical plane consisting of the data and their patterns. Based on this idea, we established the dynamic and interactivity of communication between the conceptual and the empirical plane as a dimension to characterize the conceptions of statistics. The rules-based conception is to be placed at the very bottom of this dimension, since it describes statistics as static and the conceptual level plays no role at all in this conception. The opposite pole is the investigative conception. It is dynamic by nature and lives from the constant interaction between theoretical considerations and exploration of the data. Thus, in a certain sense, it represents Bailyn's ideal. The descriptive conception and the confirmatory conception are each found in the middle ground with respect to the dynamic dimension of statistics. In them, communication takes place between the conceptual and the empirical plane, but this can be understood as a brief exchange of information rather than an engaged dialogue.

As a second dimension to characterize the conceptions of statistics, we determined the starting point of communication between the conceptual and the empirical plane. This communication can either be data-driven and thus take its starting point in the empirical plane, or it can be theory-driven and thus originate in the conceptual plane. In this context, the descriptive conception can be considered as prototypical for a data-driven communication. It takes the availability of data as an opportunity to analyze and depict it in order to enable the conceptual level to gain knowledge from this depiction. At the other end of the spectrum is the confirmatory conception. Here, hypotheses or models are developed within the conceptual plane based on literature or theory, which are then passed on to the empirical plane for testing. In this dimension, the investigative conception lies between data-driven and theory-driven approaches, as it enables both and brings them into permanent dialogue. We also place the rules-based conception in the middle, as it does not involve communication at all and thus does not initiate it. A summary of the conceptualization can be found in Figure 22.

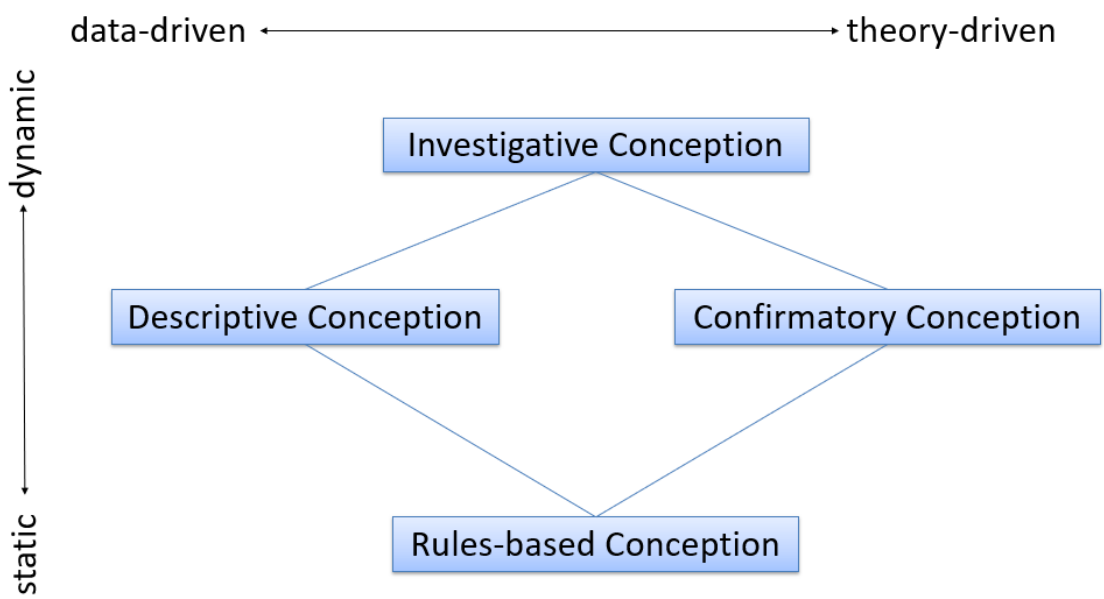

Figure 22: Visualization of the developed theoretical conceptualization about the four types of conceptions about statistics

5.2.4 Research Question and Hypotheses

As illustrated above, we know especially from mathematics education that disciplinary beliefs can make a difference. In particular, dynamic beliefs about mathematics have been found to be associated with higher achievement, while static beliefs seem to have negative effects (Geisler & Rolka, 2021; Köller, 2001). Since this finding also holds for science education, it is reasonable to assume that it also appears in statistics education. The first research question of this article is therefore: Does a relationship between students' conceptions about statistics and their learning success appear? As a hypothesis, it would be reasonable to assume that dynamic conceptions, especially an investigative conception, have a positive impact on learning success, while static conceptions, especially a rules-based conception, have a negative impact.

While there is some research linking beliefs to learning success, little is known about the effects of beliefs on learning behavior. However, if conceptions do indeed have a causal effect on learning success, effects of conceptions on learning behavior would stand to reason as an explanation for this causality. As a second research question, this article therefore asks: Do students' conceptions about statistics shape their learning behavior? It would be expected that successful learning behaviors would be more prevalent among students with dynamic conceptions, while students with static conceptions would be more likely to show less successful learning behaviors.

5.3 Methodology

5.3.1 Study Design

To answer the research question, we use an observational design that combines three types of data. We collect information about the students' beliefs about statistics using a survey. Based on the four types described above, we developed a measurement instrument with 36 items and applied various quality assurance steps (chapter 2 / study 1 of this dissertation). Students rate their agreement with the 36 statements on seven-point Likert items from do not agree at all to fully agree. From these ratings, scores are formed that determine how much the student leans towards the respective type. Based on the theoretical model, the difference between investigative conception and rules-based conception is in addition formed as a score for dynamic beliefs. As a score for data-related beliefs, the difference between descriptive conception and confirmatory conception is formed. This results in a total of six scores that describe the student's beliefs. In addition to these six scores, the survey asked for several other measures, including high school GPA, which is coded from zero (poor) to 10 (perfect).

Learning success is measured via the course's final exam. Success in this exam measured in percentage of achievable points is extracted from registry data of the university. We also use these register data to determine the gender of the students and the information whether the person has already failed the same exam in previous years. Information about learning behavior is extracted from the course's learning management system by operationalizing the emerging digital behavioral data.

Data collection took place in spring/summer 2020 at a large German public university. Prior to the first session of the course, 643 of the registered 721 students participated in the voluntary survey. Out of the 643 students with survey data, 271 students participated in the final exam. This can be seen as a common completion rate for the course, as dropout rates of about 60% were also observed in previous years in this course. Among the 271 students taking the exam, some had already failed the final exam of the course in a previous year. Since the beliefs of these students may already have changed by their prior experience, we exclude these students from our analysis. This leaves 238 students for whom survey data and exam data are available. Digital behavioral data from the learning management systems are available for all 721 registered students, resulting in a full dataset of survey data, digital behavioral data, and exam data for 238 students.

The matching of the various forms of data took place in consultation with the university's data protection officer only after a grace period following the course. The survey data were generally only available with a pseudonym, which was also stored in the digital behavioral data. The linking of the examination data was not done until months after the final evaluation of all examinations and had to be done independently of the instructor and the data analysis.

5.3.2 Educational Setting

The course Statistics for the Social Sciences 1 was used to collect the data. The course is a compulsory course for almost all students of almost all social science subjects at the university. In particular, it is attended by students of the bachelor's programs in political science, sociology, pedagogy, and sports science, with other smaller subjects joining them.

The course runs for 14 weeks, followed by a written exam two weeks later. Each of the 14 course weeks begins with a 90-minute lecture that introduces students to new content.

Since corona-related lockdowns made lectures in the classroom impossible during this study, the lectures were recorded, post-processed, and made available in smaller units of 5-20 minutes as videos on demand via the course's learning management system. Students were encouraged to watch these videos in a timely manner, but had unlimited access until the exam. After being introduced to new content through the lecture videos, students were expected to engage with the content on their own through exercise questions. For this purpose, about 20-30 small questions were provided weekly via the learning management system. The students could answer these questions in a single choice, multiple choice or arithmetic mode and received direct feedback from the learning management system. After this self-regulated learning with the practice problems, the week ends with a 90-minute tutorial in which an experienced student discusses both important lecture content and the practice problems with small groups of students. These tutorials took place as video conferences because of the pandemic and could be accessed through the used learning management system. In addition, students were provided with two quizzes of the day each day, available only that day. After completion of the 14 course weeks, students received offers for revision in repetition sessions and were offered further practice problems for their own training. In total, 401 tasks were available in the learning management system, which could be used until the exam.

The exam took place two weeks after the last tutorial as a written exam of 90 minutes length. A total of 100 points could be achieved. Of these, 45 points were assigned to various computational tasks, 30 points to interpretations of statistical results, and 25 points to questions on conceptual knowledge of statistics. According to the regulations of the university, the written exam is the only compulsory element of the course. In order to successfully complete the course, students must achieve at least 50% of the achievable points in the written exam. All other offers, i.e. lectures, exercises, tutorials, etc., are to be understood as voluntary offers that students can use to prepare for the exam, but do not have to.

The learning management system used in the course offers comprehensive support for the students as an all-in-one solution (see chapter 3 / study 2 of this dissertation). Within the same system they can watch the lecture recordings as well as download the slides, work on the exercises, access the tutorials, participate in audience response questions in the tutorials, answer a quiz of the day, view their personal learning statistics as well as ask subject-related questions to a chatbot. The learning management system used thus provides students with extensive support during the statistics course as well as an extensive treasure trove of digital behavioral data that can be used for this study.

5.3.3 Operationalization of Learning

The learning management system used stores every interaction in its log files, including the student's pseudonym, date and time. For this study, there are more than 6 million log entries of this type available. To answer the research question, these log entries are counted with respect to various characteristics to arrive at interpretable variables of learning behavior. In doing so, we distinguish two main categories of variables to describe learning behavior based on previous research (see chapter 4 / study 3 of this dissertation). First, we generate variables that are suitable to represent the quantity of (successful) learning. Furthermore, we generate variables that try to measure how learning was spaced (over the day/week/course).

To measure the quantity of learning behavior, at first the total number of initiated videos is counted (videos total). The second value counted is how many different videos the student has started (videos total district). The same is done for the practice problems, so that the number of answered questions (questions total) and the number of different questions answered (questions total district) is determined. In order to measure the success or failure in learning, the total number of correctly or incorrectly answered questions (questions total correct resp. questions total wrong) and the number of at least once correctly answered questions (questions total district correct) are counted. Regarding the tutorials, it is counted how many times a student participated (tutorial participation) and how many of those times they also actively participated in the audience response questions in the tutorial (active tutorial participation). Regarding the chatbot function, it is counted how many times a student has queried a definition from the stored glossary (chat definitions total).

To describe the spacing of learning behavior over the day, the week and the semester, five variables were formed, four of which relate to the questions and one to the videos. The latter variable calculates what percentage of all initiations of videos happened during the first 13 weeks of the course (videos during semester proportion). This is intended to assess the extent to which learning with videos took place during the semester, or the extent to which it was carried out in blocked form just before the exam. Based on the same logic, such a variable was also determined for the questions (questions during semester portion). In order to approach the question of how students space their learning over the week, a variable was to be determined that represents regular learning over the course of the week or, on the contrary, blocked learning on selected days of the week. To achieve this, the share of the most learning-intensive day of the week of the total learning volume of the week was determined (most active weekday proportion). In order to describe the spacing of learning within the individual learning day, on the one hand the share of learning in the total learning was calculated, which was mastered between 9:00am and 5:00pm (questions workday proportion). In a similar logic, the average absolute deviation of the learning time from 12:00pm noon was calculated in hours (questions mean difference to noon).

In total, there are thus seven variables that quantify learning behavior, three that describe success or failure, and five that describe learning in its distribution.

5.4 Results

5.4.1 Bivariate Relationships Between Conceptions of Statistics and Learning

To answer the first research question, we examine the extent to which the six measured conceptions about statistics are related to the students' success in the exam. Looking at the bivariate correlations quickly shows that there is little to no relationship to be identified here. Leaning toward an investigative conception correlates only with r=.022 (p=.778), agreeing with the rules-based conception correlates with r=.092 (p=.235). A confirmatory conception correlates with r=.025 (p=.750) and a descriptive conception with r=.001 (p=.992). Dynamically oriented beliefs correlate with r=-.073 (p=348) negatively with exam success, data-driven beliefs correlate likewise negatively with r=-0.031 (p=.692). Overall, therefore, no significant correlation can be found. Only for a tendency towards a rules-based conception and in close connection with this rather static than dynamic beliefs about statistics a minimal correlation can be suspected, while at least all other correlations have to be described as non-existent.

However, the absence of correlations between the measured conceptions and students´ success in the exam does not necessarily imply that no relationships can exist, nor does it necessarily negate all correlations between conceptions and learning behavior. Rather, it is important to examine whether causal relationships can be uncovered in multivariate constellations. Furthermore, the relationships between the conceptions and the learning behavior could reveal bivariate correlations. Table 6 therefore presents all bivariate Pearson correlations between the six scales of conceptions and the total of 15 variables of learning behavior. Since not all variables capturing learning behavior follow a normal distribution we also calculated Spearman correlations. However, there were no substantial differences in the results. We therefore stick to Pearson correlations for further analyses.

Table 6: Pearson correlations between the student´s conceptions of statistics at the beginning of the course and their success in the course exam (p-values in brackets)

	Rules-based conception	Investigative conception	Descriptive conception	Confirmatory conception	Dynamic beliefs	Data-driven beliefs
Videos total	**.052** (.506)	**.004** (.962)	**.003** (.974)	**.080** (.304)	**-.040** (.612)	**-.063** (.419)
Videos total distinct	**.205** (.011)	**.054** (.507)	**.133** (.104)	**.184** (.023)	**-.128** (.116)	**-.054** (.510)
Questions total	**.028** (.724)	**.056** (.474)	**.029** (.710)	**.058** (.457)	**.010** (.896)	**.026** (.741)
Questions total distinct	**.079** (.307)	**.021** (.791)	**.060** (.442)	**.080** (.306)	**-.068** (.382)	**-.006** (.940)
Tutorial participation	**.015** (.842)	**.058** (.461)	**.061** (.437)	**.024** (.756)	**.023** (.764)	**.036** (.643)
Active tutorial participation	**.045** (.560)	**.067** (.391)	**.108** (.166)	**.042** (.591)	**.006** (.938)	**.066** (.398)
Chat definitions total	**.017** (.828)	**.059** (.448)	**-.019** (.801)	**.050** (.522)	**.036** (.645)	**-.014** (.856)
Questions total correct	**.041** (.598)	**.077** (.321)	**.059** (.447)	**.065** (.406)	**.019** (.808)	**.015** (.848)
Questions total wrong	**.020** (.796)	**.028** (.715)	**-.006** (.938)	**.050** (.520)	**-.007** (.932)	**.028** (.721)
Questions total distinct correct	**.093** (.232)	**.036** (.645)	**.040** (.609)	**.068** (.387)	**-.065** (.404)	**-.016** (.837)
Videos during semester proportion	**-.014** (.860)	**.023** (.766)	**.008** (.922)	**-.001** (.994)	**.007** (.929)	**.026** (.739)
Questions during semester proportion	**.021** (.791)	**-.046** (.562)	**.056** (.480)	**.010** (.898)	**-.083** (.298)	**.048** (.551)
Most active weekday proportion	**.032** (.700)	**.047** (.568)	**.046** (.573)	**.114** (.164)	**.041** (.618)	**-.098** (.233)
Questions workday proportion	**-.056** (.491)	**-.016** (.846)	**.107** (.194)	**.046** (.574)	**.005** (.953)	**.063** (.442)
Questions mean difference to noon	**.045** (.584)	**.017** (.834)	**-.109** (.191)	**-.025** (.765)	**.005** (.951)	**-.091** (.274)

The second research question about relationships between the measured conceptions about statistics and learning behavior can be quickly answered with their non-existence in this bivariate analyses when looking at Table 6. Only two of the 90 determined correlations are significant. With so many tests of correlations performed, these are even fewer than would be expected by chance as type I errors in case of complete non-existence of correlations. It is only noticeable that both significant correlations refer to videos total district. This number of different videos watched correlates positively with a rules-based conception as well as with a confirmatory conception. The p-values for a descriptive conception and for overall rather dynamic beliefs are also comparatively low. Here, therefore, there could be a small influence of the conceptions about statistics of the students on their behavior in dealing with the videos. A look at the other variables of learning behavior shows that only four further p-values are smaller than 0.2. This again are fewer cases than to be expected at random in case of non-existing correlations. Therefore, at least for all variables except videos total destinct, no correlation to the students' conceptions can be found. Thereby, no systematic difference in the correlations and their p-values can be found even between the three bundles of variables - namely the six variables on the quantity of learning, the three variables on the success in learning, and the five variables on the distribution of learning.

5.4.2 Adding GPA and Gender Towards Structural Equation Models

Even if no significant correlations could be found in the bivariate constellations either between the students' conceptions of statistics and their learning success or between the conceptions and learning behavior, this does not necessarily mean that there is in fact no relationship at all. Thus, it is conceivable that third variables produce mutually offsetting effects that are not seen in bivariate models, but emerge in multivariate models. In order to take this possibility into account, in the following further variables are introduced into the analysis. From previous studies we know that especially the cognitive prerequisites of the students and their gender can be relevant in the context of beliefs (Chiesi & Primi, 2010; Rabin, Krishnan, Bergdoll, & Fogel, 2021). To estimate the cognitive prerequisites of the students, the grade point average of the high-school diploma (high school GPA) collected in the survey can be used. For gender, the gender registered at the university is available in the exam data.

For both high school GPA and gender, it seems reasonable to assume that they have an impact on students' conceptions at the beginning of the course, as well as on their learning behavior, and on their learning success. Therefore, to measure all relationships simultaneously, structural equation models are estimated in which gender and GPA affect all three components. In addition, conceptions of statistics are expected to have an impact on learning success and on learning behavior, which in turn is expected to also have an impact on learning success. In order to avoid having to calculate separate structural equation models for each of the six scales of conceptions, only the scales of data-driven and dynamic beliefs are used for the following considerations and are entered into the structural equation model simultaneously. All structural equation models use maximum likelihood (ML) for estimation and full imputation maximum likelihood (FIML) to cope with missing values. To ensure similar variances all variables were normed on a scale from 0 to 1. Figure 23 shows an example of the entire structural equation model for the variable questions total.

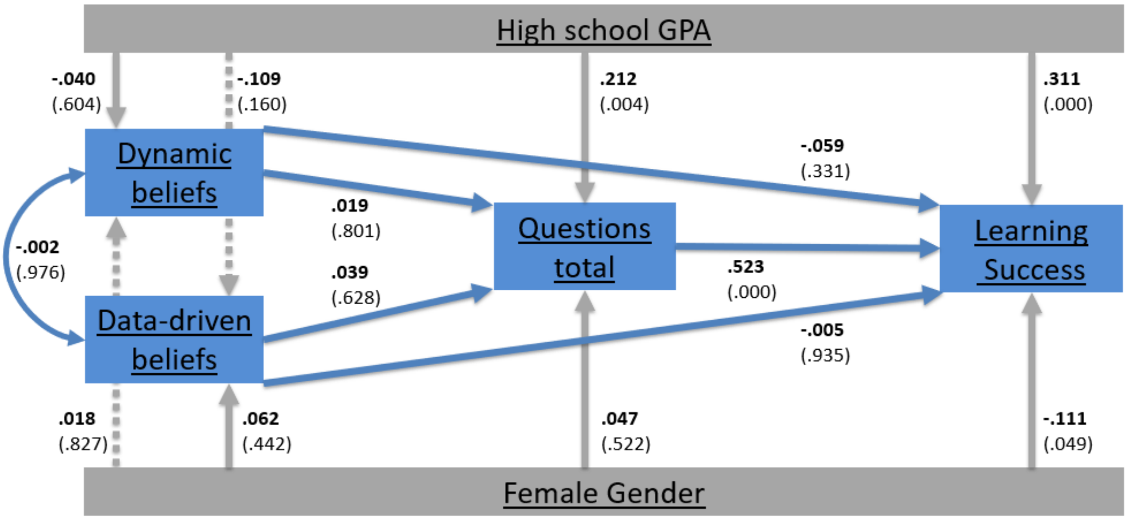

Figure 23: Visualization of the structural equation model using dynamic and data-driven beliefs as predictors of the total number of questions answered and of learning success and using high school GPA and female gender as controlling factors.

An initial view shows that the newly added variables do have an impact on students' learning. While female students performed slightly worse in the exam than males, in particular, a large effect of the high school GPA emerges. In addition, a good high school GPA also has a positive effect on learning success via its effect on the total number of questions answered. Students with a good high school GPA answer a highly significant number of questions more than others do. For female students, it appears that they work on slightly more questions in the sample considered, but this effect is clearly not significant.

The same applies for the effects on beliefs. Female students tend to have slightly higher dynamic beliefs and also slightly higher data-driven beliefs. However, these two effects are also not significant beyond the sample. A good high-school GPA, on the other hand, has a negative effect on dynamic and data-driven beliefs in the sample. Again, however, no significant result emerges.

Looking at the effects of beliefs on learning behavior and learning success also reveals no significant relationships. The effects of beliefs on the number of questions answered have p-values above 0.6, the direct effects on learning success have p-values above 0.3. At the same time, it can be noted that the number of questions answered does indeed have a large effect on learning success. This effect is the strongest in the SEM, even stronger than the effect of the students´ high school GPA.

If we look at other variables of learning behavior, the SEM does not change on the left, where effects of high school GPA and female gender on beliefs are measured (except for minor variations due to deviations in the estimation algorithm). The direct effects of these on learning success often change slightly, but without impacting interpretation. The other seven effects of the model are listed in Table 7, where each line corresponds to a different SEM with a different variable of learning behavior.

Table 7: Weights (and p-values) of the SEM when the variable of the learning behavior is exchanged (thus, each row of the table corresponds to a separate SEM).

	Dynamic beliefs on learning behavior	Data-driven beliefs on learning behavior	High school GPA on learning behavior	Female gender on learning behavior	Dynamic beliefs on learning success	Data-driven beliefs on learning success	Learning behavior on learning success
Videos total	**.002** (.978)	**-.041** (.604)	**.182** (.014)	**-.064** (.389)	**-.047** (.502)	**.024** (.733)	**.212** (.001)
Videos total distinct	**-.088** (.251)	**-.020** (.794)	**.263** (.000)	**.077** (.312)	**-.030** (.666)	**.020** (.779)	**.204** (.005)
Questions total	**.019** (.801)	**.039** (.628)	**.212** (.004)	**.047** (.522)	**-.059** (.331)	**-.005** (.935)	**.523** (.000)
Questions total distinct	**-.048** (.514)	**.024** (.748)	**.347** (.000)	**.056** (.426)	**-.013** (.803)	**-.001** (.987)	**.690** (.000)
Tutorial participation	**.038** (.617)	**.053** (.485)	**.205** (.004)	**.209** (.003)	**-.064** (.333)	**-.007** (.916)	**.408** (.000)
Active tutorial participation	**.015** (.840)	**.082** (.274)	**.208** (.004)	**.226** (.001)	**-.054** (.419)	**-.019** (.776)	**.395** (.000)
Chat definitions total	**.061** (.481)	**.010** (.910)	**.155** (.038)	**.006** (.932)	**-.055** (.441)	**.015** (.836)	**.129** (.057)
Questions total correct	**.036** (.626)	**.038** (.613)	**.303** (.000)	**-.014** (.848)	**-.072** (.215)	**-.007** (.901)	**.586** (.000)

	Dynamic beliefs on learning behavior	Data-driven beliefs on learning behavior	High school GPA on learning behavior	Female gender on learning behavior	Dynamic beliefs on learning success	Data-driven beliefs on learning success	Learning behavior on learning success
Questions total wrong	**-.006** (.939)	**.029** (.734)	**.083** (.266)	**.109** (.142)	**-.045** (.497)	**.005** (.945)	**.384** (.000)
Questions total distinct correct	**-.052** (.473)	**.018** (.804)	**.392** (.000)	**.001** (.985)	**-.007** (.890)	**.002** (.961)	**.739** (.000)
Videos during semester proportion	**-.001** (.995)	**.050** (.522)	**.216** (.003)	**-.032** (.664)	**-.046** (.465)	**-.005** (.942)	**.431** (.000)
Questions during semester proportion	**-.054** (.504)	**.067** (.405)	**.079** (.298)	**-.002** (.975)	**-.042** (.559)	**.010** (.887)	**.120** (.078)
Questions most active weekday proportion	**.034** (.668)	**-.103** (.198)	**-.105** (.176)	**-.128** (.097)	**-.035** (.607)	**-.019** (.782)	**-.328** (.000)
Questions workday proportion	**-.019** (.824)	**.075** (.367)	**.109** (.158)	**.068** (.386)	**-.042** (.543)	**.001** (.994)	**.232** (.001)
Questions mean difference to noon	**.029** (.740)	**-.102** (.230)	**-.114** (.140)	**-.132** (.095)	**-.041** (.560)	**-.003** (.962)	**-.223** (.002)

A look at the right column of Table 7 shows that the total number of questions answered is not the only variable of learning behavior that does have a significant influence on learning success. In fact, the number of definitions asked for in the chat and the proportion of questions worked on during the semester are just slightly non-significant, while all other p-values are <0.1. At the same time, the effects of learning behavior on learning success are usually the strongest effects of each of the models. If we look at the influences on learning behavior, we first notice that the high school GPA generates many significant effects, especially on variables measuring the quantity of learning or success with the training questions. Effects on the spacing of learning are smaller with only one being significant. Female students often are a bit more engaged with (active) tutorial participation being the only significant effect. At the same time spaced learning within the day and week is a bit more common among female students with effects at least below 0.1.

Looking then at the effects of beliefs on learning behavior, we find that only one effect has a p-value less than 0.2. Here, students with data-driven beliefs tend to work slightly more spaced across the week. But this effect is also clearly not significant. All other effects are already small to very small in the sample and have p-values of mostly greater than 0.4. Here, therefore, we cannot speak of an identification of effects. Rather, the overall conclusion must be that even in the multivariate model presented here, students' beliefs have no relevant effect on learning behavior. For direct effects of beliefs on learning success a similar conclusion must be drawn since all effects have p-values bigger than 0.2.

5.5 Conclusion

The present research seeks an answer to the question whether students' conceptions of statistics have an influence on students' learning success and also on their learning behavior. After several steps of analysis, it has to be stated that both parts of the research question have by now to be answered with "no" based on the data of this study. In bivariate analyses, hardly any correlations between the conceptions and learning success or learning behavior could be identified. Also, in the tested multivariate models, no significant effects of the conceptions of statistics on learning behavior or learning success were found.

As one consequence, the measurement of beliefs raises further questions. Not only has the measurement instrument used in this study not yet been tested by other researchers, so that any confirmation of the measurement quality is still pending. In addition, it has to be asked how exactly the measured beliefs are supposed to work. For example, in another study, the authors of this article found that the concepts about statistics do not have a direct effect on student motivation, but have an influence on student motivation via complex interaction effects (see chapter 7 / study 6 of this dissertation). Similarly, in the scenario studied here, it would be conceivable that students' conceptions do not directly translate into altered learning behavior or altered learning success, but that they interact with other psychological factors and possibly also with cognitive and environmental factors in complex ways and thus have a relevance that cannot be described in simple linear terms.

Beyond these measurement problems, it should also be noted that the present study relates to a single course. This course is aimed at first-year students in a social science program. To what extent the findings of this study can be generalized to other age groups and other disciplines must remain an open question. In particular, it should be noted that the course examined was a mandatory course for students. Courses freely chosen by students may be subject to very different logics. So no statement can be made for them based on this study. In addition to further research on the measurement problems mentioned above, the analysis of freely chosen courses should be a goal of further research.

The positive conclusion to be drawn from this research is that it has been possible to link psychometrically measured concepts of learning research with digital behavioral data from a learning management system and with register data from the examination system and to conduct a joint analysis. The authors see great potential for further research in the field of learning analytics in this linking of different types of data.

5.6 Concluding Remarks and Transition to Study 5

Study 4 thus provides very sobering results for both research question 2a and research question 3a of this thesis. All correlations between the beliefs about statistics and learning success are not significant and (mostly) so clearly not significant that this can hardly be explained by a to small number of cases.

The remaining hope for an association between beliefs about statistics and learning success therefore seems to lie in research question 2b of this thesis. If a (weak) link between beliefs about statistics and attitudes towards statistics could be identified, there might be a minimal link between beliefs about statistics and learning success that would be mediated by attitudes towards statistics and that is too small to be detected in Study 4. Learning behavior, however, does not appear to be a mediator between beliefs about statistics and learning success according to Study 4.

Study 6 therefore addresses research question 2b in order to identify such a possible link between beliefs about statistics and attitudes towards statistics. Before this, however, the main research question still needs to be investigated for attitudes towards statistics and statistics anxiety. Study 5 includes attitudes towards statistics and statistics anxiety in a joint model and investigates the extent to which the learning behavior variables known from Study 3, supplemented by some similar variables, are suitable as mediator variables to explain the relationship between the non-cognitive factors and learning success. Study 5 thus answers research questions 3b and 3c of this thesis, which are the core of the entire thesis. Study 5 thus is the midpoint and, in a way, the main point of this thesis.

6 Study 5: The Motivated Are the Successful – But Why? Learning Analytics Shows Affects Affect Scaling, Spacing, and Success of Learning

This chapter originated as a manuscript co-authored with Sebastian Hobert, University of Goettingen. The author of this dissertation is first author of the manuscript. The manuscript is currently under review at a peer-reviewed journal. It is used and pre-printed in this dissertation with the kind permission of the publisher and the co-author.

Berens, F., & Hobert, S. (Under reviewb). The Motivated Are the Successful – But Why? Learning Analytics Shows Affects Are Linked To Scaling, Spacing, and Success of Learning.

Abstract: Across disciplines, motivation, attitudes, and emotions have been shown to influence learning. In particular, motivation and positive attitudes are linked to higher learning success, whereas anxiety is linked to lower success. Regarding the reasons for this linkage, learning behavior is often suggested to be a mediator variable. While some survey-based research suggests learning behavior could be an explanation, studies that combine multiple factors to explain investigate why the motivated are the successful and that use digital process data and learning analytics for evidence are lacking. The present study uses expectancy-value theory and control-value theory to examine links of expectancy, value, and anxiety with learning behavior and learning success to investigate why the motivated are the successful in our classes. In an introductory statistics course for the social sciences, N=181 students were surveyed about their attitudes and anxiety at the beginning of the course, their learning behavior in the learning management system was recorded, and data were linked to exam success. Learning behavior is analyzed in three dimensions: quantity of learning (scaling), distribution of learning (spacing and academic delay), and focus of learning (focus on applied learning). Structural equation models show that all three dimensions of learning behavior are significantly associated with learning success. Furthermore, expectancy and value, mediated by planned effort, are significantly linked to scaling and spacing of learning. At the same time, high-school GPA and female gender load on most variables, revealing confounding in previous studies. The results thus demonstrate the multifaceted links of attitudes and anxiety to learning and remind us which dimensions of learning can be changed to increase learning success.

Keywords: Control Value Theory, Educational Data Mining, Expectancy Value Theory, Self-regulated learning, Statistics Anxiety, Statistics Attitudes

6.1 Introduction

In mathematics and science education, it is not a new finding that motivation, attitudes, interest in the discipline, and anxiety about the discipline correlate with learning success. As early as 1960, Feierabend (1960) concluded in a review that motivation and positive attitudes toward mathematics correlate positively with learning success in mathematics education, while mathematics anxiety correlates negatively with success. Aiken (1970, 1976) generalized Feierabend's finding in two reviews in the seventies by showing the consistency of these relationships for all ages from elementary school to postgraduate level and across U.S. borders. For science education, Schibeci (1984) noted in a review that attitudes toward science have been widely studied and are highly relevant for learning. In 1992, Schiefele, Krapp, and Winteler (1992) showed in a meta-analysis that interest is not only linked to success in mathematics and various natural sciences, but also in social sciences, foreign language learning, and literature. Further reviews and meta-analyses confirmed the correlation between domain-specific attitudes and learning success, e.g., Ma and Kishor (1997) for mathematics, Osborne, Simon, and Collins (2003) for science, and Emmioğlu and Çapa Aydın (2012) for statistics. The negative effect of domain-specific anxiety has also been repeatedly replicated, for example by Zeidner (1991) for mathematics as well as for statistics education.

An important topic of recent research is the question whether the correlations between motivation, attitudes, interest or anxiety, and learning success are due to an effect of these affective factors on learning success or whether, on the contrary, learning success changes the learners' motivation, attitudes, interest and anxiety. Using reciprocal effects models, it was shown that the correlation between academic self-concept and success in mathematics and reading is generated by causal effects in both directions (Marsh, Trautwein, Lüdtke, Köller, & Baumert, 2005). Such reciprocal relationships could also be shown for sports (Marsh, Gerlach, Trautwein, Lüdtke, & Brettschneider, 2007) and subsequently in more general meta-analyses (Marsh & Martin, 2011; H. Wu, Guo, Yang, Le Zhao, & Guo, 2021). Therefore, it can be hypothesized that motivation, attitudes, interest, and anxiety usually have a reciprocal causal relationship to learning success.

Another goal of recent studies on motivation, attitudes, interest, and anxiety and their relationship with learning success is their theoretical embedding. Commonly used for this purpose are Pekrun's Control-Value Theory of Achievement Emotions (CVT) (Pekrun, 2006; Pekrun et al., 2007) and Eccles and Wigfield's Expectancy-Value Theory of Motivation (EVT) (Eccles et al., 1983; Eccles & Wigfield, 2020; Wigfield & Eccles, 2000). Control-Value Theory tends to be used in contexts where more fluid, rather emotional affective factors are studied (Daniels & Stupnisky, 2012). Expectancy-Value Theory tends to be used for more stable affective factors, such as attitudes and interest (Emmioğlu & Çapa Aydın, 2012).

Using these theories, not only can known relationships be theoretically re-embedded and thus, in turn, empirically substantiate the aforementioned theories (Corwyn & McGarry, 2020; Klee, Buehl, & Miller, 2022; Putwain & Wood, 2022). Both theories also provide possible reasons underlying the correlations between motivation, attitudes, interest, anxiety and learning success. Already laid out in Eccles et al.´ (1983) first conception of Expectancy-Value Theory, Control-Value Theory (Pekrun, 2006) in particular sees the learning behavior of the learners as a crucial mediator between motivation, attitudes, interest and anxiety on the one side and success on the other. Following this idea, Dayupay et al. (2022), for example, report effects of expectancy and value on learner-reported learning strategies and on success in mathematics. Artino and Jones (2012) highlight effects of achievement emotions on self-regulated learning behaviors.

Both of the latter, as well as most similar studies, use learners' self-reports in surveys for the analysis of learning behavior. Especially with regard to self-regulated learning, however, self-reports are not unproblematic or can at best only reflect a part of the overall picture (Ye & Pennisi, 2022; Yingbin Zhang et al., 2022). At this point digital behavioral data that capture learning behavior in the learning management system (LMS) can be a valuable addition to learning research, as it has been shown to provide objective and detailed data that are suitable for in-depth analysis of learning behavior and thus offer added value (Gasevic, Jovanovic, Pardo, & Dawson, 2017; Ober et al., 2021).

Following this idea, some recent studies have already investigated the influence of motivation, attitudes, interest or anxiety on learning behavior through learning analytics of digital behavioral data. With regard to these studies, it can be observed that most of them either examine a longer learning process, e.g. a semester-long course, but only analyze the learning behavior through one or at least only a few facets. On the other hand, there are studies that look at learning in a much more multifaceted way, but shorten the learning to a short learning session. An example of the former is the work of Dabas et al. (2021), which analyze the influence of motivation, interest, learning strategies and gender on learning behavior and learning success. Learning behavior is quantified one-dimensionally as the frequency of logins to the LMS. S. Cheng, Xie, and Collier (2023) measure two facets of how students manage their learning time to analyze relationships to motivational beliefs and achievement. Liao and Wu (2023) look at three behavioral patterns in learning with videos and their relationship to motivation and success. On the other hand, Yingbin Zhang et al. (2022) look at one unit of learning in more depth by examining the influence of emotions on learning behavior by extracting learning strategies as well as off task behavior from the digital behavioral data.

What is lacking, however, are studies that relate factors such as motivation, attitudes, interest, and anxiety to learning behavior and success in a way that observes learning in the field over an extended period of time and uses evolving digital behavioral data to describe learning behavior in a multifaceted way. This research addresses this gap and approaches the research question:

RQ: "Why are the motivated the successful in our classes?"

To address this research question, we specifically examine how differences in learning behavior can explain the relationship between motivation and success. For this purpose, learning behavior is analyzed in the dimensions of scaling, spacing, and focus of learning to explain the correlation between expectancy, value, anxiety, and learning success. In addition to learning behavior, gender and high school grade point average (GPA) are considered possible explanations for the link between motivation and success. Understanding the relationship between motivation and success will help educational research and practice to better understand the relationship of motivation and learning and, thus, to make better use of it. At the same time, approaches can be derived that can show less motivated students ways to nevertheless be successful learners.

The research is done using the example of an introductory statistics course for social sciences held at a large German public university. A survey at the beginning of the course collects the students' attitudes and anxiety, the learning management system records their learning behavior in detail during the lecture period, and the final exam measures learning success at the end of the lecture period. Based on these data, the following analyses show that expectancy and value relate to two measured dimensions of learning behavior, which in turn relate to learning success.

6.2 Background

6.2.1 Expectancy-Value Theory

While the term motivation is widespread and commonly understood in everyday language, there is no agreed upon concept of motivation in educational research. Rather, there are several, partly parallel partly very different families of terms that provide a theoretical concept of motivation. Deci and Ryan's Self-Determination Theory (Deci & Ryan, 1985; Ryan & Deci, 2020), for example, is particularly well known, describing the experience of autonomy, the experience of competence, and perceived relatedness as factors influencing behavior. In the area of subject-related affective influences on learning, the Expectancy-Value Theory (EVT) has become a particularly established theory of motivation, which was developed by Eccles and Wigfield following some precursors (Eccles et al., 1983; Eccles & Wigfield, 2020; Wigfield et al., 2015; Wigfield & Eccles, 2000).

In their first version of the theory, Eccles et al. (1983) describe a network of relationships in which the learner's social context is reflected in assumptions the learner makes about what is expected of him or her. From this and from the attitudes of those in the social context, as well as from some other factors, the learner develops personal goals from which he or she derives the value of a particular task or discipline. This derived value is then one of two factors influencing the learner's success. On the other hand, prior experiences of the learner, together with attitudes of the social environment, influence the learner's view of his or her own learning past and, thus, attitudes about his or her own abilities with respect to future tasks. From this develops the learner's expectation of his or her ability to master the task or discipline. This expectancy is the second influencing factor besides value that is reflected in success, from which the theory then received its name Expectancy-Value Theory. In a later work concisely formulated expectancy examines the learner's response to the question "Can I do this task?" (Wigfield et al., 2015, p. 659). Value describes the worth and usefulness of the task to the learner and, thus, can be summarized in the question "Do I want to do this task?" (Wigfield et al., 2015, p. 659).

In statistics, the discipline in which this study was conducted, EVT is exceptionally important. This is particularly due to the fact that in statistics education research, learners' attitudes are by far the most commonly studied affective factors of learning (Emmioğlu & Çapa Aydın, 2012; Whitaker et al., 2022). Almost all of these studies are conducted using the Survey of Attitudes toward Statistics (SATS), which is available in two versions, one with 28 items (Schau et al., 1995) and one with 36 items (Schau, 2003). In its development and conception, this survey is explicitly based on Expectancy-Value Theory and, thus, also places all studies using it under this umbrella. The development of this widespread instrument in the 1990s also explains why the field of statistics education continues to work on the basis of EVT and not on the basis of situated EVT (Eccles & Wigfield, 2020). This work follows this approach for practical reasons, but does not recognize any model-relevant differences between EVT and situated EVT for the present purpose.

To implement the EVT, the SATS in its version with 36 items is divided into six dimensions: affect ("I will like statistics."), cognitive competence ("I will find it difficult to understand statistical concepts."), value ("Statistics is irrelevant in my life."), difficulty ("Statistics is a complicated subject."), interest ("I am interested in learning statistics."), and effort ("I plan to work hard in my statistics course.") (Schau, 2003). Cognitive competence, difficulty, and affect represent the expectancy. The other three concepts are assigned to the value of EVT. Empirical tests show that a 6-dimensional structure of the SATS can be confirmed (Schau, 2003; Tempelaar et al., 2007), but some studies show that fewer dimensions can be sufficient (Hommik & Luik, 2017; Vanhoof et al., 2011).

Especially for the construct “effort” empirical evidence shows that it does not fit well to the other two dimensions of value. Based on these findings and further theoretical considerations, Schau and colleagues developed the Students' Attitudes Toward Statistics-Model (SATS-M) (Ramirez et al., 2012). In their model, they adapt ideas from EVT about the influence of the learner's environment, personality, and experiences on his or her motivational attitudes. However, they separate effort from the other five constructs measuring expectancy and value and now describe a chain of effects in which effort mediates the relationship between expectancy and value on the one hand and success in statistics on the other.

This idea is adopted for this study. Based on EVT and the influence of expectancy and value on success described there, a model based on the SATS-M is designed as the first step of the model development of this study. In this first version of the model tested in this research, expectancy and value have an effect on the planned effort, which in turn has effect on success. The full model can be found below in Figure 24.

6.2.2 Control-Value Theory

In addition to attitudes toward statistics, emotions are the second important area of motivational influences on learning examined in statistics education research. Here, most research focuses on statistics anxiety (Chiesi & Primi, 2010; Hedges, 2017; Valle et al., 2021). Such studies on statistics anxiety are usually theoretically embedded via Control-Value Theory (Pekrun, 2006; Pekrun et al., 2007).

For their Control-Value Theory of Achievement Emotions (CVT), Pekrun and colleagues (Pekrun, 2006; Pekrun et al., 2007) construct a causal chain in which environmental factors, such as the design of a learning environment, influence how learners attribute value to learning and the subject matter, and the extent to which they expect control and feasibility. These expectations about control and value then lead to achievement emotions, which in turn have an impact on learning behavior. Learning success depends in turn on learning behavior, but also on cognitive factors such as intelligence or competencies. CVT, thus, frames the influence of emotions on learning success and sees learning behavior as a mediator for this. One emotion that can have such an effect on learning behavior and on learning success is anxiety (Pekrun et al., 2007).

CVT can complement the model begun in the previous subsection in two ways. First, it provides emotion, in this study anxiety, as a third factor influencing learning, alongside expectancy and value that come from EVT. Even though EVT and CVT are coexisting theories that have rarely been integrated so far, it can be argued that the expectancy and value constructs of EVT are on the same level as the control and value constructs are in CVT. Thus, they can be considered antecedents of anxiety in a joint model. In the refined model of this study, expectancy and value, thus, have effects on anxiety and on effort. In addition, an effect of anxiety on intended effort can be assumed.

The second significant enhancement to the theoretical model of this study is the strong emphasis on learning behavior in CVT. While (Eccles et al., 1983) consider influences of expectancy and value on learning behavior possible, learning behavior is nevertheless not part of their overall model. In Pekrun´s (2006) CVT, however, learning behavior is a fixed link in the causal chain that, together with cognitive factors, is decisive for learning success. CVT, thus, provides a possible explanation of why expectancy, value, and emotions are reflected in learning success, namely by altering learning behavior. If this study is to investigate why motivated learners in terms of EVT are successful, it makes sense to assume learning behavior as a mediator to learning success for all three factors, expectancy, value, and anxiety.

Furthermore, if learning behavior is critical to learning success along with cognitive factors such as intelligence or various competencies, this suggests the need to add a cognitive factor to the model of this study to control for its influence. A look at empirical studies shows that intelligence or competence not only affect learning success, but also have effects on, for example, expectancy or anxiety (Chiesi & Primi, 2010; Hood et al., 2012). A common approach to measuring these cognitive factors is to use the high school GPA (Johnson & Kuennen, 2006; Kowski, 2013).

Another variable that has already been shown many times to be related to many of the variables discussed in the model of this study is gender. Already in the aforementioned studies by Feierabend (1960) and Eccles et al. (1983), the investigation of gender effects was a major concern of these studies. But more recent studies have also shown that differences between genders in attitudes (Dabas et al., 2021; Lihui Sun, Hu, & Zhou, 2022), behavior (Abraham & Barker, 2015), and success (Dabas et al., 2021; Martin, Hughes, & Fugelsang, 2017) are found on a regular basis. The variable (female) gender is therefore included in the model of this study. It is assumed that, just like high school GPA, it relates to each of the components of the model. Overall, this results in the theoretical model of this study as shown in Figure 24, which will be tested varying the measurement of learning behavior used.

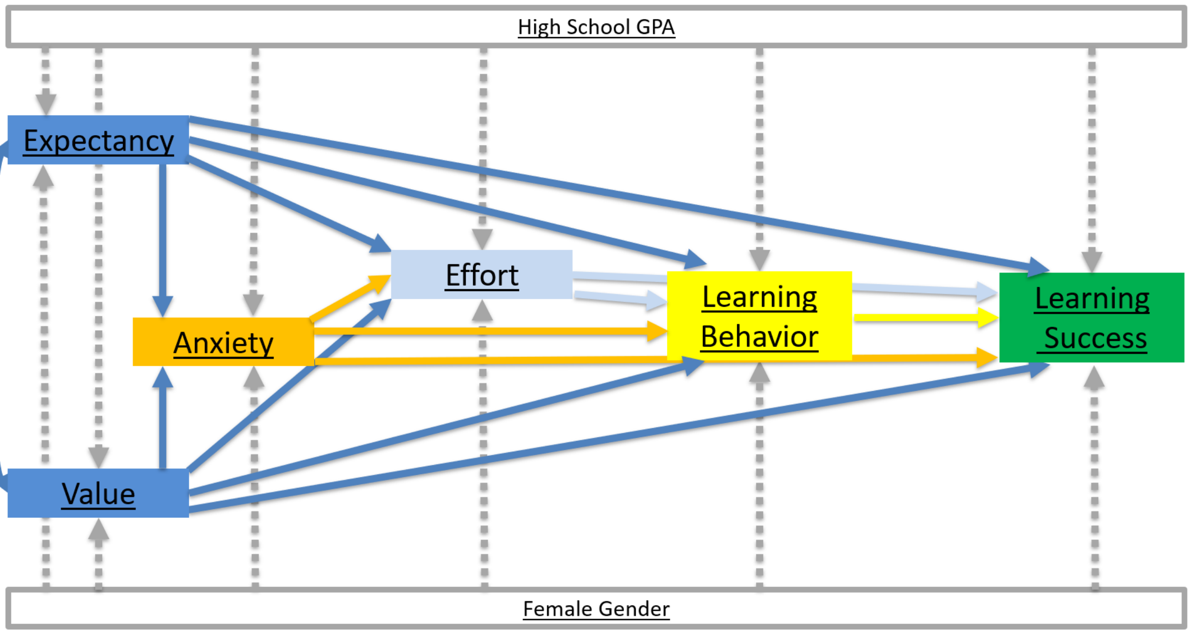

Figure 24: Full theoretical model of this research, based on EVT (blue) and CVT (orange)

6.2.3 Learning Behavior

After deriving the theoretical model, we need to specify how learning behavior is to be measured. It is evident that learning behavior is a multi-dimensional phenomenon that cannot be fully grasped but can only be observed through different lenses. In this research, we consider three dimensions of learning behavior for which there is empirical evidence that they are linked to learning success.

As the first dimension of learning behavior, we consider the quantity of learning. Especially in phases of self-regulated learning, learners can independently scale how much or how long they want to learn. But even in face-to-face teaching, learners can control whether they participate and, if so, how active they are in doing so. The fact that the quantity of learning is linked to learning success is intuitive. However, there is also empirical evidence for this. Singh, Granville, and Dika (2002) for example find evidence for a positive relationship of the amount of time spent learning with success in mathematics and science in 8th grade. Lavidas et al. (2020) find a positive relationship of both frequency of course attendance and intensity of engagement during course in an introductory statistics course at a university. Also using digital behavioral data, it could be demonstrated that higher learning activity is associated with higher learning success (Saqr & López-Pernas, 2021; Tempelaar, Nguyen, & Rienties, 2020). The quantity of learning, i.e., how much learners scale their learning engagement, is therefore the first dimension of learning behavior considered in this study.

The second dimension of learning behavior considered in this study is the distribution of the invested learning time. We distinguish between two subdimensions, which are related but also different, and for each of which a relevance for learning success is shown. The first subdimension of the distribution of learning is spacing. It has been shown that dividing a fixed amount of learning into smaller units and spreading it out over a longer period of time is more beneficial for learning success than massing learning (Bjork & Allen, 1970; Cepeda et al., 2006). The positive effect of spacing has also been shown in experiments on learning mathematical and scientific content (Lyle et al., 2022; Vlach & Sandhofer, 2012). Recent studies have also gone beyond studies in laboratory experiments and have confirmed the positive effect of spacing in the field by analyzing digital behavioral data on spaced learning (Barthakur et al., 2021; Yeckehzaare et al., 2022). In this study, spacing is therefore considered as a relevant subdimension of the dimension distribution of learning and is therefore included in the analysis as a perspective on learning behavior.

As a second subdimension of the distribution of learning, we consider the frequently studied procrastination in learning (Kim & Seo, 2015; Rabin et al., 2021; Steel, 2007). Procrastination means that learners postpone their learning activity against their better judgment (S.-L. Cheng & Xie, 2021). Procrastination has been shown to have negative effect on learning success (Kim & Seo, 2015; Rabin et al., 2021; J.-Y. Wu, 2021). Closely related to the concept of procrastination is the concept of academic delay (S. Cheng et al., 2023). Both concepts have in common that learning does not take place immediately but is postponed to a later point in time. However, while in procrastination this postponement is necessarily perceived as negative by the learner, the concept of academic delay remains neutral and includes any kind of postponement of learning, without distinguishing between reasonably planned postponements and negatively framed postponements (S.-L. Cheng & Xie, 2021). The concept of academic delay has the advantage of being much easier to measure based on digital behavioral data, as the shift in learning found in the behavioral data does not need to be scrutinized for the rationale behind the shift. Nevertheless, it has been shown that academic delay tends to be negatively associated with learning success (S. Cheng et al., 2023). In this study, ademic delay is therefore included as a second subdimension of the distribution of learning.

In addition to the quantity and distribution of learning, it is above all the "how?" of learning that influences learning success. If one understands these qualitative aspects of learning as one dimension of learning behavior, there are countless perspectives from which one can look at the quality of learning. An effect on learning success could be shown for general learning strategies as well as for subject-specific learning strategies (e.g., Dayupay et al., 2022; Liebendörfer et al., 2022). However, different learning strategies not only influence learning success, but are themselves influenced by learners' attitudes or anxiety (e.g., Rozgonjuk, Kraav, Mikkor, Orav-Puurand, & Täht, 2020). However, looking at the aforementioned and other publications on learning strategies and their influence on learning success shows that one study cannot possibly consider all perspectives on learning strategies. Nevertheless, this study wants to make a start by using the available digital behavioral data to gain insight into the quality of learning. For this purpose, this study compares how much of the learning involves own application of the learned content in contrast to pure reception and reproduction. Thus, for the third dimension of learning behavior, the focus of learning, this study considers the proportion of actively applying learning relative to all learning, understood as a first step toward examining learning strategies in the context of affective influences on learning.

6.3 Methods

6.3.1 Educational Setting of the Study

The data for the reported study was collected at a large public university in Germany. The course studied is an introductory course in statistics, which is a compulsory course for first-year students in most social science programs. Unlike in other university systems, however, this only means that students must pass a final exam about the course. All other parts of the course, whether it be lectures, tutorials, or practice materials, are to be understood as voluntary offerings that students are free to use to prepare for this final exam. The investigated introductory course in statistics consists of 14 weeks of structured course work for the students. This is followed by two weeks at the students' free disposal to prepare for the final exam. The structured program of the course consists of four parts:

1. Regularly there is a lecture of 90 minutes once a week, which introduces new content. Since data collection for this study took place during pandemic lockdowns, this lecture was replaced by video recordings. Each Tuesday, five to nine new videos were made available online, which totaled to about 90 minutes length a week. These videos were available to students on demand at any time from upload to exam.
2. Through the course's LMS, students were also provided with practice problems to learn and to deepen the content presented in the videos. Each week, about 20 exercise tasks were provided, so that a total of 249 tasks were made available during the course, which remained available from the upload until the exam.
3. After the students had time to study the videos and work on the exercises, tutorials were held on Mondays in which experienced students helped the students in small groups. These tutorials took place as videoconferences and aimed to clarify problems of the learners with the videos and to discuss problematic exercises together. The participating learners were encouraged to actively participate in the tutorials through participation in the discussions, but also through an audience response system.
4. During the two weeks at the students' free disposal just before the exam, students were offered support in the form of additional material and 152 more practice problems.

All elements of teaching were supported by the course's learning management system. It had been developed by the authors of this study early in 2019 (see chapter 3 / study 2 of this dissertation) and was used in teaching for the first time in summer 2019. The LMS is built as a multifunctional system that bundles all of the course's contents into one digital structure. Thus, the LMS can not only provide materials, such as lecture slides, but also serves as a comprehensive e-learning app. With its help, students can access the videos, participate in audience response questions during tutorials, and reach, answer, and receive feedback on the practice exercises. In addition, the LMS allows students to view their own learning statistics and a chatbot answers statistics-related questions. All interactions of students who consented to participate in research activities are stored by the system as log entries that are tagged with a pseudonym. The LMS's digital behavioral data thus enables extensive analysis of the student's learning behavior.

6.3.2 Operationalization of Learning Behavior

Learning behavior can in principle be assessed both by self-reports of the learners and by observing the learning. However, as discussed above, self-reports are not unproblematic and are often biased and incomplete (Ye & Pennisi, 2022; Yingbin Zhang et al., 2022). On the other hand, digital behavioral data provide a way to observe learning that allows for high granularity and depth (Gasevic et al., 2017; Ober et al., 2021; Tempelaar, Nguyen, & Rienties, 2020).

Regarding the operationalization and use of digital behavioral data, many approaches already exist, while at the same time it is to be encouraged to develop own measures adapted to the research problem (Riel et al., 2018). Derived from this, with respect to measuring the quantity of learning, we are guided by several ideas of counting the frequency of certain learning behaviors in the LMS (e.g., Henrie, Halverson, & Graham, 2015). Regarding the measurement of spacing, we refer to Molenaar and Wise´s (2022) and Tempelaar´s (2023) encouragement to look at time in fine granulation and multidimensionally and develop our own operationalization, based on Molenaar and Wise´s (2022) ideas about comparisons of rates. Regarding academic delay we follow Riel et al. (2018). Last, we also follow Saint, Fan, Gašević, and Pardo (2022) approaches to study self-regulated learning in its quality. Overall, we consider the following variables:

To analyze the quantity of learning, we look at the use of videos, the work with the exercises, and (active) participation in the tutorials. For each of these three aspects we look at two factors, which results in six variables describing the quantity of learning:

- **Videos total** and **Questions total** count the total number of times a video was watched / a question was answered during the course.
- **Videos total distinct** and **Questions total distinct** counts how many videos were watched / questions were answered at least once during the course. In contrast to above, repeated watching / answering is not counted multiple times.
- **Tutorial participations** counts how many tutorials the student has participated in during the course. Decisive is the login to the corresponding videoconference.

- **Active tutorial participations** counts how many tutorials the student has actively participated in. Decisive is how often the student has answered at least half of the audience response prompts asked during the tutorial.

To analyze the distribution of learning, we look at six variables that measure spacing and one variable that measures academic delay. In terms of spacing, we look at spacing across the full course, spacing within weeks, and spacing within the day. This way, we derive the following variables describing the distribution of learning:

- **Videos during term proportion** and **Questions during term proportion** determine the proportion of video watches (within videos total) / questions answered (within questions total) that were done during term rather than during exam preparation.
- **Videos most active day proportion** and **Questions most active day proportion** determine the proportion of video watches (within videos total) / questions answered (within questions total) that were done on the weekday with the most video watches /answered questions rather than on one of the six other days of the week.
- **Videos workday proportion** and **Questions workday proportion** determine the proportion of video watches (within videos total) / questions answered (within questions total) that were done between 9 am and 5 pm rather than before or after that interval.
- **Videos average delay** determines the average number of hours that pass between the upload of a new video and the first time the student starts the video.

In terms of the quality of learning, the digital behavioral data available allows to distinguish the questions answered in terms of whether students apply a learned content themselves (and then enter the result of a calculation or algorithm into the answer field) or whether they merely reproduce knowledge (by answering a single- or multiple-choice question). We interpret this information as a measure of how much students focus their learning on own application. For this purpose, we determine two variables that describe the dimension focus of learning:

- **Rate of application in questions total** determines the proportion of application questions among all answered questions.

- **Rate of application in questions total distinct** determines the proportion of application questions among the questions that got answered at least once.

6.3.3 Measurement of Expectancy, Value, Anxiety and Effort

Expectancy, value, anxiety, and effort were measured using an online survey prior to the course. The survey consisted of three parts: Questions on attitudes toward statistics, to be answered on a seven-point Likert scale; questions on statistics anxiety, also to be answered on a seven-point Likert scale; and some additional questions, for example, on the student´s high school GPA.

SATS-36 (Schau, 2003) was used to measure attitudes toward statistics because it is considered the best available instrument on attitudes in statistics education research (Nolan et al., 2012; Whitaker et al., 2022) and because it maps EVT particularly appropriately. The construct expectancy was formed from this survey as an average over the items on cognitive competence and difficulty. The mean across the items of the constructs interest, affect, and value forms the measure of the concept value of EVT. The construct effort is measured via its four items provided in SATS-36.

Testing the measurement model of the SATS yields an RMSEA of 0.074 and a CFI of 0.838. These values are not great, but they are still not unacceptably bad given the size of the model (McNeish & Wolf, 2023). Since no individual deletions of items would have yielded a relevant improvement of the model, we stick to the use of the entire instrument used in the literature for the analyses.

Statistics Anxiety is measured via the instrument STARS (Cruise et al., 1985), which has been evaluated as suitable for measuring statistics anxiety (Cruise et al., 1985; Hanna et al., 2008). The constructs worth of statistics and self-concept are not used for this study as they are too similar to attitudes. The other subdimensions of statistics anxiety, namely anxiety about statistics exams, anxiety about interpreting statistics, and anxiety about asking teachers or fellow students for help in statistics, correlate so highly that one concept anxiety is build using all items of these constructs.

As further information relevant to this study, we extract the students' high school graduation average (GPA) from the questionnaire. In the SEM it is coded in a way that high values represent a good GPA, while low values represent a poor GPA.

6.3.4 Exam, Data Matching, and Data Privacy

In this study, learning success is measured via the final examination of the course. The final exam takes place in the form of an electronic written exam in the university's e-examination center. Despite the pandemic, this was also the case in the course studied. The exam is designed to last 90 minutes. A total of 100 points can be achieved. Of these, 40 points are to be earned through calculations and algorithms, 35 points through interpretations of statistical results, and 25 points through proof of conceptual understanding. The exam is passed by those who obtain at least 50 points.

271 students took the exam in the course studied. Beside the achieved score in the exam, from the exam data we also extract the gender of students as registered by the university. Furthermore, from the exam data we infer whether the student has already taken this course exam in a previous year and failed it. Since we assume that previous participation in the course has an impact on all other measured constructs, we exclude students with previous failing of the course exam from the analysis. This leaves 238 students who took the exam for the first time in the course under study.

Hash-based pseudonyms were used to gather and merge the learning behavioral, exam and survey data used for this study. To protect privacy, the instructor did not and does not have access to the hash function used. Data curation was only started after an embargo period of three month after the exam. Additionally, all data was anonymized before analyzing it for research purposes. This procedure had previously been agreed with the university's data protection officer and complied with the university's data protection requirements.

As a result, linking survey data, digital behavioral data, and exam data was possible for N = 181 students, which is the sample for this study. Of these, 56.9% are female. 80.1% took the course in the recommended semester, 12.2% one year later, the rest even later. The average high school GPA is 2.30, which is very slightly better than the national average.

6.3.5 Data Analysis

Empirical testing of the theoretical model of this study is performed by linear structural equation models (SEM) estimated with version 0.6-13 of the R package lavaan (Rosseel, 2012). Each SEM tests the theoretical model described in Figure 24. However, since learning behavior is operationalized in a total of 15 different ways, the model from Figure 24 is estimated a total of 15 times accordingly. Finally, we estimate an integrated model that uses one variable per dimension of learning behavior. To show the influence of motivation on success-relevant learning behavior, we select those variables within each dimension that best predict learning success.

All SEM presented below were estimated by maximum likelihood (ML) estimation. All underlying variables that are not by themselves proportions were previously normalized to the interval [0,1] to facilitate SEM estimation through more similar variations. In each case, we report the standardized weight of an arrow of the SEM along with the corresponding p-value, which is reported in parentheses.

Because some variables studied (especially of learning behavior) have some skewness of the data, we also estimated all reported SEM by robust maximum likelihood (MLR) estimation. In addition, we also recoded all studied variables into rank variables and recomputed all SEM over the rank variables using both ML and MLR estimation. Because none of the three alternative methods for calculating the SEM produced substantially different results, we report the initial results using ML estimates of normalized data.

6.4 Results

6.4.1 Motivation Is Linked to the Scaling of Learning

To pursue the goal of this study of relating expectancy, value, and anxiety to different dimensions of learning behavior and to learning success, the theoretical model summarized in Figure 24 is tested for all operationalizations of learning behavior individually. We begin with the number of videos watched by the students in total as a mediator variable between the affective variables and learning success. Figure 25 shows the resulting model.

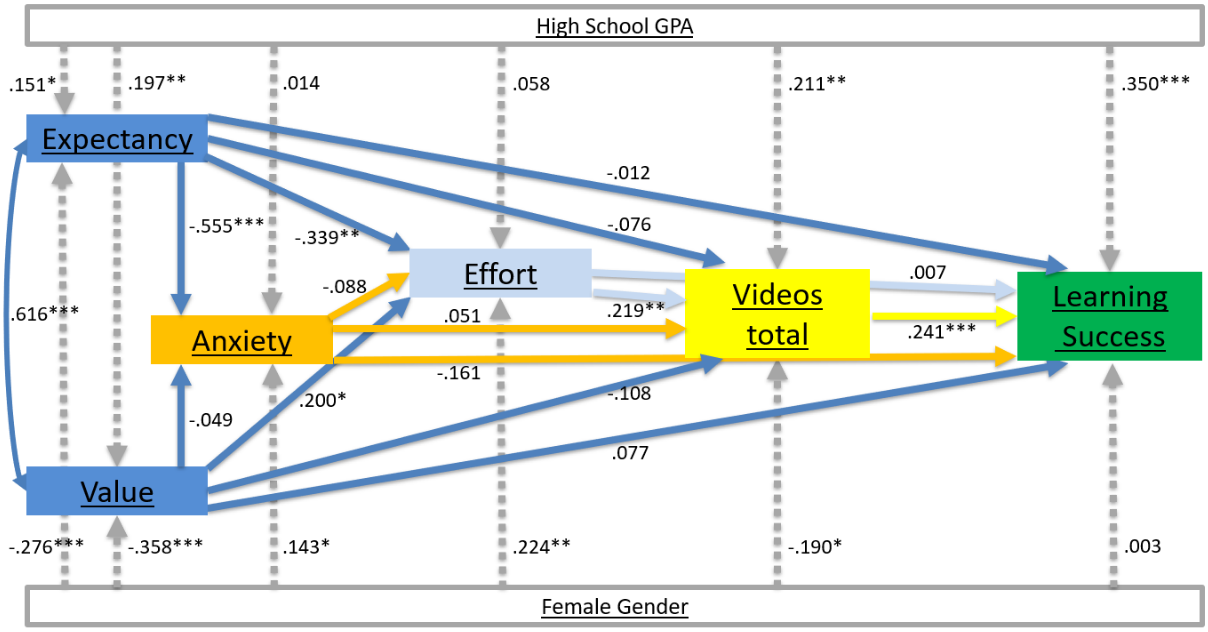

Figure 25: Expectancy, value, anxiety and effort are linked to students´ total number of watched videos

First, the high importance of the control variables for the model is striking. The high school GPA is significantly linked to expectancy, value, the number of videos watched, and learning success. The direct link to learning success is even the strongest among the links, despite controlling for the others. Female gender is significantly linked to all affective factors and watching videos; the gender differences regarding expectancy and value are particularly strong.

A closer look at the relationships between the affective factors reveals that expectancy is significantly linked to anxiety, but that value has basically no link to anxiety. In turn, anxiety itself is not significantly linked neither on the intended effort, nor on the learning behavior or the learning success. Expectancy and value, on the other hand, both are significantly linked to the intended effort, with a high expectancy beeing negatively related to effort, while a high value is positively linked to the intended effort. Expectancy and value do not have any other direct links to learning behavior or learning success. However, since intended effort in turn is significantly linked to learning behavior, expectancy and value are associated with learning behavior mediated by intended effort. Furthermore, a quite strong significant link between learning with videos and learning success is revealed. Overall, the first SEM examined shows that expectancy and value, mediated by effort, are linked to the scaling of learning and thus to learning success, while we do not find that for anxiety.

If we now change the considered variable of learning behavior to videos total district, some relationships in the model change, in which the variable of learning behavior is not directly involved. Comparing all six models with a variable of scaling, however, shows that these changes are mostly small and that the overall interpretation does not change. The conclusions drawn from the example of total videos therefore also apply to the other five measurements of scaling. In Table 8, therefore, only the five relationships are shown in which learning behavior is directly involved. Each column of the table is to be understood as an excerpt of a separate SEM.

Table 8: SEMs of expectancy, value, anxiety and effort on the quantity of learning and of this behavior on success; each column corresponds to a SEM as shown in Figure 25 with the header of the column as the analyzed variable of learning behavior.

	Videos total	Videos total distinct	Questions total	Questions total distinct	Tutorial Participations	Active tutorial participations
Learning behavior on success	**.241** (.000)	**.243** (.001)	**.537** (.000)	**.701** (.000)	**.412** (.000)	**.399** (.000)
Effort on learning behavior	**.219** (.003)	**.251** (.001)	**.164** (.030)	**.196** (.007)	**.224** (.002)	**.231** (.002)
Anxiety on learning behavior	**.051** (.618)	**.088** (.400)	**-.169** (.084)	**-.153** (.111)	**-.045** (.649)	**-.015** (.878)
Expectancy on learning behavior	**-.076** (.502)	**-.093** (.406)	**-.182** (.104)	**-.071** (.513)	**-.024** (.824)	**.031** (.778)
Value on learning bahavior	**-.108** (.271)	**-.074** (.451)	**-.032** (.750)	**.005** (.962)	**-.034** (.722)	**-.052** (.585)

A look at Table 8 shows that the various operationalizations of scaling have high commonalities. (Active) Participation in tutorials is somewhat stronger linked to learning success than working with videos, while working on questions is linked even stronger. Thus, overall it can be seen that scaling is highly relevant for learning success in all subtypes. Looking at the antecedents of scaling, significant links are found for the planned effort in all SEM, which differ only slightly in magnitude. Looking at the links of anxiety to learning behavior, it is noticeable that anxiety is slightly negatively linked to scaling when working with quizzes, even if this relationship is (just) not significant. Expectancy and value have no significant direct link to scaling. Overall, it can be seen for all variants of scaling that expectancy and value are linked to learning behavior via effort and that behavior is linked to learning success, while anxiety has no such link.

6.4.2 Motivation Is Linked to the Distribution of Learning

If we now look at the distribution of learning, the comparison of the total of seven models with the six models on scaled learning shows that the relationships in which learning behavior is not directly involved are again not substantially different. The changes in some parameters are somewhat larger than within scaling, yet the interpretations mentioned remain basically the same.

Table 9: SEMs of expectancy, value, anxiety and effort on the spacing of learning and of this behavior on success; each column corresponds to a SEM as shown in Figure 25 with the header of the column as the analyzed variable of learning behavior

	Videos during term proportion	Questions during term proportion	Videos most active day proportion	Questions most active day proportion	Videos workday proportion	Questions workday proportion
Learning behavior on success	**.459** (.000)	**.132** (.057)	**.167** (.013)	**-.323** (.000)	**.295** (.000)	**.218** (.003)

Effort on learning behavior	**.165** (.034)	**.155** (.047)	**.141** (.078)	**.051** (.517)	**.153** (.051)	**.164** (.035)
Anxiety on learning behavior	**.124** (.219)	**.185** (.081)	**-.065** (.540)	**.176** (.104)	**-.175** (.091)	**-.389** (.000)
Expectancy on learning behavior	**.119** (.302)	**.051** (.664)	**.171** (.149)	**.312** (.009)	**-.125** (.278)	**-.253** (.033)
Value on learning bahavior	**-.123** (.224)	**.160** (.118)	**-.146** (.159)	**-.165** (.112)	**-.055** (.592)	**-.016** (.881)

If we look at the direct relationships of spacing in Table 9, we first see that a high proportion of learning during the course as expected is positively linked to learning success. For the reception of videos this is even stronger than for own work with questions. Intended effort is comparably strong significantly linked to both variables. However, although not significant, anxiety is also positively linked to spacing within the course, at least more strongly than anxiety is relatedlinked to Sscaling. No direct links can be found for expectancy and value.

Looking at spacing within the week shows surprisingly that for working with videos massing learning on one day of the week is linked to higher success, while with working with questions it is the other way around. Effort is positively linked to the apparently productive massing of videos, while the other three factors are not linked to spacing within the week. On the other hand, expectancy is the only factor that is significantly positively linked to the productive spacing of tasks over the week, but anxiety and value also have rather small p-values, with value being negatively linked.

For the distribution of learning activities over the course of the day, videos and questions alike show that working between 9 and 5 p.m. is positively linked to learning success. Planned effort is positively linked to this favorable learning behavior, while anxiety is negatively linked. Expectancy also is negatively linked to working with questions between 9 am and 5 pm, while for value there is no link.

Overall, then, the picture for spacing is not quite as consistent as for scaling. Nevertheless, the clear tendency can be seen that effort is linked to spacing. Anxiety also has at least a little association here. Expectation and value, on the other hand, have hardly any direct link.

Looking at academic delay, there are many similarities with spacing over the course. A high average time delay in watching videos is strongly and highly significantly negatively linked (-0.501***) to learning success. Planned effort is significantly negatively linked to delay (-0.176*), while the other three affective factors have no direct link.

6.4.3 Motivation Is Linked to the Focus of Learning

Finally, looking at the focus on active application as a third dimension of learning behavior, we see that both operationalizations produce almost identical results. As the results from Table 10 show, a stronger focus on own application is strongly positively linked to learning success. However, none of the measured affective factors shows associations to serve as an explanation for differences in this learning focus.

Table 10: Effects of expectancy, value, anxiety and effort on the focus of learning and of this behavior on success; each column corresponds to a SEM as shown in Figure 25 with the header of the column as the analyzed variable of learning behavior

	Rate of application in questions total	Rate of application in questions total distinct
Learning behavior on success	**.430** (.000)	**.428** (.000)
Effort on learning behavior	**.013** (.864)	**.012** (.873)
Anxiety on learning behavior	**-.103** (.304)	**-.093** (.358)

Expectancy on learning behavior	**-.181** (.110)	**-.131** (.251)
Value on learning bahavior	**.066** (.509)	**-.001** (.993)

6.4.4 Integrated Model

The 15 models tested show that learning behavior is significantly linked to learning success and that at the same time affective factors are linked to learning behavior, mostly via the intended effort. This also makes clear that all estimated models are incomplete, since they always consider only one aspect of learning behavior, while omitting other aspects that have been shown to be relevant in other models as well. A complete model would therefore have to include many dimensions of learning behavior. At the same time, however, many of the measured variables of learning behavior are highly correlated with each other. Therefore, they cannot be estimated in a joint model without measurement problems. In order to meet the tension of simultaneously representing as many dimensions of learning behavior as possible in the model and producing as little collinearity between the variables as possible, an integrated model is tested as the last step of data analysis, in which the three dimensions of learning behavior studied in this research are represented. To select the most relevant variable for learning success within each of the dimensions scaling, spacing and focus of learning for this purpose, the variable that is most predictive of learning success is selected to represent the dimension. Of interest then is how strongly these relevant variables, under each other´s control, are related to motivation and anxiety.

Thus, the integrated model places questions total district, videos during term proportion and proportion of application in questions total in the place of learning behavior. In the estimation, links of all affective factors to all three dimensions of learning behavior, as well as correlations between the dimensions, are allowed. Figure 26 shows the integrated model, with all nonsignificant effects omitted for the purpose of clarity of the model.

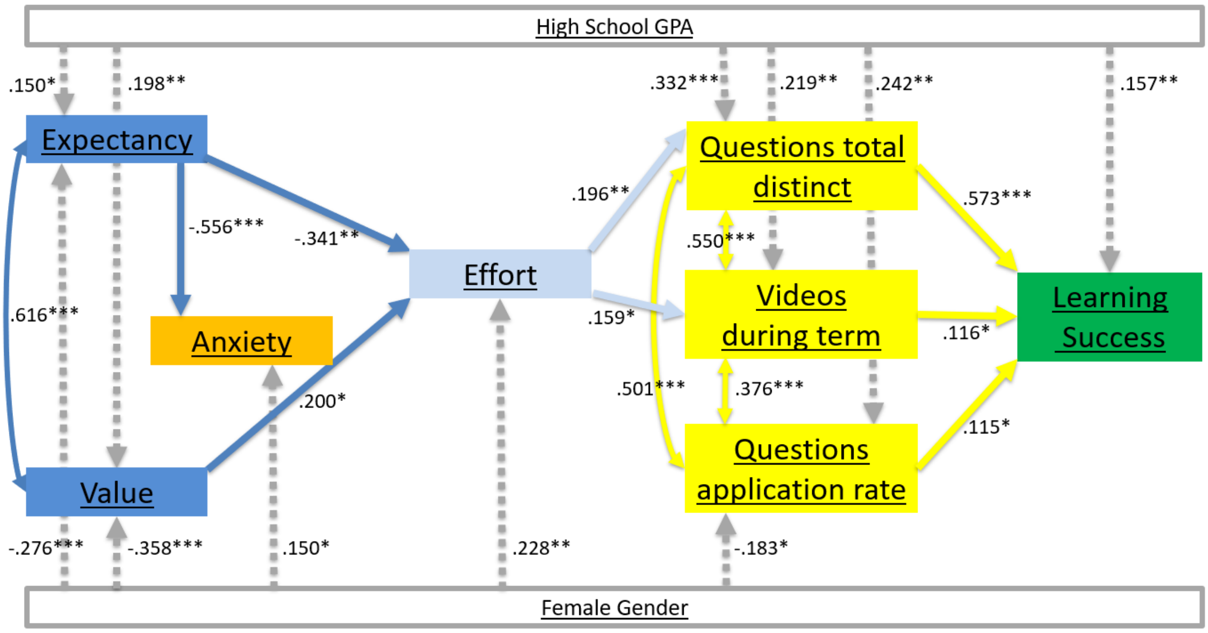

Figure 26: Integrated model with all significant links of expectancy, value, anxiety and effort to scaling, spacing, focus of learning and success; non-significant links are part of the estimated model, but are hidden in the graph for clarity.

The integrated model shows that scaling (Questions total distinct), spacing (Videos during term proportion), and focus of learning (Rate of application in questions total) all are linked to learning success, with the link of scaling clearly being the largest. Looking at the affective factors and the three dimensions of learning behavior, we find that expectancy and value, meditated by intended effort, are linked to scaling and spacing of learning. No such link can be found for the focus of learning. Furthermore, it is noticeable that anxiety has no significant link to the other elements of the model. Looking at high school GPA and female gender shows that both variables are significantly linked to large parts of the model.

6.5 Discussion

Summarizing the results, it can be stated that the research question "Why the motivated are the successful ones in our classes?" must be answered with more than one part of an answer. First, it can be concluded that high school GPA is positively linked to expectancy and value, as well as to learning success in the overall integrated model (Figure 25). In part, therefore, the correlation between expectancy or value and learning success is a spurious correlation that can be explained by shared links to high school GPA. The characteristics represented in a factor like high school GPA, such as but not limited to intelligence, thus in their totality are positively linked to motivation and success and are thus part of the explanation for their positive correlation.

However, the results also show that learning behavior does indeed contribute to explaining the relationship between motivation and success. It could be shown that scaling, spacing and focus of learning are linked to learning success. Furthermore, scaling and spacing are useful as an approach to explain the link of expectancy and value to learning success, since they are related to expectancy and value through the intended effort. In contrast, links to the focus of learning could not be found. Likewise, no link between anxiety and learning behavior could be identified. The latter is particularly interesting since anxiety correlates even higher with learning success than expectancy and value. However, the relationships of anxiety seem to be different in that at least not the investigated dimensions of learning behavior account for the stronger link.

Overall, then, it appears that three pillars provide explanation for the positive correlation between motivation and success in learning: As a first pillar, the correlation is spurious, as it can be linked to characteristics measured in high school GPA. As a second pillar, motivation is linked to an increase in the quantity of learning, which is linked to success. And as a third pillar, motivation is linked to spaced learning and the lack of delay of learning, which also is linked to more success.

Regarding Eccles and Wigfield´s (2020) Expectancy-Value Theory it can be stated that the postulated influence of expectancy and value on learning success can be a possible explanation for the links found. The importance of learning behavior as an explanation for this influence, as suggested by Eccles et al. (1983), also fits the results of our models well. While only based on the statistical examination of the empirical data presented here, it is not possible to conclude unidirectional causal effects of expectancy and value on learning behavior or success, the results found do nevertheless make a causal interpretation, as assumed in the EVT, seem conceivable. Thus, the case studied provides empirical evidence for EVT. In addition, EVT can be inspired by this study regarding the importance of learning behavior for the causal relationships in the model. Also, the role and rationale of students' intended effort should be critically examined in conjunction with the SATS-M (Ramirez et al., 2012).

With regard to Pekrun´s (2006) Control-Value Theory, only little support can be found in the case under study. On the one hand, the examined construct anxiety indeed shows a negative correlation to learning success, on the other hand, hardly any links of anxiety to learning behavior can be identified. One reason may be the high correlation of anxiety with expectancy, which make sense due to the conceptual proximity of the concepts. Another reason may be that links between anxiety and learning behavior occur on qualitative levels that could not be investigated in this study. At the same time, it is still surprising that students do not react to their existing anxiety by, for example, increasing their learning engagement. However, CVT is valuable, among other things, because it conceptualizes two main reasons for learning success: learning behavior and cognitive factors.

For further research, there would be great added value in replicating the conducted study under different circumstances. In particular, the course studied is a compulsory course for the students. It would be worth investigating whether the relationships mentioned would be similar in a voluntarily chosen course. In addition, there would be value in complementary studies that can analyze learning behavior in even finer detail, especially with respect to the quality of learning.

Regarding the limitations of this study, it should be noted that it involves data from only one course. In previous research, however, it has been shown that effects of attitudes and anxiety can be strongly dependent on the style of teaching (Bateiha et al., 2020; Herman & Kerby-Helm, 2022; Intepe & Shearman, 2020). Replications or extensions of the study are therefore needed.

In conclusion, it can be stated that the presented study finds empirical evidence that explains the correlation between expectation, value, and anxiety on the one hand and learning success on the other. Existing research that had already examined similar models through purely survey-based data is augmented with fine-grained digital behavioral data from an observation of a course actually delivered in the field. The research thus shows that expectancy, value, and anxiety are important factors of teaching and learning not only because they correlate with learning success in an abstract way, but because they are related to learning behaviors of students in a very concrete way. The research also serves as a reminder of how multi-dimensional learning behaviors are and how strongly they are linked to learning success. Thus, learners can increase their chance of success by adjusting their learning behaviors. Instructors can focus on their students' expectancy and value to promote such a behavioral change.

6.6 Concluding Remarks and Transition to Study 6

Study 5 thus shows, in relation to the main research question of this work, that the assumption is supported, at least for attitudes towards statistics. Variables of the quantity of learning as well as variables of the distribution of learning seem suitable for explaining at least part of the link between attitudes towards statistics and learning success. For research question 3b, the assumed model is therefore supported.

No such relationships can be found for statistics anxiety. This is probably mainly due to the strong links between attitudes towards statistics and statistics anxiety. Nevertheless, it remains that for two of the three levels of non-cognitive factors, no mediating role between these non-cognitive factors and learning success can be found for learning behavior.

However, the approach that remained at the end of study 4 and that was formulated in research question 2b remains: to look for connections between beliefs about statistics and attitudes towards statistics. Such a connection could be used to argue for the relevance of beliefs about statistics, although their indirect effect on learning success would only be very small. Study 6 therefore examines research question 2b of this thesis. The study initially chooses a quantitative approach that attempts to measure direct linear relationships between beliefs about statistics and attitudes towards statistics. In a second step, the connections are examined in a more nuanced way in qualitative interviews.

7 Study 6: How Students' Statistics Beliefs Influence their Attitudes: A Quantitative and a Qualitative Approach

This chapter originated as a manuscript co-authored with Kelly Findley, University of Illinois at Urbana-Champaign and Sebastian Hobert, University of Goettingen. The author of this dissertation is first author of the manuscript. The manuscript has been published as a peer-reviewed chapter in an edited volume. It is reproduced with permission from Springer Nature and the co-authors. The original publication can be found under

Berens, F., Findley, K. P., & Hobert, S. (2023). How Students´ Statistics Beliefs Influence their Attitudes. A Quantitative and a Qualitative Approach. In G. F. Burrill, L. de Olivera Souza, & E. Reston (Eds.), *Research on Reasoning with Data and Statistical Thinking: International Perspectives. Advances in Mathematics Education (pp. 303-316)*. Springer, Cham. https://doi.org/10.1007/978-3-031-29459-4_22

Abstract: Negative attitudes of students toward statistics in introductory courses are a widespread phenomenon. The reasons lie first and foremost in students' experiences with mathematics, but they are also more diverse than just that. In addition to known reasons, this paper proposes that one's organizing beliefs about the nature and applicability of statistics may explain one's attitudes toward statistics. This study builds on our previous work, from which we developed a statistics beliefs instrument that reflected four broad statistical conceptions found in our data and corroborated in the statistics education literature. For this next stage of work, we surveyed 471 students and interviewed 14 students studying the social sciences in a German university. Quantitative findings show slight correlations between students' beliefs toward statistics and their attitudes toward statistics. In particular, static, rules-based beliefs about statistics have negative impact on the students' attitudes, while more open, investigative beliefs on statistics have a positive impact on all dimensions of attitudes. Qualitative insights show that beliefs work primarily by emphasizing certain parts of students' attitudes and prior experiences, thus shifting their weighting in the composition of the overall attitude.

Keywords: mathematics attitudes, Mixed Methods, statistics attitudes, statistics beliefs, statistics conceptions

7.1 Introduction

Students' views about and interest in a discipline can greatly affect their own learning experiences and knowledge acquisition (Roesken, Hannula, & Pehkonen, 2011; van Griethuijsen et al., 2015). This relationship has been frequently highlighted in statistics education research. Not only is there a close connection between students' attitudes and their performance in statistics courses (Emmioğlu & Çapa Aydın, 2012; Finney & Schraw, 2003; Ramirez et al., 2012), but it is well established that positive attitudes should be a goal of its own for statistics education (Gal et al., 1997).

Statistics education researchers have argued that students' negative attitudes toward statistics may be the result of certain limiting beliefs about statistics (Finney & Schraw, 2003; Zieffler et al., 2008). However, the use of the term beliefs in the existing literature varies, reflecting the multi-faceted con-struct of beliefs (McLeod, 1992). Much of the existing literature that links stu-dents' learning of statistics to beliefs (including the research cited previously) focuses on "self-efficacy beliefs"—views that learners have about them-selves and their abilities (Finney & Schraw, 2003; Gal et al., 1997). But students may also have disciplinary beliefs that reflect how they see usefulness, structure, and purpose in disciplinary work (Justice et al., 2020; Wild et al., 2018).

Disciplinary beliefs play a critical role in students' motivation for learning (Kloosterman, 2002), how students view disciplinary advancement and knowledge construction (Muis, 2004; Tsai et al., 2011), and how students think about problem solving (P. Bell & Linn, 2002; Ozturk & Guven, 2016). Even with the same curriculum and the same teacher, students see tasks quite differently depending on the disciplinary beliefs and conceptions they bring to these contexts (Presmeg, 2002).

Unpacking and understanding students' disciplinary beliefs and conceptions has gained much attention in recent years as the field grapples with the growth of data science and the merging of several areas of study. In this chapter, we want to look more carefully at students' disciplinary beliefs about statistics. In addition, we will be relating their beliefs to their attitudes of statistics to see whether disciplinary beliefs may explain negative attitudes among new learners at the college level.

7.2 Background

7.2.1 Attitudes about Statistics

One's attitude toward a discipline refers to their disposition toward that discipline and directly stems from their experiences and perceptions of that discipline (Fishbein & Ajzen, 2009). Attitudes are more stable than emotions, which better reflect the in-the-moment responses that students may have in the learning process; they are also more dispositional than beliefs, which better describe students' perceptions of their own abilities, in addition to the cognitive propositions they have constructed (Gal et al., 1997; McLeod, 1992; Philipp, 2007).

Ramirez et al. (2012) reviewed a set of statistics attitudes instruments and identified six attitudinal subcategories that were reflected in the item constructions of these surveys. They described the connection between these subcategories in comprising one's statistics attitudes as follows:

"Before devoting the time and energy (Effort) to learn and do statistics, our model indicates that students evaluate their skills (Cognitive Competence) and the Difficulty of statistics and statistics tasks. They choose to expend Effort on statistics tasks and courses that they like (Affect) and are interested in doing (Interest) while they avoid others. They also consider how useful statistics is and will be in their lives (Value)" (Ramirez et al., 2012, p. 61).

Ramirez et al. identified the Survey on Attitudes toward Statistics (SATS-36) as best representing these core components of statistics attitudes (Schau et al., 1995; Schau, 2003).

While other authors (e.g., Tempelaar et al., 2007) find negative attitudes among their students, it is not easy to identify the reasons for that. One explanation might be students' previous experiences with mathematics and the role they perceive that mathematics may play in their statistics coursework (Dempster & McCorry, 2009; Nasser, 2004). Ramirez et al. (2012) also posit that other student characteristics, such as gender or race, as well as past experiences with statistics. These factors interplay with students' attitudes and in some ways predict their attitudinal responses. The relationship between these components can be mapped out, as shown in Figure 27.

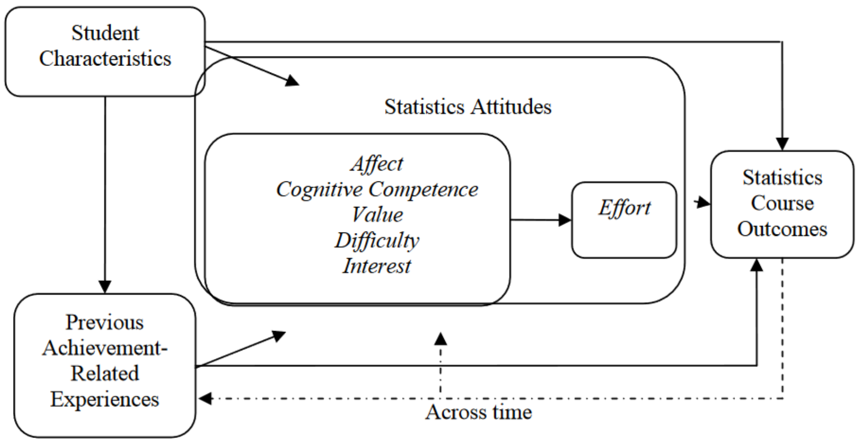

Figure 27: Students' Attitudes Toward Statistics Model (Ramirez et al., 2012).

In addition to these factors, however, there may be value to understanding how students' previous experiences and background may contribute to their attitudes about statistics. In particular, we ask how their beliefs about the discipline of statistics might explain certain attitudes. We take a look at this con-struct in more detail next.

7.2.2 Beliefs about Statistics

Beliefs is arguably a messier construct to define, in part because beliefs can vary widely in their construction and status. On one hand, a belief can be dis-positional and deeply embedded in one's affects or desires (e.g., "I believe I can complete this task," "I believe everything happens for a reason"); on the other hand, a belief may be thought of as a conviction held with a varying level of certainty, typically through some type of induction (e.g., "I believe it might rain based on the weather I see") (McLeod, 1992; Philipp, 2007; Southerland, Sinatra, & Matthews, 2001). Thus, the usefulness of this construct in research de-pends on a careful description by the authors of what it means in the context of their study. In our work, we are not specifically concerned with the affective dimensions of a belief, but rather beliefs as representing ideas and constructions. In our case, we are narrowing in on students' beliefs of the discipline of statistics, including how they think about its structure, its application, and the work of experts in the discipline.

Philipp´s (2007) use of the term conception may also be a helpful idea to at-tach to this particular focus on beliefs. According to Philipp, a conception is "a general notion or mental structure encompassing beliefs, meanings, concepts, propositions, rules, mental images, and preferences” (p. 259). Thus, students may have broad disciplinary conceptions that encompass their disciplinary beliefs and organize them into a framework for interacting with the discipline.

Several previous studies have directly investigated students’ conceptions of statistics and offered unique framings (Gordon, 2004; Justice et al., 2020; Reid & Petocz, 2002; Rolka & Bulmer, 2005). We found much alignment across the findings in these studies, including a clear novice conception, an expert conception, and two different intermediate conceptions. One of these inter-mediate conceptions was more theory and method centered, while the other was more data and interpretation centered. These category descriptions are highlighted in Table 11, along with the conception dimensions they align with from each study.

Table 11: Alignment between Previous Research and Statistical Conception Categories

Conception Categories	**Research Alignment**
Novice Conceptions	Gordon • Process • Mastery Justice et al. • “Paint-by-numbers” Reid and Petocz • Statistics is individual numeric activities • Statistics is using individual techniques • Statistics is a collection of techniques Rolka and Bulmer • Statistics as a collection of tools and procedures

Intermediate Conceptions (theory centered)	Gordon • Tool Justice et al. • Step-by-step Painting Class Reid and Petocz • Statistics is the analysis and interpretation of data Rolka and Bulmer • Statistics as tools with a context
Intermediate Conceptions (data centered)	Justice et al. • Realist Perspective Reid and Petocz • Statistics is a way of understanding real-life using different statistical models Rolka and Bulmer • Statistics as a means of understanding a complex world or as part of a unified picture with understanding data
Expert Conceptions	Gordon • Critical Thinking Justice et al. • Picasso Reid and Petocz

	• Statistics is an inclusive tool used to make sense of the world and develop personal meanings Rolka and Bulmer • Statistics as integrated into the world itself

Despite different data collection approaches, there are striking similarities in the conception categories described by all three of these papers. Novice conceptions recognize nothing particularly unique or clearly unifying about statistics (e.g., statistics as disconnected procedures). Intermediate conceptions see growing cohesion and connection to contextual application, either in the direction of its analytical affordances, or in the context of making sense of data more contextually. Expect conceptions add a clear emphasis on meaning-making, with statistics seen not simply as a tool, but also as a way to see and interact with the world differently.

We highlight these conceptions because we see the beliefs we are measuring in our study as indicative of broader conceptions that students have about statistics. By connecting certain belief patterns toward these higher-level conceptions, we hope to more clearly map the potential connection between beliefs and attitudes.

7.2.3 Research Question

This paper examines whether students' disciplinary beliefs play an important role in the development of statistical attitudes. It is therefore assumed that students have different ideas about the nature of statistics and different personal definitions of the term statistics. These ideas about what statistics means would sensibly play a role in the frame of reference over which negative attitudes are formed. Thus, we wish to answer this question: Is there an association between students' disciplinary beliefs as they relate to statistics and their attitudes toward statistics?

The approach of this paper is to use students' beliefs as explanatory variables. We hypothesize that the more static perspectives characterizing novice conceptions have a negative influence on attitudes, since this view grants students' little autonomy. Additionally, we expect conceptions that focus on statistics as a meaning-making activity to be characterized by higher student autonomy, thus being associated with more positive attitudes.

7.3 Methods

The data for this study were gathered in an introductory course on statistics for social scientists at the University of Goettingen. Participants come from nine different disciplines in the area of social science, most frequently political science or sociology. Prior to their statistics course students already took a course on qualitative and quantitative empirical methods in social sciences of 14 weeks with eight hours per week.

Data collection took place in connection to the enrolment of the course. Out of 707 students enrolled in the course 471 took part in the survey and completed the items to an extent that the survey could be evaluated. About 50% of the participating students report to be in their second semester of university studies, with an average over all of 4.1 semesters. 61% of the respondents report being female. These characteristics are typical for this course.

For the qualitative part of the study, a random sample of 90 of these students was invited. Of these, 14 students were willing to participate and got interviewed through semi-structured interviews. As part of a bigger interview the following questions were asked (in German) in a standardized manner to answer this paper's question:

1. What do you generally think about statistics?
2. In your studies, you are required to take statistics. What do you think about that?
3. Do you think statistics is useful? Why? Why not?
4. Do you think statistics is difficult compared to other courses you take? Why? Why not?
5. What do you think, how good will you be at statistics?
6. What would you say is statistics? How would you describe its nature? Or could you define it?

7. If you compare statistics to other courses you take, what is special about statistics?
8. When someone does statistical work, what do they do?
9. How do you think you can recognize a statistical expert? What distinguishes him or her as a person and what competencies does he or she have?

In addition, the interviewers asked follow-up questions about what was said.

The interviews were transcribed and pseudonymized. The transcripts were initially analyzed separately by the first author and by an independent person. In the analysis, in-vivo codes were first generated from the interviews related to students' attitudes, justifications for those attitudes, and/or students' disciplinary beliefs. These codes were then inductively grouped into themes. Themes and their occurrences were then coded by consensus in order to collectively conduct further analysis. Regarding students' disciplinary beliefs, consensus assigned one of the four conceptions to be the main conception. For some of the respondents, a second conception was assigned as a minor conception if codes of a second conception appeared, which however was not coded as often as the main conception.

7.4 Quantitative Instrumentation

As introduced earlier, the SATS 36 instrument, developed by Schau (2003) and widely established in statistics education, was used to survey student attitudes. In the SATS-36 instrument, Schau proposes a six-dimensional structure of statistics attitudes. The dimensions are presented as Affect (I will like statistics.), Cognitive Competence (I will find it difficult to understand statistical concepts.), Value (Statistics is irrelevant in my life.), Difficulty (Statistics is a complicated subject.), Interest (I am interested in learning statistics.), and Effort (I plan to work hard in my statistics course.).

For the theoretical model on statistics beliefs described above, a questionnaire was developed by the first two authors in order to identify different disciplinary beliefs students resonated with. Since this beliefs instrument is not yet published, we first provide some background about the structure of this instrument.

In addition to existing research precluding this project, we have carried out various qualitative preliminary work in our attempts to create an instrument that captures students’ disciplinary beliefs about statistics. To create this instrument, the first author collected data through focus group interviews with German students of the social sciences. Data from these interviews were used to describe various student conceptions of statistics. Further work with this concept was carried out with students of mathematics in the USA, so that the authors have developed a theoretical framework that represents students' belief in four categories. Data from both the focus groups and individual student interviews showed high alignment to the conception categories found in previous literature (Findley & Berens, 2020).

The four categories can be seen as a 2x2 matrix in which two data-driven (left column) and two theory-driven (right column) conceptions can be found, as well as two more application-oriented (top row) and two more pure (bottom row) conceptions of statistics. The first is a rules-based conception, which understands statistics as a collection of rules and procedures that, when ap-plied correctly, determine solutions to statistical problems. In this conception, however, application is understood only in the sense of computations or algorithmic statistics, not with respect to real-world problems. Data are not necessary either. Secondly, the confirmatory conception describes a picture of statistics in which theories and models can be verified or falsified by comparison with data. Reference to real-world problems arises through the tested theories or models, which can come from a wide variety of application fields. They are the starting point for statistics, which then compares data to evaluate the plausibility of the theories or models. The third, descriptive conception views statistics as a large set of tools that can be used to represent data to readers. As in the rules-based conception, statistics is seen here as a toolbox. Likewise, the distance of statistics to real-world problems is a common characteristic. A difference, however, lies in the starting point, which the descriptive conception sees in data. These data can be summarized and depicted objectively by statistics in order to reflect reality dispassionately. The fourth, investigative conception understands statistics as a process of data exploration in which an attempt is made to generate information for decision-making from data by means of a circular procedure. Here, data are again the starting point, but they do not only come from the real world, but it is the task of statistics to investigate this real world and to gain new, valuable insights about it.

Table 12: Disciplinary Conceptions of Beliefs Instrument

The Investigative Conception Exploring data with the goal of generating questions and gathering insights	**The Confirmation Conception** Testing theories and claims using formal methodology in order to come to a conclusion
The Descriptive Conception Reporting summaries and representations of data in order to share information clearly	**The Rules-based Conception** Following steps and procedures in order to find correct answers

For this purpose, nine items were developed for each of the four conceptions on statistics, which were based on the qualitative preliminary work. Three out of nine items were related to the nature of statistics in each conception, three related to the process of doing statistical work, and three described the characteristics of statistical experts and the skills that are most important. The first of the survey's items were tested using principal component analysis and confirmatory factor analysis. All analyses reveal a need for further work on the survey design. Nevertheless, the analyses indicate that the basics of the survey's ideas are reflected in the results. There-fore, the resulting data set can be used for further analyses, even if a revised version of the survey is used in the future.

7.5 Results

If we first look at the students' ratings broadly, we see that students provide somewhat high ratings for all four conceptions. Nevertheless, students are slightly more inclined towards a descriptive conception (mean of 5.06 out of 7 on a Likert scale), and less inclined towards an investigative conception (4.45). All in all, however, all four conceptions of statistics are relatively close (rules-based 4.89, confirmation 4.77). We also see that the standard deviations are relatively similar, even though here again the investigative conception has the lowest standard deviation. Nevertheless, the standard deviations and different quantiles indicate heterogeneity, which can be used to ex-plain different attitudes.

7.5.1 Quantitative Influence of Beliefs on Attitudes

To identify the influence of students´ conceptions on their attitudes, six multiple linear regressions are performed. The six dimensions of attitudes about statistics proposed by Schau act as the dependent variables. The four conceptions we identified to organize students' patterns in beliefs are the dependent variable to explain the attitudes. The results of regressions can be found in Table 13.

Table 13: SATS-36 explained by students´ conceptions in 6 linear regressions (one regression per column, p-values in brackets, all attitudes are coded in a way that high values represent positive attitudes)

	Affect	Competence	Value	Difficulty	Interest	Effort
Rules-based	-0.488 (0.000)	-0.410 (0.000)	-0.130 (0.094)	-0.306 (0.000)	-0.160 (0.116)	0.190 (0.042)
Confirmation	0.093 (0.344)	0.125 (0.197)	-0.012 (0.884)	0.082 (0.194)	-0.074 (0.506)	-0.110 (0.266)
Descriptive	0.039 (0.665)	0.070 (0.431)	0.161 (0.039)	-0.044 (0.446)	0.308 (0.003)	0.322 (0.001)
Investigative	0.232 (0.006)	0.187 (0.024)	0.354 (0.000)	0.089 (0.095)	0.618 (0.000)	0.178 (0.039)
Intercept	3.663	4.274	2.775	3.841	1.352	2.150
R^2	0.094	0.066	0.087	0.093	0.145	0.089

First and foremost, the effect of the investigative conception on attitudes is particularly remarkable. In all cases, it has a positive effect on students' attitudes, sometimes with effects that can be rated as very large and usually highly significant. So, the investigative conception appears to have a particularly strong and positive effect on students' attitudes.

In contrast, the rules-based conception has a negative effect on students' attitudes. Some of the effects are again quite high. Only the intended effort is positively influenced by this conception. Holding a rules-based conception, therefore, seems to be a big burden for a student decreasing the students´ affect and perceived competence while increasing the perceived difficulty and planned effort in learning.

Regarding the descriptive conception of statistics results give a more mixed picture. The descriptive conception has a positive effect on the personal interest in statistics and the perceived value of statistics. Both are related concepts focusing the usefulness of the application of statistics. The likewise positive effect of a descriptive conception on the planned effort in learning fits well with these findings. However, the pair of perceived difficulty of statistics and the perceived own competence is not related with a descriptive conception. Also, the general affect toward statistics can neither benefit nor is damaged by a descriptive conception.

For the confirmation conception no effects on students´ attitudes can be found.

Looking at the R^2 of the six models one can see that the students´ interest in statistics can best be predicted by the students´ conceptions. The other five R^2 are on the same level. It should be noted, however, that all R^2 are not particularly high, indeed quite low. The first thing to say about this is that disciplinary conceptions were never intended to be the strongest predictor of attitudes. They are only one factor in a multicausal model. Even with this in mind, however, the explanatory power is not very high. Looking into qualitative data will therefore provide an explanation of how beliefs influence attitudes and why this influence cannot be measured as high quantitatively here.

7.5.2 Insights from the Qualitative Interviews

Among the 14 interviewees three were classified as mainly holding a rules-based conception of statistics with two of them holding descriptive conceptions as a minor conception. Another three students mainly held a confirmation conception with all three other conceptions appearing as a minor once. Two students mainly expressed an investigative conception of statistics, one with a descriptive conception as a minor. The biggest group of interviewees were six students mainly holding a descriptive conception of statistics. Three of these also held rules-based conceptions as a minor, one holds investigative conceptions as a minor. Overall, then, it appears that the application-oriented beliefs of confirmation and investigative conceptions are relatively rare. Only two of the students only held application-oriented conceptions, another four combine them with rules-based or descriptive conceptions. The attitudes associated with these conceptions are shown in Table 14.

Table 14: Attitudes and conceptions of the 14 interviewees

Pseudonym	Main conception	Minor conception	Attitude
Daniel	rules-based		Negative
Emma	rules-based	descriptive	Rather negative
Nick	rules-based	descriptive	Rather positive
Alissa	descriptive		Neutral
Marc	descriptive		Rather positive
Chen	descriptive	rules-based	neutral
Sophie	descriptive	rules-based	positive
Ben	descriptive	rules-based	positive
Mia	descriptive	investigative	neutral
Sam	confirmation	rules-based	Negative
Selma	confirmation	descriptive	Rather positive
Alex	investigative	descriptive	Rather positive
Simon	confirmation	investigative	Rather positive
Hannah	investigative		Very positive

Regarding the reasons for the respondents' attitudes, many different arguments were coded. In particular, the interviewees talked about their previous experiences with mathematics, about the benefits of statistics for further studies, science and personal career perspectives, and about the societal role of statistics. Not every interviewee expressed a direct connection between his or her conception of statistics and these reasons, but for nine interviewees such connections could be coded. Daniel, Nick, Sophie, Ben, and Sam's attitude toward statistics argued directly from the assumption that statistics is mainly math and aims to calculate formulas and algorithms.

Daniel for example started his answer to the first question in the list of questions above saying “Well, I have to be honest, I am not a fan of statistics. Simply because I don't like math very much and statistics is really math for the most part…”. Later in the same answer, he said: “that´s statistics, that´s what you have to expect, it is formulas”. Similarly, for question 6, his answer starts “For me, it's really just math, so that's why it's/. I must also say that I really don't know whether statistics is something for me in this case, because it' s just formulas”.

Among students who put mathematics and a strong focus on formulas and their calculation in the center of the reasoning, all five held a major or minor rules-based conception. Only two of the remaining nine also held a rules-based conception. This connection, however, does not have to imply negative attitudes. Sophie and Ben reported that they had always liked mathematics and therefore look positively on statistics and formulas. It is striking, however, that interviewees with other conceptions do not draw this direct connection, almost identity, between mathematics and statistics. This means that the influence of previous experiences in mathematics on attitudes towards statistics was also smaller for them. Hannah, for example, reported very negative previous experiences in mathematics, but did not let these flow into her attitudes towards statistics at all. Thus, the influence of a rules-based conception of statistics did not seem to be a direct positive or negative influence on attitudes towards statistics here. Rather, the conception moderates which weighting certain other reasons, such as prior experiences with mathematics, get in the overall evaluation of one’s attitudes.

In the cases of Ben, Sophie and Marc, their interviews brought an element of a descriptive conception in direct connection with attitudes towards statistics. In particular, the attitudinal dimension of the value of statistics is addressed by describing statistics as an important tool to describe society and its change. Thereby, statistics is attributed a societal significance, which it unfolds through its use in societal debates.

Marc for example answered the first question saying, “I actually think that's quite important. Especially in the social science field. I think through statistics you can actually reflect how society is developing and also give a certain picture that is true to reality, so to speak, of how people think about certain topics.” In his answer to question seven, he added: “[in statistics] you try to give a depiction, but I think that has a certain credibility compared to other things, because you can actually create statistics with it. I think people are also more convinced when you say that this percentage of the population has done this and that…”. Once again, it is not the descriptive conception itself that leads to more positive or negative attitudes here, but the descriptive conception focuses on the factual description of (social) reality, which in turn is considered valuable.

Selma, Simon and Hannah also had a focus on the value of statistics in their interviews. They emphasized not so much the importance of statistics for social debates, but rather the use of statistics in science. They described statistics as valuable because it can be used to generate advances in knowledge in research. Simon reflected this idea well: “[It is] important in any case because it reflects data that is then analyzed that disproves or proves certain statements and you need this toolbox of statistics to understand that properly, which I would not have thought before, but that makes absolute sense to learn and grasp statistics” and added “if you have learned statistics, then you also understand whether it is correct and how one has researched. And that really makes sense. I think that I can use this well in the future” answering the third question. A look at the conceptions shows that Selma, Simon and Hannah were three of the five people with more application-oriented conceptions. The conceptions of confirmation or investigation put the use of statistics for research in the foreground and thus justified a high value of statistics. Once again, the conception of statistics thus acted as a moderator in the choice of the importance of other influences on attitudes toward statistics.

7.6 Conclusion

In summary, the quantitative results show some negative effects of the rules-based conception, as well as slightly positive effects of the descriptive conception and very positive effects of the investigative conception. Some of the effects are certainly not small, but they explain only a small part of the heterogeneity of students' attitudes toward statistics. The qualitative results provide an explanation for this. In the students' arguments, the conceptions of statistics do not appear as direct justifications. Rather, the conceptions moderate which other arguments play a prioritized role in the overall composition of attitudes. Thus, learners' conceptions of statistics play an important role in explaining learners' negative (and positive) attitudes. However, this occurs only slightly directly, but rather as a moderating factor of other variables.

The results of this study should be treated with caution since the results come from only one German university. Thus, it can be concluded from the existing investigation that it is valuable to further investigate the question of the influence of the beliefs on attitudes, especially if analyzed together with other relevant factors as for example attitudes toward mathematics or attitudes toward research. Therefore, further investigation with other student groups is needed. Given the great importance of negative attitudes toward statistics, this further work seems meaningful and promising.

7.7 Concluding Remarks and Transition to Study 7

Study 6 shows in its quantitative part that there is indeed a (weak) relationship between beliefs about statistics and attitudes towards statistics. This can be attributed in particular to the fact that rules-based beliefs have a negative effect and investigative beliefs have a positive effect on attitudes. Study 6 thus largely supports what Bond et al. (2012) also found qualitatively.

In its qualitative part, however, Study 6 also illustrates how these effects might work in detail. It shows, for example, that many students have different reasons for their attitudes towards statistics. Their beliefs about statistics than play a role in determining the weighting of these reasons. Study 6 thus provides an answer to research question 2b of this thesis.

This answer is by no means complete, however, as it does not fully reflect the overall situation of the learners and is only oriented towards the dimensions of the SATS (Schau, 2003) in the selection of the attitudes surveyed. Study 7 therefore investigates the question of the effects of beliefs about statistics on attitudes further in three in-depth case studies. It is important to note that Study 7 is only the starting point for a large qualitative panel study that will run for years, but for which data collection is still ongoing.

8 Study 7: Lois Lane, Superman, and Iron Man. How Perspectives of Statistics Relate to Students' Identities and Career Pursuits

This chapter originated as a manuscript co-authored with Kelly Findley, University of Illinois at Urbana-Champaign and Nicola Justice, Pacific Lutheran University. The author of this dissertation is third author of the manuscript. The manuscript has been published as a peer-reviewed paper in conference proceedings. It is reproduced with permission from the International Association on Statistics Education and the co-authors. The original publication can be found under

> Findley, K. P., Justice, N., & Berens, F. (2022). Lois Lane, Superman, and Iron Man. How Perspectives of Statistics Relate to Students' Identities and Career Pursuits. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap. Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina*. International Association for Statistical Education. https://doi.org/10.52041/iase.icots11.T3F1.

Abstract: Incoming university students' unique experiences with and perspectives of statistics may shape their learning trajectories, recruitment, and retention in the discipline. We completed qualitative case studies with three first-year statistics majors. Each student shared their views of "who" statistics is to them, how they imagined a career in statistics, and recalled their prior experiences and motivations for studying statistics. We contrast the students and remark on how their varying experiences relate to perspectives on their chosen field. Results varied in the perceived nature and content of statistics, and experiences engaging with rich data sets were associated with more substantive perspectives of the discipline. Future research will use follow-up interviews to examine the trajectories of students' changing perspectives, participation, disciplinary identities, and interests.

8.1 Introduction

The American Statistical Association's (ASA) Guidelines for Undergraduate Statistics Programs highlights several important practices for the next generation of statistics students, including increased experience in data science and computing, working with complex data, using diverse models and analytical approaches, and communicating insights and methods to diverse audiences (American Statistical Association Undergraduate Guidelines Workgroup, 2014). As students enter undergraduate university programs with varied experiences and perspectives on the nature of statistical work (e.g., Gordon, 2004; Justice et al., 2020), we see two reasons why it is important to understand student perspectives about the discipline. First, research conducted from constructivist theories of learning suggests that students learn best when they connect new content with their own experiences and conceptual frameworks (e.g., Bransford, Brown, & Cocking, 2000). Therefore, it is important for teachers of statistics to understand what experiences and perspectives students bring into their courses so that they can build on them. Second, early classroom experiences that help shape students' perspectives of the utility and creative nature of statistics may incline more students to explore careers in the discipline (National Center for Science and Engineering Statistics, 2021).

This study explores the statistical perspectives of three newly enrolled university students intending to major in statistics. We asked these novice statistics majors to describe their perspectives of what statistics looks like and to reflect upon their prior experiences with and initial interest in statistics. This is an initial report based on the first interviews of a longitudinal study aiming to understand the first-year experiences of statistics students and how the content of their first-year courses interacts with their perspectives.

8.2 Background

We use the term perspective to capture a broad swath of constructs, including conceptions a student may have constructed about a topic, as well as the student's perceptions, impressions, and loosely held associations with it. Perspectives involve what the student perceives in relation to who and where they are and has nuances of being less-developed and more amenable to change than fully-formed conceptions. It is important to note that students' perspectives about statistics are descriptive (e.g., statistics involves calculations), and therefore not the same as the students' attitudes towards statistics, which are by nature evaluative (e.g., statistics is difficult; stats is fun).

8.2.1 Research on Experts' and Students' Perspectives of Statistics

Most statisticians agree that the discipline of statistics involves using data to grapple with uncertainties involved with real-world problems. Davidian and Louis (2012) offer a definition of statistics used by the ASA: "statistics is the science of learning from data, and of measuring, controlling for, and communicating uncertainty" (p. 1). As Wild and Pfannkuch (1999) document, the process of doing statistics is much more complex than knowledge of methods or the ability to follow a step-wise analytical process. It is also the process of developing productive dispositions (e.g., curiosity, skepticism, perseverance) and demonstrating a deeper awareness of how to connect statistical principles to real data (e.g., transnumeration, acknowledging variation, reasoning with models). Statistics is dynamic, data-oriented, and innately interested in drawing meaning (DeVeaux & Velleman, 2008).

Several studies have examined how students conceptualize statistics, with two common example conceptions being statistics as a series of tests-and-procedures and statistics as a way of making meaning in the world (e.g., Bond et al., 2012; Gordon, 2004; Reid & Petocz, 2002; Rolka & Bulmer, 2005). Students' conceptions of statistics are often viewed hierarchically, with expert conceptions of statistics being more dynamic and focused on meaning.

Rather than a strict hierarchical structuring, Justice et al. (2020) and Findley and Berens (2020) examine students' conceptions in the form of two spectrums. In both frameworks, the authors offer one spectrum representing objective, static views of data against views that acknowledge variability and uncertainty. A second spectrum contrasts theory-centered, procedure-driven views against real-world-centered, data-driven views.

In this paper, we look more broadly at students' perspectives of statistics. We examine how three prospective statistics majors have found their way to the discipline and identify the broad disciplinary perspectives that have guided their motivation to pursue careers in statistics. We make connections between their statistical perspectives, their participation in the discipline thus far, and what these connections reveal about their developing disciplinary identities.

8.3 Methodology

In this exploratory research study, we conducted qualitative case studies (Creswell & Poth, 2016) to understand and relate the perspectives of incoming freshman majoring in statistics. Three students responded to our solicitation to interview, which was sent to all first-year intended statistics majors at a large Midwestern U.S. university. We make no attempt to generalize these results to a larger population; we see these three students as what Creswell and Poth (2016) describe as instrumental cases, designed to help understand the phenomenon of students entering and remaining in statistics programs.

Two of the authors led semi-structured interviews together for each student. Modeling loosely after the Draw a Scientist Test (Chambers, 1983), we asked, “Who is statistics?” and invited students to draw a picture that personifies statistics and explain their drawings. We also asked students to identify skills or character traits a successful statistician might need, their prior experiences with coursework or projects, and their motivations for studying statistics. The semi-structured format allowed for a conversational style where we could follow up on important ideas and encourage students to elaborate. Interview transcripts and students’ drawings formed the primary data sources for the study. Secondary data were the researchers’ memos recorded throughout data collection and analysis.

Our analysis was shaped by the theoretical lens of disciplinary appropriation (Levrini, Fantini, Tasquier, Pecori, & Levin, 2015) in the context of legitimate peripheral participation (Lave & Wenger, 1991). That is, we examined to what extent students see links between statistical work and their developing identities, with an eye toward how experiences may involve legitimate peripheral participation in statistics. With this perspective, the extent to which the discipline is perceived to be responsive to new ideas and allow for creativity will relate to the extent to which the students participate and believe themselves to belong as legitimate members or contributors.

The first round of analysis involved three statistics education researchers completing independent In Vivo coding. Next, the researchers met to discuss their codes—initially line-by-line and eventually with a more holistic approach—comparing themes and challenging each other’s interpretations. To continue to ground results in the data, direct quotes were re-examined to determine whether interpretations and claims were supported by data.

8.4 Results

All three participants depicted statisticians as different types of heroes, and we will use three to illustrate our results. (a) Lois Lane, the heroine in the superman saga, is an investigative reporter with a commitment to finding truth and sharing that truth with the world. She has no supernatural powers, but she is nonetheless a hero as she uncovers important stories. (b) Clark Kent has two contrasting identities: a friendly, neighborly business employee who aspires to a simple life; and Superman with out-of-this-world strength and knowledge. Whether he likes it or not, the world depends on Clark to use his powers. (c) Tony Stark is a brilliant, confident engineer-hero who uses his knowledge of science and technology to create powerful tools—most notably the Iron Man suit. Tony finds satisfaction in his innovative creations and finding new shiny things to play with.

To be clear: none of the students claimed to be superheroes themselves. These three fictional personalities emphasize certain powers and missional orientations that were key to understanding each student's perspective. Lois (named after Lois Lane) depicted a statistician who served as a data journalist—uncovering truth and making it known to her audience. Clara (named after Clark Kent) depicted a statistician with a double life: he aspires to simple, quiet home life often intruded upon by work demands when others depend on his valuable skillset. Tony (named after Tony Stark) painted a picture of a modern, "sexy," versatile, and popular data science. His statistician is a data engineer who uses machine learning and cutting-edge algorithms to do incredible things.

8.4.1 Lois Lane: Statistician as a Data Journalist

Lois is drawn to the discipline of statistics through opportunities she anticipates in the job market as well as her own personal success in her high school Advanced Placement (AP) statistics course. She likes that statistics is balanced between technical skills and soft skills; balanced between using math, but in a very applied and useful context; and balanced between working alone at a computer and working with people. She sees statistics as modern and relevant. In particular, she thinks that statistics is useful for helping businesses and other fields make effective decisions.

Lois personifies statistics as carrying a phone because she is "always keeping track of stuff" and communicating. In her perspective, there is a single underlying Truth that can be found in the data, and it is the job of the analyst to find that truth and communicate it to clients. She says of data, "I think it's like a sure-fire way to help a business grow because it's like it's proof [emphases added]." When recounting her experiences in her AP class, she seems to identify a threshold for evidence in determining Truth and value: "it was cool seeing statistical significance … how much evidence there needs to be for something to actually be relevant." She sees the statistician as a data journalist committed to sorting through the irrelevant information in order to find underlying Truth in the data generating process.

In her perspective, creativity comes with the challenge of communicating results to an audience. In that sense, the data journalist is a helper with two purposes: first, to serve as a consultant for which the input is the data and the output is the answers that reveal the Truth underlying the data. Secondly, to communicate those results in a way that the client can understand, make sense of, and implement in advantageous ways (improving business practices, gaining more sales, furthering science, etc.). In both purposes, the data journalist serves as an intermediary helper or bridge between the client and Truth that will help move the business or research forward.

With regard to analytic methods, Lois envisions there being a correct analytical approach to take, and that she has much to learn. These perspectives largely stemmed from an internship experience in which she recorded data to a spreadsheet and gained awareness of the larger process of statistics. She feels she did not know enough of the "data stuff" to explore results and play with the data. She frequently positioned herself as a novice with quotes like, "I'm pretty much a beginner …" and "I didn't get to work with [the data] at all because they knew I hadn't had much experience …" The primary skills that Lois sees herself as lacking are computing skills. She also views coding skills as a primary source of appeal on the job market, and the skill that distinguishes adept statisticians from novice learners.

Overall, Lois' perspectives on statistics relate to her view that there is an objective process to doing statistics. As a future statistician, it is her job to carry out these methods with an open mind and get that information out to others. Lois also views herself as a novice who is on the periphery—or perhaps even outside—of statistical practice. Contributing to these feelings are her self-expressed lack of higher-level content knowledge and coding skills. She is eager to learn more statistical methods and coding and to gain more experience with real data.

8.4.2 Clark (Clara) Kent: Statistician as a Data Superman

Clara recognizes that data analysts play a critical role in the business world. Much like Lois, Clara also found success in AP statistics, and chose statistics in part because it mixes different skill sets. She thinks of statisticians and data analysts as being good at math, but also getting to be creative and interactive. Clara's personification of statistics was a young, nerdy guy living in a big city who she characterized as a “workaholic.” At the same time, he is trying to balance a fast-paced environment at work with the desire for a simple, leisurely lifestyle. Clara describes statistics as someone who has a dog “against his better judgment,” and as someone who likes to stay at home on the weekend. Rather than having the most modern technology, Clara describes him as having an antique analog watch and carrying a briefcase with his hard copies of spreadsheets.

Clara thinks there may be an objective Truth underlying the claims that statisticians strive toward, but she emphasizes that it is hard to identify that Truth in statistics. Claims involving data are not only restricted by levels of confidence, but also clouded by biases and bad assumptions. She explains, “I was really shocked at how much of a gray area there is …” and, “you can pretty much manipulate data into however you want it to be … it definitely made me more skeptical of things that I see. They'll be like, oh, 99% and I'll be like, oh really?”

Although Lois had some experience collecting real data, Clara largely drew on her perceptions from AP statistics, as well as impressions she gathered from a relative who works with data analysts. Clara had few specific data experiences to ruminate on in her interview. This difference in previous experience with data seemed to explain a nuanced distinction in Lois’ and Clara’s perspectives: Lois views the role of a statistician as filtering out noise to identify patterns and report the facts (a transparent reporter). In contrast, Clara views data analysts as finding creative ways to say something relevant from data (a knowledgeable sage).

Clara recognizes a data analyst as using some type of skillset to do their work, but this skillset is still largely a black box for her. She did not bring up coding in her responses (except to say that her first-semester data science class involved coding). She did say that data analysts were good at math, and she frequently brought up communication. She emphasized that whatever it is they did, it was extremely valuable to companies and in high demand.

We see Clara’s views of the modern data analyst much like a dual-personality superhero. This person offers incredibly important skills at work, but whose exact work is difficult to understand. Additionally, Clara offers a picture of someone living two disconnected lives, saving the company during the week, and recharging as an ordinary person on the weekend. Clara’s disconnect between these two personalities might stem from her lack of authentic data experiences. Without more detail to contextualize statistical activities, she was not sure how to characterize the work of statistics. We see Clara still largely as an outsider who is peeking into the world of statistics based on some positive class experiences and anecdotal perspectives.

8.4.3 Tony Stark (Iron Man): Statistician as a Data Engineer

Like Lois and Clara, Tony also participated in AP Statistics in high school. For Tony, however, the AP Statistics experience was only the initiation point for his journey with statistics. His teacher recommended the website, Kaggle, as a way to continue learning. Tony took up this impulse with great interest and started various data projects. Tony frequently experienced roadblocks with algorithms he had not learned, but he found help and inspiration via internet searches and YouTube videos. Even though Tony describes himself as a beginner, he reports that he has been able to achieve high accuracy scores with his models. “That was just something that was very cool,” he comments. The topic relevancy of his projects, as well as the challenges he overcomes, motivate his pursuits.

Kaggle as the main encounter space with statistics has also shaped Tony's view of statistics, which he also referred to as “data science” seemingly interchangeably in the interview. He sees statistics as an attempt to model data in a way that allows for the most accurate predictions possible. Statistical expertise here is knowing many methods, algorithms, and codes that can be used to model the data. Doing data science mainly involves trying out the different possibilities and combining them in such a way that an optimal prediction is achieved.

In contrast to Lois and Clara, the answer to a statistical problem is not represented by a tangible truth, but rather as an optimal solution. The optimization criterion is the accuracy of the prediction of the model. The value of statistics comes more from the fact that predictions can be useful. This perspective seems a significant shift from the others by avoiding an epistemological consideration in the statistician’s role. To Tony, statistics (or perhaps more specifically, data science) is pragmatic.

He personified statistics as "the new girl on the block." She parties on weekends, wears cool T-shirts, communicates clearly, and stays active. At the same time, she is also intelligent and able to explain complicated issues in an understandable way. It is here that we see balance once again in the personification—statistics is cool and popular, but also smart and nerdy. This balance is important: to be truly useful to a business with statistics, the statistician must be able to present the meaning of her results to other stakeholders. “Communication is probably like number 1 … like story telling. If you're a good storyteller, kind of the ability to tell a story, umm, using data, like why things occurred in the past. What might that mean about the future.”

Overall, we see Tony as a data engineer. The task of the data scientist is then to design a “machine” that can do incredible tasks efficiently and find an optimized solution. In his descriptions, Tony reminds us of Marvel's Tony Stark (i.e., Iron Man). He sees statistics as agile, powerful, and modern. At the same time, both are engineers who build something from known basic materials that is more powerful than solutions before and capable of improving the world. With his experience working with real data and self-learning methods, Tony appears already to be on an inbound trajectory with respect to the discipline.

8.5 Conclusion

In one way or another, all three students depicted statistics personas with unique, relevant, and useful powers. Whether by uncovering the truth that few others could find (Lois), using mysteriously acquired knowledge to offer sage insights when the company was counting on him (Clara), or designing tools for prediction that few others have mastered (Tony), all three saw the work of a statistician as meaningful for making positive change.

All three liked that statistics uses math and saw themselves as being good at math, yet perceived statistics to be more exciting, interesting, or useful than pure, theoretical math. When asked about the content of their AP courses, all mentioned some form of statistical inference, with little emphasis on large projects with messy datasets. AP Statistics then was a critical entry point to the discipline, but one with little or no experiences working with more authentic data-based questions.

One major difference between our three cases was the extent to which they have engaged with open-ended projects or grappled hands-on with messy, or difficult data sets. Tony has spent hours tinkering with difficult models and trying out different approaches to improve their performance. Lois’s internship experience collecting data to answer a business question gave her some insights into the investigative cycle outside of classic textbook problems. Clara, on the other hand, shared that she had no experience working with larger data sets on more open-ended problems (in her course, students could either choose a project or take the AP exam). We hypothesize that this gap in her experience may explain her under-developed perspective of what a statistician actually does in their daily work, which was evident by her focusing more attention on her statistics person's lifestyle.

In addition, there was a striking difference in the value and role of computer coding. Although Lois and Tony recognized a need for coding skills, Clara's picture of statistics was stuck in a world without digital devices. Her statistician was somewhat proud of his ability to use his analog watch (which he thought was somewhat of a dying art), and he carried paper spreadsheets of data in his briefcase. In this sense, his powers were a sort of ancient wisdom that he had the key to uncover.

These differences may explain the varying connectedness of the identities of the statisticians presented with the interviewees themselves. Lois presented a modern, young, business-like statistician who was female (like herself). Lois seemed to identify with the statistician's organization, consultant role, and strong quantitative skills. Tony presented his image of statistics as a clever, sexy new-girl-on-the-block with brand name clothing. She was versatile, in hot demand, and full of new ideas. Clara's statistician was more dichotomous and disconnected: sage hero that everyone depends on by day; low-key, friendly homebody in the evenings. It was difficult to find ways that she might identify with the daytime superhero, and she seemed to identify only with his "simple life" outside of work watching movies and cooking food with neighbors. We asked Clara: what genre of movies does he watch? Her answer? Superhero movies!

8.6 Future Research

We hope to follow up with these three students again. Second-round interviews would create an opportunity to check in on their first-year experiences and explore how their evolving disciplinary identities might relate to their ongoing engagement in legitimate participation in the discipline. We hypothesize that the identities and trajectories observed in our interviews with regard to experiences working with real data or acknowledging the role of computing will be more pronounced as students gain more experience in their major courses. Additionally, a larger-scale study involving more students could detect trends in what early experiences in statistics are most critical to helping students find their identity in the discipline and how to attract and retain a more diverse group of students who will enter the story of becoming their own kinds of statistical heroes.

8.7 Concluding Remarks and Transition to Study 8

The approach of Study 7 is somewhat broader than would have been necessary for this thesis alone. Background to this study is a project together with Nicola Justice, Kelly Findley and Christopher Kinson, in which students who have chosen to major in statistics are followed in a qualitative panel design with comprehensive interviews from the beginning of their studies until their transition to the workplace. The aim is to investigate what beliefs about statistics the students bring with them to their studies (evaluated following the approach of Bulmer & Rolka, 2005), what experiences these are based on, how these beliefs develop over the course of their studies and what influences they have on the processes of appropriating statistics (Levrini et al., 2015), on the feeling of belonging to statistics (Gravett & Ajjawi, 2022), and on learning and study decisions.

The material used for Study 7 comes from the first three students recruited for this larger project from their interviews at the beginning of their studies. For these participants, the fourth interviews could just be conducted after completing their third year of study in May 2024.

This thesis is helped by the findings from this project that, on the one hand, beliefs about statistics are influenced by the students' experience of working with real data and having carried out projects on their own responsibility. At the same time, these beliefs have an impact on the extent to which such experiences are sought in the further course of study and which courses the students choose. Thus, in addition to general attitudes towards statistics, as shown in Study 6, beliefs also influence more specific attitudes towards certain forms of statistics learning and, as a consequence, attitudes towards various differently characterized courses in statistics studies. Other studies that have since emerged in the project also show that beliefs play a role in how, how comprehensively and how quickly students adopt statistics for themselves and make it part of their own identity (Berens, Findley, Justice, & Kinson, 2023; Findley, Justice, Kinson, & Berens, 2025). Study 7 and the associated studies thus supplement research question 2b with the finding that beliefs about statistics can have an effect not only on attitudes towards statistics in the narrower sense but also on other attitudes towards studying statistics. At the same time, it must be taken into account that these findings were obtained in a setting that differs greatly from the setting of all other studies in this thesis. The students interviewed were not studying social sciences, a subject that many perceive as being not statistically oriented, but were majoring in statistics. Furthermore, the data comes from the USA and thus from the American university and study culture. These differences raise the question of whether similar findings would be found among students of social sciences in Göttingen. However, they do prove that such complex structures of relationships between beliefs and attitudes are plausible.

This only leaves research question 4 of this thesis completely open, which asks about the development of attitudes during the course and the mechanisms behind these developments. Study 6 in particular, but also Study 7, already indicate that attitudes can be more multidimensional than the SATS measures (Schau, 2003) suggest. This finding is now taken up in order to investigate in a qualitative interview panel over the course of a semester how attitudes towards statistics develop and which facets of attitudes are the cause and driving force behind these developments.

9 Study 8: What Changes Students´ Attitudes?. A Qualitative Panel Study on How and Why Attitudes Toward An Introductory Statistics Course Change

This chapter has been published as a peer-reviewed paper in conference proceedings. It is reproduced with permission from the International Association on Statistics Education. The original publication can be found under

Berens, F. (2022). What Changes Students´ Attitudes?. A Qualitative Panel Study on How and Why Attitudes Toward An Introductory Statistics Course Change. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap. Empowering & Educating Today's Learners in Statistics. Proceedings of the 11th International Conference on Teaching Statistics (ICOTS11 2022), Rosario, Argentina*. International Association for Statistical Education. https://doi.org/10.52041/iase.icots11.T3G1.

Abstract: Change in attitudes toward statistics during introductory courses has been shown to have a major impact on learning success. While attitudes pre- and post-introductory courses have therefore been widely studied, there is a lack of research on how attitudes evolve during the course. Only Kerby and Wroughton (2017) include a measurement at midterm and find some roughly v-shaped progressions of attitudes. The present study shows a different development in a qualitative panel study with six students and four interviews per student. The main reason for changes in attitudes is identified as the current course content, which changes how mathematical, difficult, interesting, and valuable statistics is currently perceived to be. The findings show the reciprocal relationship between topic-related emotions and domain-related attitudes.

9.1 Introduction

In 1997, Gal et al. (1997) were the first to argue that attitudes are relevant for statistics education. They outline the importance of attitudes for learning in a statistics course, the influence on an individual's willingness to choose more in-depth courses or programs and the effect on a person's propensity to make practical use of statistics outside of education in everyday life or at work.

Since that time, there has been much development. A milestone was certainly the development of the Survey of Attitudes Toward Statistics (SATS) (Schau et al., 1995; Schau, 2003), which measures six dimensions. Affect captures (positive) feelings toward statistics, difficulty asks about the perceived general/abstract difficulty of statistics, and cognitive competence captures perceptions of one's own ability with regard to statistics. The value of statistics is the perceived relevance and usefulness of statistics. In addition, the interest in the subject of statistics is surveyed as well as the amount of effort planned or already invested in the statistics course.

Using this instrument, researchers have argued that students' attitudes towards statistics do effect on their success in statistics courses (e.g., Abbiati et al., 2021; Chiesi & Primi, 2010; Nasser, 2004). Research also suggests that changes in attitudes predict success even better than attitudes themselves (Abbiati et al., 2021; Whitaker et al., 2022). In addition to the instrument, SATS dimensions of attitudes became a fixed canon for further, also qualitative, research (e.g., Songsore & White, 2018).

Despite these findings, little attention has been paid to changes in attitudes during a statistics course. While some studies have attempted to elicit attitudes through changes in course design (e.g., Bateiha et al., 2020; Herman & Kerby-Helm, 2022), (qualitative) studies on the reasons for changes in attitudes are lacking. This research aims to fill this gap by conducting a qualitative panel study following six students through an introductory statistics course.

9.2 Background

9.2.1 The Development of Attitudes in a Statistics Course

Even in the development of the SATS instruments, it was clear that attitudes are not very time-stable facets of the learning process, but rather changeable personality traits (Schau, 2003). Both the original grounding of the SATS in Eccles et al.'s Expectancy Value Theory (1983) and later refinements of the theoretical grounding (esp. Ramirez et al., 2012) not only suggest the possibility of change over time, but even call for it as an educational goal.

Consequently, many studies have looked at changes in attitudes toward statistics. Their results, however, show a mixed picture: While, for example, Abbiati et al. (2021) and Herman and Kerby-Helm (2022) see positive trends in introductory statistics students' attitudes over time, Schau and Emmioğlu (2012), for example, find a negative trend. Meanwhile, Bateiha et al. (2020) and Cladera et al. (2021) find no significant change at all.

What several of these studies have in common is the finding of high rates of attitude change among students. However, these changes (partially) cancel each other out in the averaging across students, as improvements and deteriorations occur in similar amounts (Abbiati et al., 2021; Whitaker et al., 2022). From this, it can be concluded that not only do course-related factors have an influence on the development of attitudes, but also that there exist individual differences between students. This hypothesis is also supported by C. Xu et al. (2020). They found that the instructor has a high impact on the development of students´ attitudes, and yet they find very heterogeneous development of attitudes within courses. Thus, the explanation for attitude changes must be at least partly at the individual level.

9.2.2 The Role of Attitudes and Their Development for Learning Statistics

While research on individual reasons for changing attitudes is still largely open, the importance of these changes has been documented. It has been known for some time that attitudes toward statistics at the beginning of a course have predictive value for course success (Chiesi & Primi, 2010; Nasser, 2004). However, more recent studies show that changes during the course are even more predictive of exam success (Abbiati et al., 2021; Whitaker et al., 2022).

This finding further emphasizes the importance of changes in attitudes toward statistics. Yet, only Kerby and Wroughton (2017) take a closer look at the development of students' attitudes toward statistics. In a panel study that adds a midterm survey between the pre and post measures of attitudes, they find different results for the different dimensions assessed by SATS. While affect improves slightly in the first half of the course and then remains roughly constant, value and interest describe a v-shaped progression. No significant changes can be found in the other dimensions. For the v-shaped progression in particular, Kerby and Wroughton (2017) assume a reality shock at the beginning of the semester that causes students' value and interest to drop, but students then recover later in the course. However, the researchers cannot provide evidence or deeper insights into the reasoning behind these changes.

9.3 Methodology

As previous research demonstrates a high importance of changes in attitudes towards statistics, this study aims to uncover individual patterns of reasons for individual changes. In order to comprehensively represent and analyze individual reasons and their contexts, a qualitative study design is used. Since the changes are to be traced in the process and not only retrospectively, the study is conducted as a qualitative panel.

To implement the qualitative panel, 70 students from an introductory course in statistics for social scientists at a large German university were invited for interviews. The course is obligatory in the programs of these students. In the first interviews directly before the course began, 14 students participated. Subsequent interviews took place after weeks five and nine of the course, as well as after completion of the 14-week course. During the first five weeks, students were introduced to descriptive statistics (e.g., central tendencies, dispersion, graphical representations of distributions). In the following four weeks, the theoretical introduction to inferential statistics including confidence intervals and hypothesis testing followed. In the last course section, bivariate statistics was introduced (Pearson's correlation, bivariate linear regression, relationships of two binary variables).

All interviews were held as guided, semi-structured interviews. The interviews always began with the question "What do you think about statistics in general?" and continued with several consistent questions about the six dimensions of attitudes outlined in the SATS. This was followed by questions about the students´ learning behavior and satisfaction with the statistics course. Finally, interviewers were free to ask any open follow-up questions on topics that arose.

Six students participated in all four interviews. However, analyses of all 14 first interviews failed to establish that the attitudes of the six remaining students differed substantially from those of the eight students who dropped out of the study. The six students with the full four measurement points are therefore considered the six cases of this study.

The analysis of the 24 interviews follows Saldaña´s (2003) manual for the analysis of longitudinal qualitative data. Saldana therein proposes a three-step procedure: First, the interviews are to be coded per case as closely to the material as possible using in vivo coding and paraphrasing and then summarized into case reports. In the second step, the material is coded across individuals and time points for themes and patterns. In this step, it is also possible to use quantifying codes to make comparisons between persons and time points more tractable. In a third step, the case reports and the themes and patterns are synoptically combined to form an overall picture.

In the present work, in vivo coding and paraphrasing according to Miles, Huberman, and Saldaña (2020) was used for the first step. This resulted in case reports, which are briefly summarized in the first part of the results.

The themes and patterns were examined using a qualitative content analysis creating categories, definitions, anchor examples, and coding rules for attitudes and reasons for attitudes following Mayring (2015). First, deductively obtained categories were used. For this purpose, the six dimensions of the SATS became first categories. Based on Songsore and White (2018), subcategories of value were assumed to be value for future schooling, value for a future career, and value for everyday life. Inductively from the material, the subcategories value for society and value for science were added. In addition, the category of effort was split into two subcategories, general effort and effort relative to other courses. A category on prior experience with mathematics was added.

In order to identify patterns in addition to themes, two additional dimensions of categories were developed inductively from the material. First, it is easily seen that all statements can be positive or negative to varying degrees. Therefore, the sentiment of a statement was coded as very positive, rather positive, neutral, rather negative, or very negative following the example of a Likert scale. Secondly, the follow-up interviews show that students repeatedly evaluated specific content topics instead of statistics in general. This was therefore coded as the frame of reference, which can be either topic-focused or statistics in general.

Overall, each quote in the material was coded in all three dimensions, namely theme, sentiment and frame of reference.

9.4 Results

9.4.1 Case Reports of the Six Interviewed Students

Sophie, pseudonym for the first interviewee, starts her statistics course with positive attitudes towards statistics. She attributes a high value to statistics both for her future schooling and for everyday life. She sees mathematics as a central element of statistics, but this does not bother Sophie because she has "always liked math…and never had big problems with it." Therefore, she does not consider statistics to be difficult, but expects to put in a lot of effort in the course. After five weeks in the course, it is confirmed for Sophie that she is doing well with the math in the course. She has also become even more aware of the social importance of statistics, evaluating it as "very useful, as you can also see now in election times." Her attitude therefore improves to the highest sentiment. After nine weeks of the course, Sophie describes her attitude as unchanged overall and "still finds it very useful to learn statistics." However, she compares topics with each other and reports that she found the most recently taught introduction to "inferential statistics more interesting than the topic before," while she considers descriptive statistics more valuable. Even after the end of the course, her attitude remains very positive, and the value of statistics in particular continues to dominate this view as she finds it "very useful to learn statistics…and a helpful different way of thinking about social questions."

Chen starts the course with an overall neutral attitude toward statistics. He considers statistics to be important "to understand social contexts" and for science, but not so much for himself personally. Chen also considers statistics to be mathematical. Since he "never was good at calculations," he considers statistics to be rather difficult. This fear was then confirmed for Chen after the first five weeks of the course. Since his problems with mathematics are even greater than expected, his attitude slightly tips toward the negative. However, Chen continues to see the general value as a good reason for statistics, as he sees "a legitimacy in statistics, but it is not necessarily fun." After nine weeks, Chen primarily discusses inferential statistics, which he finds "not very interesting, but sometimes useful…and really used in science…and in papers I read." His overall attitude therefore returns to a medium level and even becomes slightly positive at the end of the course because "statistics is an area you can make clear using examples…from scientific papers."

The difficulty was also no longer considered to be so high as Chen saw he "was able to pass the exam."

Daniel enters the course with negative attitudes. He equates statistics with mathematics and "never liked nor was good at math." He also states that he does "not need statistics so much because I rather want to take the qualitative research path." After five weeks of the course, Daniel's view has not really changed. He explains that statistics "is necessary in general and to understand sociology…, but not really necessary for me." After nine weeks, Daniel's attitudes towards statistics are at an all-time low. He "doesn´t like it…often left early…hardly understood half of it…and didn´t listen." As a result, Daniel drops out of the course after eleven weeks. In the interview after the course, Daniel reports that he still doesn't like statistics, but now realizes that statistics has value. He understands "why you have to learn it. Simply so that you understand how to deal with quantitative data." Daniel therefore wants to start the course next year with more commitment.

Emma starts with rather negative attitudes. The main reason is mathematics, which leads Emma to consider statistics as "very difficult." At the same time, however, Emma "find(s) it interesting." After five weeks, this view has been reaffirmed for Emma. She perceives statistics as "still rather difficult" but finds it "not so uninteresting" as "maybe you need it later in your job, when you have to analyze something or so." After nine weeks, Emma's attitude has changed more. Overall, it falls slightly into the negative again. On the level of individual dimensions, her interest has increased further because she now also sees a high value in statistics, which is justified by the topic of inference. At the same time, however, she finds this topic more difficult—perhaps too difficult. She feels overwhelmed, "it´s over my head … it's just all a bit much for me." She retains this feeling until she drops out of the course a few days before the exam. However, in the last interview, she still expresses a high value of statistics, calls it "relevant and interesting" and discusses how bivariate statistics was even more interesting and valuable than the content before.

Selma starts with rather positive attitudes. She finds statistics "valuable and actually interesting. It can be complicated…, but it is useful and interesting to somehow give information about a population." After five weeks, there is some slightly positive attitude development, as she now perceives the value of statistics (as well as its difficulty) higher than before. After nine weeks, because of inferential statistics, Selma is "definitely less motivated. Before, the topics were probably a bit less complicated and easier to comprehend." Nevertheless, she describes statistics as very valuable. The last part of the course, however, Selma likes more, because she finds bivariate statistics less difficult than previous topics. Selma likes statistics now and finds it very useful: "In general … statistics is actually a great thing, you can definitely use it, especially in the field of politics, when it comes to elections…It's definitely helpful…but there are certain difficulties."

Alex starts with positive attitudes. Statistics is described as interesting and "quite relevant for me...because I need it to do research and to be able to evaluate it." Alex does not consider statistics to be difficult. After five weeks of the course, positive attitudes even increase. "By now, I think it's great. I am completely thrilled!" Especially because of the topic of the first weeks and because of experiences in another course, statistics is now described as outstandingly important. Alex still does not consider statistics to be difficult. After nine weeks, Alex has experienced a slight drop in positive attitudes. He still finds statistics important as "it is necessary for what I want to do later on," but the topic of inference is very difficult for him. His motivation suffers, and he "didn´t do as much as before." By the end of the course, however, this development recovers because the rest of the course was no longer as difficult for him. He continued to find the course "interesting … and definitely needed for scientific work."

9.4.2 Themes and Patterns Across the Interviews

Looking at the categories assigned during the qualitative content analysis, the first thing to note is that the category value for future schooling is by far the most frequent. For Emma, this category only appears in her last interview, in which she sums up the course after dropping out. In the total of 20 interviews of the other five students, the category occurs at least once in every interview and is almost always positively evaluated. Only Daniel initially makes negative statements and revises his opinion in the last interview. It is also evident that about half of the statements in the follow-up interviews attribute the value for future schooling to specific content. Especially in the third interviews this can be observed for the introduction to inferential statistics. The value of statistics for future schooling thus appears overall as an important component of attitudes, which, apart from the very negative Daniel, also keeps the more reserved students on track.

Other categories of value are also coded frequently. The value for everyday life occurs more frequently for Sophie and Selma and only once for Chen and Daniel. However, these categories do not represent only positive attitudes but are a reason for the overall negative attitudes in Chen and Daniel. The same applies to the value for a future career. For Selma this is a noticeably frequent theme that changes in parallel with her overall attitude. Also important to her is the value for society, which she evaluates more consistently positively, which keeps her on track. Sophie and Chen also argue positively for the value of statistics for society to some extent. The value for science drives Alex's positive attitudes, in particular, but also can be seen in Selma’s and Chen's attitudes.

The second most frequently coded category after value was difficulty. Often the categories are related to prior experience with mathematics. Whereas each person’s coded statements for mathematics are sentimentally similar, each person’s sentiment towards difficulty is subject to slightly greater fluctuations. This can be explained, among other things, by the fact that difficulty is often attributed to specific topics. Half of the interviewees (Emma, Selma and Alex) emphasize the particular difficulty of inferential statistics. Previous experience with mathematics, on the other hand, played no role in these third interviews. Personal cognitive competence is also often associated with difficulty. However, the interviewees do not distinguish between the general difficulty of statistics content and personal difficulty with learning the content but use the expressions practically synonymously.

Codes for the category of interest can also be found in all six students. Although Emma and Alex talk about interest in all interviews, Selma only talks about it in the first interview. Sophie, Chen, and Daniel each talk about their interest in the second and third interviews. It is noticeable that in these interviews in particular, interest is discussed in relation to the specific topic, although with very mixed evaluations of both topics.

Affect is coded least often overall. Sophie and Emma express nothing in terms of affect, Selma and Alex do so once each and Chen twice. Only in Daniel's case are there several codes for this category. No codes were evident in his first interview, but thereafter, affect is very pronounced and in direct context to the overall assessment of his attitudes.

Effort in some ways represents a special dimension, as it relates more to the course than to statistics itself. Nevertheless, it does seem to affect students' attitudes as a whole, as can be seen from the many codes in this category. It is striking that the comments on effort in the first interviews refer exclusively to the general effort, while in the later interviews comparisons to other courses are drawn more and more frequently.

9.5 Conclusion

This study of six students in an introductory statistics course complements existing findings on attitude development. Although simple pre and post comparisons may not show significant changes, there are more dramatic attitude changes throughout the course progression related to specific topics. Although Daniel, Emma, and Alex ended the course with the overall attitude they had at the beginning of the course and Sophie, Chen, and Selma experience only slight improvements, stability with the first interview can only be found in a total of four out of 18 follow-up interviews. In the course of the interviews, it becomes apparent that the second interviews are coded more positively than the first, the third often more negatively, and the fourth again somewhat more positively.

With regard to the reasons for these attitudes and their changes, it should first be noted that the value of statistics for the students´ future schooling is a frequent and stable cornerstone. In the context of a compulsory course, a baseline level of attitudes only builds up from this. The heterogeneity of attitudes among the six interviewees is mainly related to the perceived value for everyday life and to the difficulty of statistics, the latter mostly in connection with previous experience with mathematics.

It is also evident that statistics is not perceived as a uniform block, but that students relate their evaluations to specific topics. It is noticeable that the introduction to inferential statistics is perceived to be both more valuable and more difficult. Depending on the previous perceived difficulty, a tipping point towards dropping out is often reached then, or attitudes rise again as value and interest increase.

For instructors, this can mean that they can act proactively to positively affect their students´ attitudes, particularly in the area of the value of statistics. The value for the students´ future schooling is generally shared but should be kept present in order to have a motivational impact. The value for everyday life and society seems less clear but is closely related to the overall evaluation in the explanations. Here, instructors can adjust instruction by providing convincing examples. Researchers should be concerned that inferential statistics continues to be perceived as very difficult, despite the fact that mathematics is not seen as a problem here. New approaches to good subject matter didactics would be helpful here.

At the same time, however, it must be noted that the six students interviewed have unique individual backgrounds and developments. Their opinions and explanations are therefore not necessarily representative, and certainly not for other, differently organized courses. Nevertheless, they can point out sites of fracture, which are worth reflecting on both for teachers and researchers.

9.6 Concluding Remarks and Transition to Study 9

With Study 8, all the research questions of this thesis are now adressed, so that final discussions could be held and conclusions drawn. However, another study was conducted unplanned, although it neither provides a direct answer to one of the four research questions nor prepares such an answer. Rather, study 9 arose due to the sudden emergence of the Covid-19 pandemic while working on this thesis and was intended as a spontaneous reaction to the uncertainty that arose at the time regarding academic teaching under the conditions of emergency remote teaching. The study compares the first digital behavioral data collected with the system from study 2 in the summer semester 2019 with the behavioral data from the summer semester 2020, which was changed by the pandemic. Even if this comparison of traditional teaching with emergency remote learning does not directly address a research question of this thesis, the results provide an interesting outlook on the susceptibility of the investigated relationships between non-cognitive factors and learning. To a certain extent, Study 9 thus provides a frame that helps to interpret the results obtained.

10 Study 9: Learning during COVID-19: (Too) Isolated yet Successful

This chapter originated as a manuscript co-authored with Sebastian Hobert, University of Goettingen. The two authors share first authorship of the manuscript. The manuscript is currently under review at a peer-reviewed journal. It is used and pre-printed in this dissertation with the kind permission of the publisher and the co-author.

> Berens, F., & Hobert, S. (Under reviewa). Learning during COVID-19: (Too) Isolated yet Successful.

Abstract: In a world changed by the COVID-19 pandemic, in 2020 universities had to completely rethink and immediately transform their teaching to a fully online setting. As a result, students had to learn from home and organize their learning by themselves. In a natural experiment, data about learning processes indicate that students learned more engaged and achieved higher learning success in this new situation compared to the traditional learning process. However, the students experienced a three-fold isolation: (1) physical isolation, (2) social isolation, and (3) learning isolation, which resulted in a stressful learning experience. In conclusion, these affective challenges indicate that this exceptional learning setting should not be normalized, even though positive outcomes were achieved during COVID-19. Beyond the situation in the pandemic, it can be deduced that students who learn at a distance need additional support to not (only) learn in isolation.

Keywords: COVID-19 pandemic, Distance education and online learning, Emergency Remote Teaching, Learning Analytics, Natural Experiment, Post-secondary education

10.1 Introduction

Starting early 2020, the COVID-19 pandemic altered the familiar trajectories of many aspects of life. Its effects were not only felt on the health front but also in people's social lives through lockdowns, social distancing, and other safety measures (Desvars-Larrive et al., 2020; World Health Organization, 2020). To prevent the further spread of SARS-CoV-2 , also education has undergone a fundamental change, especially at the beginning of the pandemic (Litao Sun et al., 2020; Weeden & Cornwell, 2020). Schoolchildren were taught from home, often for the first time in their lives, which was a major challenge for parents, teachers, and learners as well (Saavedra, 2020). The pandemic also transformed learning at universities worldwide (Chatziralli et al., 2021); Liu, 2020). In-class lectures, seminars, practice units, and labs had previously dominated university learning, which have than been closed. Within a short space of time, the entire teaching of universities shifted to a digital learning space. Learning in person was replaced by synchronous and asynchronous digital teaching formats, organized via video recordings, real-time video conferences, forum discussions, chats, and many more (Johnson et al., 2020; Wayne et al., 2020).

This rapid transformation poses a fundamental question about learning: How did the pandemic situation affect students' learning?

On the one hand, answering this question provides insights into the pandemic situation and its influence on learning at universities and on students as individuals. On the other hand, viewing the pandemic as a natural experiment can help to understand the effects of students having to learn at a distance and/or completely digitally involuntarily or at least contrary to their actual preference. After all, the pandemic situation was characterized by the fact that students who had opted to study in person on campus now had to learn entirely online and at a distance. And while we know that there are structural differences between students who choose to study on-campus and those who choose to study remotely (e.g., DeVaney, 2010; Gundlach et al., 2015), there is no complete separation between these groups, as external circumstances also lead students to study remotely who would not otherwise have chosen to do so (Fidalgo et al., 2020). Analyzing the pandemic as a natural experiment helps to identify possible risks for these students. In addition, the analysis can be a warning where future prophets prophesize the vision of a completely digital university and the superfluousness of campuses.

10.2 Background

10.2.1 <u>Online Learning</u>

The use of digital tools and the delivery of teaching online may first lead one to believe that teaching and learning during Covid-19 follows patterns that have long been studied in the field of online learning. There, findings emerged that can actually provide guidance for teaching and learning during Covid-19 (Bragg et al., 2021). For example, it has been shown that proactive communication by instructors is especially important in online teaching in order to cultivate a relationship with the learners and thus ensure learning success (Beege et al., 2022; Flanigan et al., 2022).

However, before evidence on how best to conduct teaching and learning during Covid-19 can be explored, it is necessary to outline how (and how successful) teaching and learning took place during COVID-19. One step in this direction is research examining courses that were offered in parallel both online and face-to-face, most recently referred to as dual mode teaching (Soesmanto & Bonner, 2019). Simple comparisons of the two modes, however, have not yet yielded clear results. Some studies found no differences in learning behavior and success (Shotwell & Apigian, 2015; Soesmanto & Bonner, 2019), while in other studies online students performed better (Pei & Wu, 2019) or worse (Gundlach et al., 2015) than face-to-face students. One reason for these differences may be that in general, and also in the studies mentioned, assignment to online or face-to-face teaching does not happen randomly, but is usually freely chosen by the learners. This leads to the fact that the learners differ considerably at the beginning of the courses (DeVaney, 2010; McPartlan et al., 2021). However, if learners who choose a course online differ from learners who attend in presence, it becomes obvious that the view on online learning is also not directly comparable to the situation during Covid-19. After all, during the pandemic situation neither teachers nor learners could choose, but all had to teach and learn online. Moreover, since this digital teaching had to be done at very short notice and unprepared, the term Emergency Remote Teaching became established.

10.2.2 Emergency Remote Teaching

Already early in 2020, the term Emergency Remote Teaching (ERT) developed for the digitally prepared and online delivered teaching during the pandemic-related lockdowns and restrictions (Bozkurt & Sharma, 2020). ERT differs from other online formats in at least three respects. First, teachers and learners alike have not freely chosen this format, are inexperienced in it, and have not been trained for that format. Second, there was little time to prepare for this format and to adapt previously prepared content and materials. Third, online learning is embedded in a life situation that is also substantially changed by the pandemic and its social consequences. Thus, learners are not only learning at a distance for a particular course, but are severely hindered in continuing their previous daily lives as a whole (Bond et al., 2021; Bozkurt & Sharma, 2020). The consequences of Emergency Remote

Teaching for learning must therefore be described and studied as a novel phenomenon (Tang et al., 2021).

Therefore, very early on, first studies appeared that asked students about their experience during Emergency Remote Teaching via surveys or interviews. The results were that ERT had a negative effect on motivation and perceived sense of belonging, while it increased stress and anxiety (Ezra et al., 2021; Petillion & McNeil, 2020). In addition, some students reported that concerns outside of learning, for example about health or economic development, also negatively impacted their learning experience (Alvarez, 2020). The only positive aspect sometimes seen was increased flexibility (Matarirano et al., 2021). But students also note that increased flexibility paradoxically also poses significant challenges to learning (Rahiem, 2020). As a consequence, the comparison of satisfaction between the new Emergency Remote Teaching and before usually results in a negative evaluation, for example in Knudsen's study (2020). Aldhahi et al. (2022) add that satisfaction is particularly strongly dependent on personal time management skills. Oinas et al. (2022) broaden the view to other self-regulatory abilities and also find high effects on learning behavior and satisfaction.

Moreover, dissatisfaction with the learning process is not the only negative outcome for the students. In various studies, students also report that they perceive their own learning success in Emergency Remote Teaching to be lower (Sharma et al., 2021; Shin & Hickey, 2021). Interestingly, however, Bawa (2020) and Iglesias-Pradas et al. (2021) do not report any negative effects in their studies on the actually measured learning success. Bawa (2020) does not find significant differences between ERT and before (whereby within the examined sample ERT even performed slightly better). Iglesias-Pradas et al. (2021) compare ERT and teaching before in even 43 courses, finding significantly higher learning success in ERT. For both mentioned studies, however, it should be noted that the comparison of learning success before ERT with learning success during ERT cannot be made completely without problems, as it remains unclear to what extent the measurement instruments of success were comparable and to what extent teaching changed beyond ERT in terms of personnel and pedagogy.

Moreover, it remains open how learning behavior has changed beyond the necessary modifications of ERT, and which of these are possible explanations for not finding a decline in success. Thus, an important gap remains that will be filled answering the four research questions of this study:

- RQ1: What is the impact of ERT on students' learning behaviors?
- RQ2: What is the impact of ERT on students' learning success?
- RQ3: Can explanations for the learning success be derived?
- RQ4: How can students' personal experience be related to the above?

These questions will be answered using a natural field experiment with ERT being the treatment in a course otherwise kept as comparable as possible. Learning behavior is examined using learning analytics, learning success is measured via the official exam, and personal experience is collected via a survey.

10.2.3 Learning Analytics as an Approach in Educational Research

New research areas of technology-enhanced learning, particularly learning analytics and educational data mining, can address these questions in a special way. Educational data mining aims to analyze large learning-related datasets with statistical and data science methods in order to understand learning (Chatti et al., 2012; Kavitha & Raj, 2017). Learning analytics with a comparable approach aims to understand individual learning success to improve learning processes (Chatti et al., 2012; Gašević et al., 2022; Long & Siemens, 2011). Both approaches observe learning non-invasively and use data generated in the learning process, for example, using e-learning systems. In this way, learning can be observed not only in limited physical spaces and over short periods of time, but also in real learning processes that may last for months (Schumacher & Ifenthaler, 2021; Zhang et al., 2018). Prerequisite for the application of such approaches is the existence of large amounts of learning-related data, which requires a technical infrastructure whose process data provide this insight into learning (Avella et al., 2016; Gray et al., 2022). The digital behavioral data stored by, for example, a learning management system can then be counted to operationalize them into meaningful variables of learning behavior. These reflect learning without having to rely on self-reports of questionable data quality.

10.3 Methods

10.3.1 Setting of the Experiment

The reported natural field experiment was conducted at a large public university in Germany. Teaching takes place in lecture periods of 14 weeks with additional two to three weeks for examinations after that. Attending lectures, tutorial sessions, or office hours is not a prerequisite to attend or pass an examination. Thus, the entire teaching is intended to support the students' learning processes. Students may decide on their own whether to make use of these offers or to learn self-educated.

The conducted study aims to investigate the learning behavior of students in a course on introductory statistics. The course targets students from many different undergraduate study programs related to social sciences and covers basic statistical concepts and intends to build a basic understanding of statistical principles. The course takes place every summer semester, i.e., every year from mid-April to mid-July. Each year approximately 700 – 800 students enroll in the course's learning management system and approx. 30 to 40 % of them register for the examination.

The course consists of five elements of teaching. This is a typical setting that exists in other courses as well. The five elements are:

1. The lecture takes place weekly from April to July in the extent of 90 minutes. In 2019, the lecture took place on Tuesdays starting at 04:15 pm.
2. Tutorial sessions of 90 minutes in smaller subgroups take place weekly (ideally 20-30 students actively participate). In 2019, Thursdays eight (identical) sessions were offered: two tutorials at 8:15 am, four tutorials at 12:15 pm, and two tutorials at 04:15 pm. Due to a bank holiday in Germany (in both 2019 and 2020), the number of tutorial sessions is reduced to twelve within the first 13 weeks, resulting in a total of twelve tutorial weeks.
3. Exercises provided by the lecturer to be used by the students to deepen the content and prepare for the examination. In the used e-learning app, students get access to about 20 quizzes a week for this purpose, which add up to 249 quizzes during the semester.

4. At the end of the lecture period, students were given the opportunity to self-assess their state of learning using a gamified quiz. Additionally, instead of classic tutorial sessions, the teaching assistants repeated the course content in three 90-minute presentations. Further, the students were provided with 152 additional quizzes for self-study using the e-learning app.
5. The examination takes place as an electronic assessment in the university's examination center. The examination is intended to be solved in 90 minutes.

10.3.2 Technical Infrastructure Used for the Study

In spring 2019, the authors introduced a novel e-learning app in the course (Hobert & Berens, 2024). This app supports students in as many parts of learning as possible. Whereas the previous learning management system mainly focused on management tasks (i.e., sharing documents, managing participants, and making announcements), the novel e-learning app extended these management tasks by providing actual e-learning tools as formative quizzes with automated feedback, an audience response system to collect instant, anonymous feedback, access to a glossary for looking up definitions and important terms used in the lecture, and a video player to provide recordings via the e-learning app.

Besides these functionalities provided to students, the e-learning app also integrates learning analytics functionalities. Learning analytics was incorporated into the system by design and offers opportunities for collecting data about the students' learning activities. For instance, data about system usage can be collected if the students opted in.

10.3.3 Intervention Induced by the Pandemic Situation

Shortly after the decision of the lockdown, the authors designed the natural field experiment. As the authors started researching the learning behavior of the students participating in the course in the summer term 2019 to evaluate the e-learning app (Hobert & Berens, 2024), the planned continuation in 2020 only needed to be adjusted slightly due to the COVID-19 situation. Due to the required transformation of teaching in 2020, the second observation period (summer semester 2020) incidentally became a treatment group.

To provide the students in 2020 an adequate learning environment during the COVID-19 situation, the authors decided to keep the overall structure of the courses by offering all five components described above. For each component, it was analyzed how it could be provided best to support the students. Whereas the components 3 and 5 were offered identically like before, adjustments need to be made for components 1, 2, and 4.

As the university recommended switching large-scale lectures with more than 100 students to asynchronous video recordings, the lecturer provided the lecture content via the e-learning app described above. Luckily, the lectures were recorded in 2019, which enabled the lecturer to post-process and uploaded the recordings. This ensured that the students in 2020 got access to exactly the same lecture contents compared to 2019. Only the presentation changed from synchronous in-class teaching in 2019 to asynchronous video recordings in 2020. In 2020, the video recordings were made available weekly following 2019‘s schedule. The videos were available to the students until the examination.

As the tutorial sessions are held in smaller groups, they could be provided synchronously as real-time video conferences. Thus, they were held in a synchronous setting like in 2019, and only the format changed from in-class sessions to real-time video conferences. If the students could not attend the synchronous video conference, they could watch a recording of the tutorial sessions afterward, which was only done by few students (less than 10% of tutorial participations). For organizational reasons, the tutorials also had to be postponed in the weekday and now took place on Mondays between 10 am and 4 pm. Since students in 2019 made great use of the opportunity to ask the tutors individual subject-related questions after the tutorials, and this was not possible in the same way in 2020, an open office hour of the tutors was offered on a weekly basis. The offerings of the examination preparation were not changed, but they also had to be shifted from in-class settings to video conferences.

These changes from in-class teaching to online teaching are to be considered the intervention of the natural field experiment, while all other factors (including the e-learning app) were kept as stable as possible. In particular, all 401 quizzes provided remained exactly the same. Also for the examination, the same instrument was used in 2019 and 2020 as it had not been made accessible after its use in 2019.

10.3.4 Operationalization and Data

To measure the students' learning behavior, data generated by the e-learning app was used. The e-learning app generates a pseudonymized log entry for each individual action taken by a student who opted-in to participate in research activities. The log entries encompass the type of action, a timestamp, and some (technical) metadata.

For each student, the following operationalization was used in this study based on data gathered using the e-learning app:

- **Completed quizzes**: number of quizzes solved by a student including all repetitions.
- **Distinct completed quizzes**: number of distinct quizzes solved by a student excluding all repetitions.
- **Success rate of completed quizzes**: percentage of correctly completed quizzes.
- **Success rate of distinct completed quizzes**: percentage of distinct quizzes completed correctly at least once.
- **Number of participated tutorials**: number of tutorial sessions a student attended. In 2020, this number was increased by the number of students watching the recording of the tutorial.
- **Rate of active tutorial participations**: number of tutorial sessions a student actively took part in. In contrast to above active participation was counted only if a student participated in audience response activities.
- **Number of watched videos** (only 2020): number of videos a student started to watch including all repetitions.
- **Number of distinct watched videos** (only 2020): number of videos a student started to watch excluding all repetitions.
- **Number of active lecture participations** (only in 2019): number of lectures a student actively took part in measured by participation in the audience response questions.

10.3.5 Additional Information on the Conducted Survey at the End of the Digital Semester

During the semester in 2020, conversations with students, teaching assistants, and other lecturers revealed that the transformation to digital teaching induced by COVID-19 could be a threat to the students' wellbeing. For instance, increased stress, insecurity, and anxiety were discussed as possible effects. To analyze whether these effects are also relevant for the students in the digital experimental group, the authors decided to approach these affective layers of the transformed digital learning utilizing a survey. This survey was designed by the authors and made be available via the e-learning app. A total of 97 students responded to this survey, which corresponds to a response rate of over 90% among students present in the video conference the survey was advertised in and approx. 13.5 % of all students enrolled in the course.

In addition to questions to evaluate the course and the e-learning app, six items concerning the transformation of teaching were asked: The six statements to be evaluated were (originally posed in German): 1. I could spend less time on university issues than in other semesters due to (corona-related) care duties. 2. It was very burdensome for me to work so much at home instead of being able to leave the house and meet people at the university. 3. It was harder for me to motivate myself during this semester than it was in traditional semesters. 4. I had more problems with my self-organization this semester. 5. This semester, I benefited from more freedom in the organization of my studies. 6. The digital studies have increased my level of stress compared to a traditional semester.

10.4 Results

In the pre-pandemic semester in 2019, 710 students were registered for the course whereas in 2020, 722 students signed up for the digital course. In 2019, seven students chose not to give permission for research activities on their data; in 2020, only one student did so. Therefore, this study operates with n1=703 students for 2019 and n2=721 students for 2020. For each of these 1424 students, extensive data collected with the e-learning app on their individual learning behavior is available (e.g., on practicing with the provided 401 distinct formative quizzes). In total, this provides a data basis of 10,957,536 learning-related log entries.

10.4.1 Students´ Learning Engagement

The comparison of the two experimental groups was based on the students' learning behavior. In 2019, students completed a total of 161,373 quizzes (including repetitions) using the 401 unique quizzes compared to 201,893 quizzes completed in the digital semester. Adjusted for the number of participants, this means that in 2020, almost 22% more quizzes were completed per person as can be seen in Table 15. Also, by focusing on distinct quizzes, the difference rises to 22.9%. A contrary pattern was observed for the tutorials. In 2019, 30.5% of the students on average took part in tutorials with 19.8% active participants. In 2020, only 22.4% of students participated in tutorials, with 13.1% active participants. In 2019, an average of 28.2% of the students attended the lectures. In the digital semester, an average of 44.7% of the students watched lecture recordings weekly. This advantage of the digital semester is also reflected in the fact that students watched videos again on average 1.1 times. Relevant differences in heterogeneity between the two years cannot be found in any of the variables considered. The first observations about the learning behavior thus show that the learning engagement of the students was higher in almost every aspect in the Emergency Remote Teaching. Only the participation in the last remaining live element, the tutorials, is lower than before. This interesting finding will be deepened in the following by looking at the distribution of learning.

Table 15: Summary statistics of learning behavior in both years.

	2019			2020		
Indicator	**Mean**	**Median**	**Std.Dev.**	**Mean**	**Median**	**Std.Dev.**
Completed quizzes	229.5	81.0	314.9	280.0	101.0	360.8
Distinct completed quizzes	90.5	42.0	107.6	111.3	48.0	129.8
Success rate of completed quizzes	0.485	0.474	0.159	0.454	0.461	0.179
Success rate of distinct completed quizzes	0.825	0.871	0.168	0.743	0.821	0.249
Number of participated tutorials	2.17	1.0	2.74	2.68	1.0	3.61
Rate of active tutorial participations	0.102	0.032	0.167	0.131	0.0	0.235
Number of watched videos	-	-	-	78.3	57.0	74.4
Number of distinct watched videos	-	-	-	36.8	25.0	35.2
Number of active lecture participations	2.80	2.0	2.96	-	-	-

10.4.2 Distribution of Students´ Learning

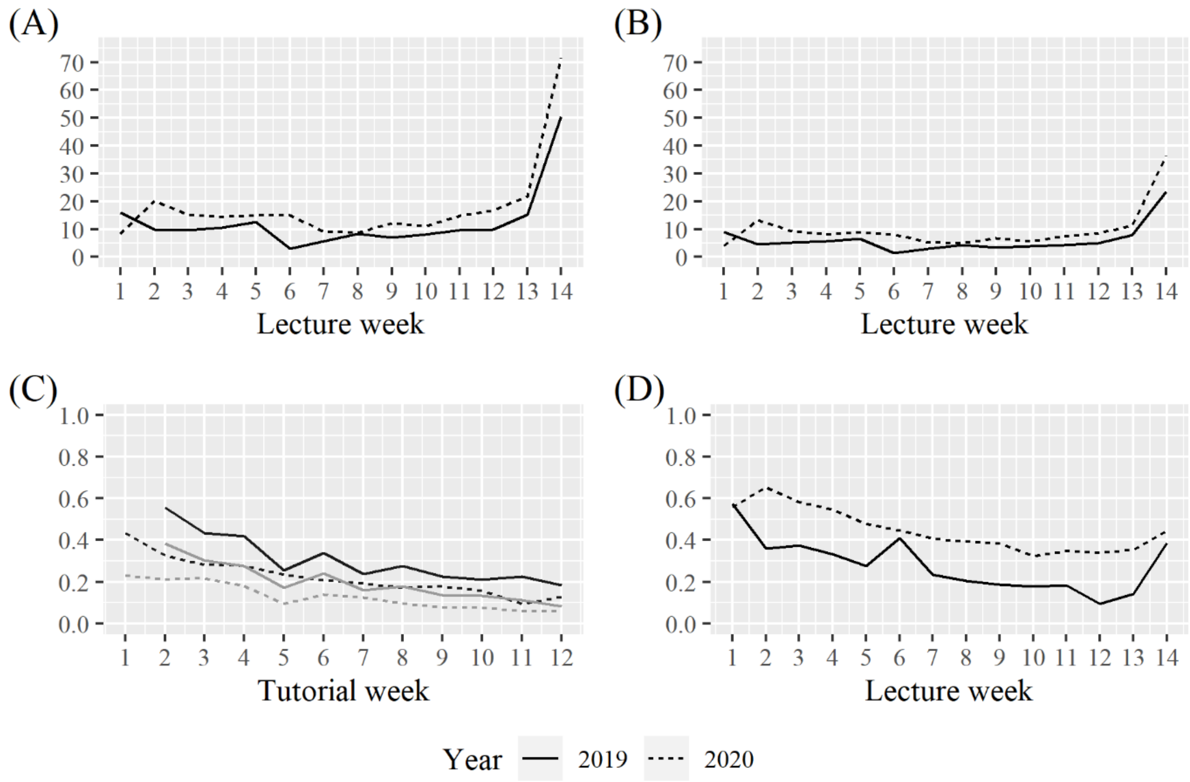

Figure 28: Development of the learning behavior during the lecture periods on a weekly basis. (A) Average number of completed quizzes (including repetitions) per student. (B) Average number of completed distinct quizzes (without repetitions) per student. (C) Rate of participation in tutorials and in grey rate of active participation. (D) Rate of participants in lectures (2019)/watching videos (2020).

Figure 28 shows that the identified differences between both years regarding learning engagement can be observed over the entire period after week one. Students in the digital semester completed more quizzes in each single week and worked on more distinct quizzes. Additionally, more students watched lecture videos throughout the entire semester, compared with in-class lectures in 2019. Only participation and activity in tutorials were consistently lower in the digital semester. Although this finding is surprising, it does not appear to be an isolated case. Jeffery and Bauer (2020) in their research also find that participation in the remaining live events is declining and students are withdrawing more than necessary into self-study.

Figure 29 supports the impression gained by illustrating that students in the digital semester learned in a much more distributed manner over the course of the day and the week and were much more engaged in learning on the weekends. The heatmaps visualize the distribution of learning activities across weekdays and times of the day.

In 2019, the highest activity can be recognized on Tuesdays and Thursdays during synchronous learning activities (lecture resp. tutorial sessions). Additionally, increased activity can be detected on days before the synchronous learning activities took place (Monday and Wednesday). This can be explained by preparatory activities of the students (e.g., repeating the contents of the previous week on Mondays or solving quizzes as preparation for the tutorial sessions on Wednesdays). It is clearly visible that the learning activity followes the schedule of the course and that learning decreases rapidly during the weekend.

In 2020, the learning activity is more homogeneously distributed as there are no clear maximum points. This can be explained by the absence of synchronous lectures and lower participation in tutorial sessions. In contrast to 2019, there is no substantial drop in the learning activities at the weekend – even though the activity decreases slightly.

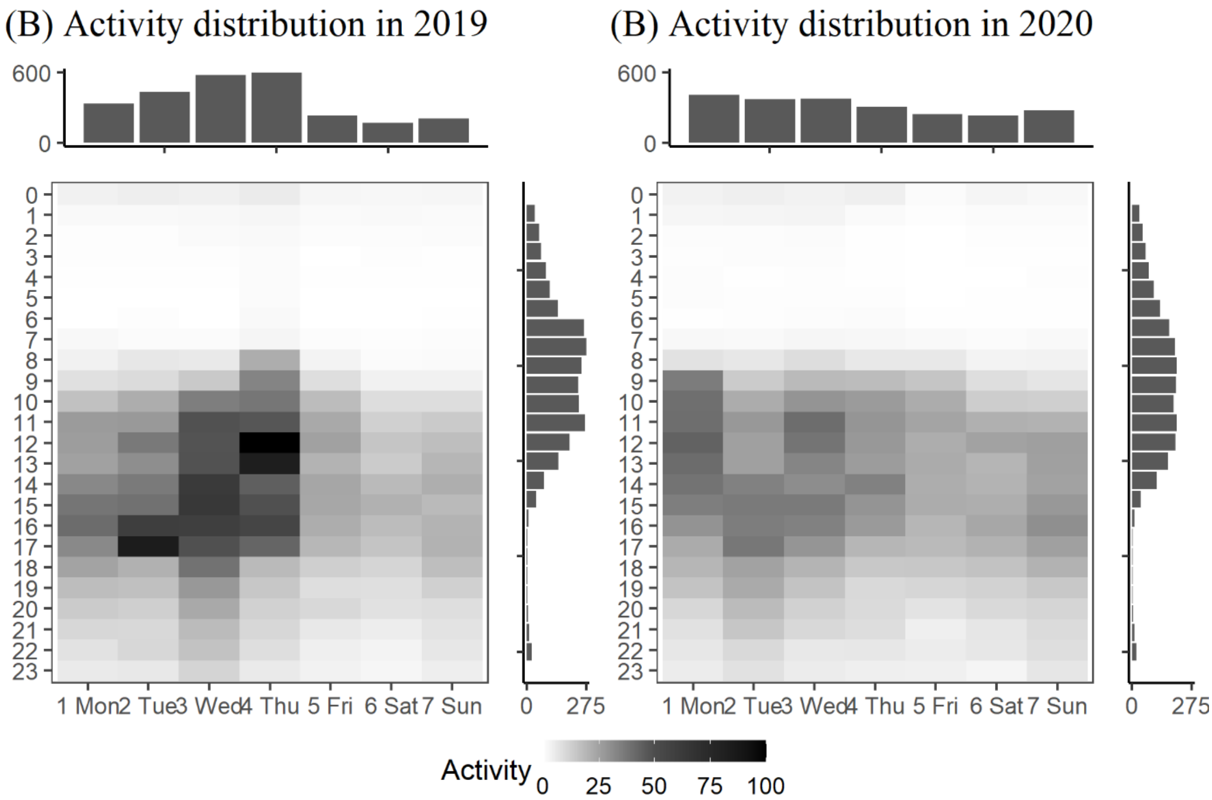

Figure 29: Heatmap visualizing the distribution of learning activities across weekdays and time.

Regarding learning behavior, we draw the overall conclusion that students are substantially more engaged in the self-regulated asynchronous phases of learning than in the traditional year and spread their learning more across the week. However, they are less active in synchronous tutorial sessions. Considering that substantially less synchronous teaching was offered in the digital semester, this finding is counterintuitive. It was assumed that students would attend as much synchronous teaching and social interaction as possible instead of learning individually. Instead, however, students apparently encounter physical isolation by isolating themselves even further than necessary by learning in isolation.

10.4.3 Students´ Learning Success

Given the generally higher asynchronous learning engagement in the digital semester, the consequences on students' learning success must be studied. Three indicators are important for students' learning success. The first indicator is the courage to take the examination. Since all course offerings are voluntary and there is no fixed schedule for the exam either, students do not have to take the exam but can postpone taking it without any negative consequences if they feel better this way. As a result, in the year of traditional teaching, only 34.0% of the 703 students registered for the course took the examination, which must generally be considered normal for the course. In the digital semester, 37.6% of the 721 students took the examination. Consequently, the digital semester scores slightly better in this indicator. The second indicator is the examination pass rate. Seventy-seven per cent of the students passed the examination in the traditional semester whereas the pass rate in the digital semester is higher at 83.0%. The third indicator are the average points received on the examination, which point in the same direction. In the traditional semester, an average of 54.5% of the points were achieved, whereas in the digital semester it is 60.5%. This means that students in the digital semester performed better in all three success indicators. Thus, in this sense, the Emergency Remote Teaching can be considered successful. Thus, the counterintuitive finding of Bawa (2020) and Iglesias-Pradas et al. (2021) is confirmed in our data as well. At the same time, the increased learning engagement provides an approach to explain the increased learning success in ERT. In order to examine this more closely, in the following we examine to what extent the learning behavior variables can be used as an explanation for the learning success.

10.4.4 Explanations of the Learning Success

Table 16: Pearson correlation matrix between learning behavior and score in the final examination in both years.

Year	Completed quizzes	Distinct completed quizzes	Success rate of completed quizzes	Success rate of distinct completed quizzes	Rate of partici-pated tutorials	Rate of active tutorial partici-pations	Watched videos	Rate of distinct watched videos	Rate of active lecture partici-pations
2020	0.522	0.729	0.501	0.409	0.469	0.471	0.230	0.273	-
2019	0.461	0.635	0.411	0.426	0.395	0.348	-	-	0.351

Thus, students in the digital semester were not only more engaged in asynchronous activities but also more successful in their examinations. The smaller number of synchronous sessions and the lower participation in them did not have a too negative impact on success as also the correlation with the synchronous learning engagement in tutorials shows. The correlations from Table 16 indicate that self-regulated learning correlates most highly with success in the examination. The number of distinct quizzes completed stands out as the most highly correlating indicator, but the total number of completed quizzes also correlates highly. (Active) participation in synchronous events correlates with examination success in both tutorials and lectures in the middle range only. Watching lecture videos correlates only slightly. The comparison of the two years shows that self-regulated learning is even more important in the digital semester as it correlates stronger with examination success than in the traditional semester. The corresponding linear regression model in Table 17 shows that because of the many correlations between the predictors, the importance of the number of distinct completed quizzes increases even more, while many other elements of student learning have no significant effect on examination success. Overall, it is to be expected that the students increased their total learning engagement and focused (unknowingly) on the more predictive activities of practicing quizzes. In this way, the higher learning success may actually be explained by changed learning behavior.

Table 17: Multiple linear regressions explaining score in the final examination by all learning behavior measures at the same time in both years.

	2019					2020				
Indicator	**Coeff.**	**Std. Err.**	**Beta**	**t-value**	**p-value**	**Coeff.**	**Std. Err.**	**Beta**	**t-value**	**p-value**
Completed quizzes	-0.056	0.025	-0.307	-2.296	0.023	-0.026	0.015	-0.135	-1.710	0.089
Distinct completed quizzes	0.563	0.091	0.848	6.178	0.000	0.481	0.060	0.689	8.053	0.000
Success rate of completed quizzes	0.276	0.116	0.189	2.386	0.018	0.428	0.097	0.233	4.408	0.000
Success rate of distinct	0.066	0.084	0.059	0.787	0.432	0.068	0.075	0.048	0.905	0.366

completed quizzes										
Rate of participated tutorials	-0.082	0.053	-0.135	-1.541	0.125	-0.002	0.051	-0.004	-0.045	0.964
Rate of active tutorial participations	0.063	0.058	0.080	1.090	0.277	0.062	0.059	0.099	1.043	0.298
Watched videos	-	-	-	-	-	-0.006	0.011	-0.031	-0.544	0.587
Rate of distinct watched videos	-	-	-	-	-	0.029	0.039	0.042	0.741	0.459
Rate of active lecture participations	0.042	0.039	0.071	1.089	0.278	-	-	-	-	-
Intercept	0.086	0.061	-	1.408	0.161	0.030	0.061	-	0.488	0.626

10.5 Discussion

Summarizing the data-driven learning analytics, digital teaching and learning during the rapid transformation caused by the COVID-19 pandemic seems to have been successful from a cognitive perspective. This, however, neglects the affective perspective on students' life in general and learning in particular (Dhawan, 2020; Fairlie & Loyalka, 2020). Research on previous crises and the latest insights on the COVID-19 situation indicate that major effects on affective layers exist (Mucci et al., 2016; Salari et al., 2020). Not only the concerns about personal health but also the societal changes like physical distancing, result in increased stress, insecurity and even anxiety (Barzilay et al., 2020). A non-consideration of these effects does not seem appropriate when analyzing learning. Therefore, a holistic evaluation of learning during COVID-19 should incorporate an analysis of the affective layers.

This also gets reflected in the survey of the students participating in the digital course in 2020. The students reported an increased burden due to lack of social contact (65.6% very high or high burden). They further experienced an increased level of stress (54.2%) and 58.3% of the students experienced difficulties motivating themselves to learn.

These results indicate that besides the physical isolation caused by the COVID-19 lockdowns, the students experienced problematic social isolation. Surprisingly, they did not react by increasing their participation in synchronous learning. They put themselves in a 'learning isolation' instead of participating in the offered cooperative tutorial sessions. Combining both perspectives, an overall evaluation of learning during COVID-19 shows positive impacts in terms of students' learning engagement and examination success, but negative consequences for students' affective life due to a (partially self-chosen) three-fold isolation.

This finding can be interpreted in two ways with regard to the COVID-19 lockdowns. On the one hand, the emergency remote teaching formats have been effective to the extent that students have learned something, and in the course studied, even more than before. ERT was therefore a successful response to the situation. At the same time, ERT was a burden for many students, which put their resilience to the test and for whom it is unclear whether a much longer phase of ERT would not produce worse results, as resilience cannot be maintained indefinitely in an exceptional situation.

At the same time, these results are highly relevant for students who, in dual-mode teaching situations or in other formats, do not have the choice of attending classes in person because, for example, they have care work to do or commuting distances in rural areas are unreasonable. These students could find themselves in a similar situation of involuntary distancing. This study shows that greater support for these students could be necessary, but also leaves open the need for research into what support could best help these students.

10.6 Concluding Remarks on Study 9

Even though study 9 did not directly answer any of the research questions of this thesis, it does provide an interesting perspective on how vulnerable the system of relationships between non-cognitive factors and learning behavior and success is. The external shock inflicted by the pandemic on the usual learning scenario affects all levels of learning. Learning success seems to be higher in this first year of the pandemic than before. The measured variables of learning behavior have also changed for the better. Theoretically, these two changes still fit together well, as a positive effect of increased learning engagement on learning success seems reasonable. However, the measured attitudes (motivation) and emotions (stress, strain, isolation) have deteriorated. As an external shock, the pandemic therefore appears to be stronger than the positive relationship that has otherwise been found between non-cognitive factors and learning. Especially for possible intervention studies, this is a reminder that the findings on the main research question cannot be used to assume positive effects on the other parameters for every intervention that leads to a positive change in non-cognitive factors or learning behavior.

11 Last Part: Discussion: Attitudes Matter, Beliefs Probably Too

11.1 Overview Over the Results of the Nine Studies

Before all the results of this thesis, which help to provide answers to the research questions of this thesis, are summarized and discussed for each research question, the results of the individual studies are first summarized again with regard to their relevance for this thesis. These summaries are provided for each study separately. In the next subchapter, the results are then synthesized and discussed with regard to each research question.

Study 1 of this thesis aimed to identify and describe types of beliefs about statistics, to conceptualize these types together in a theoretically sound model and to develop a measurement instrument that can measure learners' beliefs about statistics in line with this conceptualization. The goal of identifying and describing types of beliefs about statistics was pursued through qualitative focus groups and interviews and their evaluation using a grounded theory approach. This way, four types of beliefs about statistics could be identified:

In the rules-based conception, statistics is primarily understood as a collection of formulas, rules and procedures from which one must select the correct statistical tools and apply them precisely in order to solve a data-related task correctly. In the confirmatory conception, statistics is primarily understood as a systematic approach to prove or disprove scientific or real-world hypotheses, models or theories on the basis of data by systematically applying suitable statistical procedures and tests. In the descriptive conception, statistics is primarily understood as a collection of possibilities to present data and information from the real world in a clear, understandable and informative way through visualizations, tabular representations or key figures in order to convey a clear picture of reality. In the investigative conception, statistics is primarily understood as a flexible and cyclical process of analyzing data that aims to find insights in data and link them to real-world or scientific theories in order to find explanations for existing phenomena or solutions to real-world problems.

For a theoretical framing of the four described types of beliefs about statistics, the communication behavior in these belief systems between the data level and the theoretical level is used. In the investigative conception, there is a constant dialog between data and theory, which can be initiated by either side at any time. In the descriptive conception, communication is initiated by the data, which regularly pass on information to the theoretical level, which processes the information but does not respond to it. In contrast, in the confirmatory conception, communication starts from the theory, to whose hypotheses the data only react with a simple yes/no answer without entering into a real dialog. In the rules-based conception, no communication takes place at all, as the value of such communication is not seen.

This communication metaphor creates a two-dimensional model of beliefs about statistics, in which the intensity of communication forms one dimension. This dimension can be described as isolated vs. communicative, unconnected vs. connected, static vs. dynamic or novice vs. advanced. The rules-based conception lies at the bottom of this dimension close to the first adjectives, the investigative conception at the top close to the latter adjectives. The second dimension is characterized by whether the level of data or the level of theory is the dominant actor in communication. On one side of the spectrum, the data are the dominant actor, determining and shaping the communication by initiating it, keeping it active and controlling it, while the theory only reacts. On the other side of the spectrum, by contrast, are beliefs that see theory as the driving force of communication. The descriptive conception is entirely on the data-driven side of this spectrum, while the confirmatory conception is entirely on the theory-driven side. Since all four types of beliefs are close to the middle in the respective dimension in which they were not described as a prototype of one end of the spectrum, the overall result is a diamond-like arrangement, which earned the model the name "The Diamand Model of Statistics".

To measure the degree to which learners (or other people) tend towards each of the four types of belief systems, a survey instrument was developed. The instrument consists of a total of 36 items, nine for each type of belief. These items are organized in such a way that nine questions or topics were identified (e.g. What happens during the statistical process? or How do you become an expert in statistics?) and then each of the four types of beliefs provides a statement on these questions or topics based on their world view. Psychometric and qualitative tests of the instrument showed that it does not work exceptionally well, but at an acceptable level. The psychometric parameters are roughly in the same range as they are for the measurements of attitudes towards statistics and statistics anxiety in Study 5. On this basis, Study 1 publishes the developed instrument and makes it freely available to the scientific community.

Study 2 of this thesis aimed to develop a prototype for a chat-based digital tutor, to test it in the field and to develop it further in order to derive a design theory for the development of learning platforms as intermediaries between instructors and learners. To this end, a design concept was developed through literature reviews and expert workshops. This was then technically implemented and tested in a statistics course. The further development of the first software solution was driven by comprehensive quantitative and qualitative user surveys, resulting in a comprehensively evaluated prototype and an underlying design theory.

The first important aspect of this prototype for the students was that it was able to respond to their statistical questions in a course-specific manner via a comprehensive chat component. The system thus provided the students with added value, which improved student support outside of classes. From a computer science perspective, it was particularly important to develop a chatbot that not only supports a one-off communication situation, but also acts as a companion over a long period of time.

As a second important aspect for the design theory and also for the use of the system in teaching, the system was designed as a comprehensive solution that combines all digital learning offerings and aids for students within one system. This meant that students could not only chat with the system, but also download lecture slides, watch lecture videos, take part in audience response polls, complete quizzes as homework or for practice, solve a daily quiz, take part in lecture hall games and view their own usage statistics. In this way, a system was created that constantly supports students in their learning, regardless of time and place. As all these support services were combined in one system, it established itself as a multi-layered, helpful digital tutor that became a constant companion in the statistics course.

In addition to the added value of this digital system for the students, as a third important aspect about the design was that it enabled comprehensive work with digital behavioral data of the learners. The breadth of the supports offered and their integration into one system ensured that comprehensive and fine-grained behavioral data was generated, which all came together in the same database. The system developed was thus from the start developed as a basis for subsequent learning analytics projects and served as such in several subsequent studies.

One of these studies was Study 3, which aimed to use the digital behavioral data obtained to investigate the relationship between learning engagement and spaced learning on the one hand and learning success on the other, using video-based teaching as an example. For this purpose, the digital behavioral data of the system from study 2 were processed in such a way that the use of videos in student learning could first be comprehensively described before a series of variables for quantifying learning engagement and spacing of learning were operationalized, whose relationship to learning success was examined in the final step.

In the first step, the study showed that students during the pandemic in the summer semester of 2020 learned intensively with videos, very often with more than a single time watching the video. This learning with videos was widely spread over the course of the semester, but also intensified in the exam preparation phase. For this last phase of the course, it was also shown that many videos were played at a higher playback speed, which was hardly ever the case before.

In the second step, the study showed that not only course-wide descriptive overviews can be obtained from the behavioral data, but that variables of learning behavior can also be operationalized at an individual level. It was possible to generate variables of learning engagement as well as of the spacing of learning. With regard to the spacing of learning, an important innovation compared to previous research was that different levels of the spacing of learning could be considered separately. Thus, the study considered learning during the semester versus directly before the exam as one level of spaced learning, concentration or non-concentration on one day of the week as a second level and, thirdly, the centering of learning on the classic working hours between 9 a.m. and 5 p.m. or learning outside this period. In addition, the academic delay in learning with videos was also operationalized.

In the third step, the study examined the relationship between the operationalized variables of learning behavior and learning success. It was found that all the variables examined correlated significantly with learning success. Only when the variables were controlled against each other some of the significances disappeared due to the high correlations between the behavioral variables. What is interesting about the observed correlations, however, is that not all of them had the expected sign. While learning engagement was always positively related to learning success and academic delay had the expected negative relationship, a mixed picture emerged for spaced learning. Spreading learning throughout the semester rather than studying heavily before the exam is positively related to success, as expected. However, spreading learning within the week correlated negatively with learning success. Studying outside of 9 a.m. to 5 p.m. also correlated negatively. In terms of the theory of spaced learning, both must be surprising. At the same time, however, it remains that all the considered variables of learning behavior correlate with learning success and could therefore be potential mediators for the relationship between non-cognitive factors and learning success.

This finding of correlations of learning engagement and spacing of learning with learning success, and thus the possibility that these are mediators of relationships between non-cognitive factors and learning success, is taken up by Study 4. The aim of Study 4 was first to investigate whether a correlation between the types of beliefs about statistics developed in Study 1 and learning success can be found, as expected from corresponding findings in other disciplines. Building on this, the aim was then to investigate whether and to what extent the variables of learning behavior operationalized and associated with learning success in Study 3 and some similar variables can be considered as mediators for this correlation. For this purpose, the digital behavioral data of the system from study 2 were operationalized as in Study 3 and linked with a survey of the students and the results of the course examination.

In the first part of the study, only very small correlations were found between beliefs about statistics and exam success, some of which were perceptible and some very clearly non-significant. Apart from a possile type 2 error, this suggests either that such a correlation does actually not exist or that it is so small that the number of cases used (n = 238) did not have the necessary power to detect it.

After these results, it was less surprising that beliefs about statistics did also not correlate with learning behavior, at least less frequently than type 1 errors would be statistically expected. The hypothesis that there is a relationship between beliefs about statistics and learning success that is mediated by learning behavior could therefore not be confirmed in any way in Study 4. However, this does not mean that there cannot be (other) effects of beliefs about statistics on learning behavior, for example as very small effects mediated by attitudes towards statistics or as effects on other types and operationalizations of learning behavior.

Just like Study 4, Study 5 took up the finding of correlations between the operationalized learning behavior variables and learning success from Study 3 and pursued the idea that learning behavior could be an explanatory mediator between non-cognitive factors and learning success. In Study 5, however, this approach was not tested for beliefs about statistics as in Study 4, but for attitudes toward statistics and statistics anxiety. To test this approach, Study 5 first developed a theoretical model that combines Expectancy-Value Theory and Control-Value Theory and thus relates attitudes towards statistics and statistics anxiety to learning behavior and learning success within the same model. The attitudes towards statistics, measured in expectancy and value, constitute the starting point of a causal chain. Both have an effect on statistics anxiety in the first step. All three of these variables then have an effect on the effort intended by the students. In the following, an effect on learning behavior is assumed for all four variables and, finally, an effect on learning success for all variables considered. This causal model is supplemented by the control variables high school graduation average and female gender, for which effects on all variables included in the model are permitted.

Like Study 4, Study 5 used the system's digital behavioral data from Study 2 and its operationalization from Study 3, as well as some additional, similarly operationalized variables. It combined them with data from a survey conducted among the students on their attitudes towards statistics and their statistics anxiety and with the results of the course examination. This way, the theoretically developed model could be empirically tested on the basis of n = 181 students. In a first step, the learning behavior contained in the model was filled in with one of the variables operationalized from the digital behavioral data at a time. This way, the theoretical model was estimated six times with a variable of learning behavior that reflects learning engagement. All six models show that expectancy and value are significantly associated with the intended effort. Intended effort, in turn, is significantly linked to actual learning engagement. This in turn, as already shown in Study 3, is linked to learning success. For attitudes towards statistics, measured here in the two main categories of expectancy and value, it can therefore be said that, with the addition of the further mediator of intended effort from the Model of Students' Attitudes Toward Statistics, a mediating role of learning engagement between attitudes towards statistics and learning success can indeed be supported by the data.

The same is true for the six models in which a measure of spaced learning fills the place of learning behavior, and also for a model with academic delay as a variable of learning behavior. In all seven models, attitudes (six times via intended effort, once directly) have a link to learning behavior in the sense that positive attitudes are associated positively with learning behavior, which in turn was shown to be associated with higher learning success in Study 3. Thus, spacing of learning as a variant of learning behavior also plays a mediating role in the positive relationship between attitudes towards statistics and learning success.

This cannot be found for two newly generated variables, which were intended to be a first attempt to measure the quality of learning. Both variables correlate with learning success as hoped, but not with one of the attitude variables or statistics anxiety. Nevertheless, a third reason for the relationship between attitudes towards statistics and learning success can be found in the model alongside the mediating roles of learning engagement and spaced learning. The high school graduation average correlates positively with all measured variables except statistics anxiety. Based on this finding, part of the relationship between attitudes towards statistics and learning success may also be based on a spurious correlation induced by the high-school graduation average and thus perhaps in the background by variables such as intelligence or prior knowledge.

There are hardly any correlations for statistics anxiety in any of the models. Expectancy is relatively strongly associated with statistics anxiety. Other than that, there are no significant connections that would hold up robustly across several models. This raises the question of whether statistics anxiety has its own contribution and its own effects at all or whether negative expectancy already explains all relationships.

A final overall model, in which three variables representing the three dimensions of learning behavior - engagement, spacing of learning and quality of learning - are included, confirms the findings of the individual models. As mediators, learning engagement and spacing of learning can each explain part of the relationship between attitudes towards statistics and learning success. A spurious correlation induced by the high-school graduation average explains another part. Nevertheless, part of the correlation remains unexplained. In contrast, no separate role for statistics anxiety can be identified in relation to learning behavior or learning success.

After study 5 confirmed the relationship between attitudes towards statistics and learning success and was able to identify mediators, study 6 takes this up in order to investigate whether there could be consequences of beliefs about statistics that could not be found in study 4. These could lie in particular in the idea that beliefs about statistics have an effect on attitudes towards statistics. Study 6 of this thesis therefore aimed to discover relationships between beliefs about statistics and attitudes towards statistics. To this end, both survey data using the instrument from Study 1 and qualitative interviews were conducted.

In the quantitative part of the study, the four types of beliefs about statistics were regressed as explanatory variables on six dimensions of attitudes towards statistics that are represented in the Survey of Attitudes Toward Statistics: the perceived difficulty of statistics and the perceived own competence as facets of expectancy from the Expectancy-Value Theory, the value attributed to statistics, the interest in statistics and the affects towards statistics as facets of the value component in the Expectancy-Value Theory and sixthly the effort intended by the students.

It was found that there are indeed relationships between beliefs about statistics and attitudes towards statistics. A rules-based conception is negatively associated with attitudes towards one's own affects and competence and towards the difficulty of statistics, while it is positively associated with the intended effort. Attitudes towards the value of statistics and interest in statistics do (just) not reach significance. An investigative conception, on the other hand, is positively linked to all the dimensions of attitudes considered, with only perceived difficulty just failing to reach significance. A confirmatory conception, on the other hand, has no significant links, and a descriptive conception only one link to perceived value. Overall, this shows that novice beliefs about statistics are linked to negative attitudes, while more advanced beliefs are linked to more positive attitudes. For the other dimension of the Diamond model from Study 1, no correlation with attitudes can be observed.

In the qualitative part of Study 6, numerous examples were found that illustrate how beliefs about statistics can affect attitudes towards statistics indirectly or via interaction effects. For students with a strongly rules-based conception, previous experience with mathematics played a greater role in the justification of attitudes than for other students, in both positive and negative directions. In the descriptive and confirmatory conceptions, different application ideas were emphasized that increase the value of statistics for these respondents. Thus, it could be shown through various examples that beliefs about statistics do not necessarily have effects by changing attitudes towards mathematics or about the value of statistics for science, but that they can change priorities and emphasize different aspects of statistics more or less, which in turn can change the tendency of attitudes. Overall, Study 6 thus showed that a small linear relationship can be found between beliefs about statistics and attitudes towards statistics, but this could possibly just be a consequence of multiple interaction effects that actually operate in a more complex way.

Study 7 therefore looked at the effects of beliefs about statistics from a broader perspective in order to gain a more holistic view of students' thinking about statistics and their studies. To this end, Study 7 aimed to describe students' beliefs about statistics, to reconstruct the contexts in which they arose and subsequently to understand what role they play in their studies. Overall, the aim was to shed light on a variety of aspects of studying, including the motives for choosing their degree program, elective decisions during their studies, such as the choice of voluntary courses, and which forms of learning are preferred during their studies. A particular focus was to be placed on understanding the role beliefs about statistics play in the process of appropriating statistics during the course of study and how students develop a sense of belonging to statistics.

However, achieving all of these goals, ideally in a longitudinal observation of students, would have been beyond the scope of Study 7 and this thesis. As a first step, Study 7 therefore focused on in-depth qualitative interviews with three students at the beginning of their statistics studies. The case studies showed how different beliefs about the nature and purpose of statistics lead students to engage with their future subject in different ways before starting their studies and how this results in different expectations of studying statistics. One case, for example, regarded statistics as a problem-solving and modeling process similar to an investigative conception and therefore worked on small statistical puzzles in his free time before starting his studies. Another student had a more descriptive conception and therefore, not coincidentally, did an internship in the field of market and survey research. These cases show that beliefs about statistics can have a variety of consequences for studying and learning beyond impacting the attitudes towards statistics.

While all the studies mentioned so far understood the non-cognitive factors investigated to a certain extent as static, Study 8 breaks with this assumption, as it is known from many studies that non-cognitive factors (can) change. Study 8 aimed to describe the changes in attitudes towards statistics during an introductory statistics course in more detail than has previously been done and to explore the reasons that explain these changes. To this end, a qualitative interview panel with six students and four data collection points was conducted for Study 8.

With regard to the development of attitudes during the course, Study 8 was able to show in the case studies that this is not a linear development between the measurement at the beginning of the course and the measurement at the end of the course. While attitudes at the end of the course were often at a similar level to those at the beginning of the course, in some cases even slightly more positive, most students experienced a significant drop in attitudes in the meantime. Contrary to what is assumed in the literature, however, this was rarely a shock at the beginning of the course, but rather a downturn in the third survey period, which was primarily driven by the previously taught introduction to inferential statistics, which was perceived as abstract and complicated.

A look at the reasons showed, as in Study 6, that the dimensions of attitudes measured in the Survey of Attitudes Toward Statistics can be broken down into different facets in the students' perceptions. With regard to the value of statistics in particular, the students made an explicit distinction between the value for their further studies, the value for everyday life, the value for their future career, the value for society and the value for science. These different facets of value play a different role for the students in their contribution to the overall development of attitudes towards statistics. Seeing statistics as valuable for further studies was a widely shared and a very stable position that kept many students on track even in difficult phases. The value of statistics for everyday life and the value for their future career were assessed very heterogeniously and therefore contributed more to the heterogeneity of attitudes between students. Different perspectives on the value of the specific topics covered for everyday life and career were also a factor behind the changes in attitudes.

Another driver behind the developments in attitudes was difficulty. Even more than the value of statistics, this was seen as topic-specific and therefore fluctuated relatively strongly, with the introduction to inferential statistics often being perceived as the most difficult.

The third driver behind the development of attitudes towards statistics is the effort involved, which students perceive as very high, especially higher than in their other courses, which generally leads to a dampening of attitudes.

These results show that different facets of attitudes towards statistics may have different roles. Some appear more as a consensus between students, others are drivers of heterogeneity, some are more stable over time, others are drivers behind the change in attitudes and still others are much more content-related and not a general characteristic of statistics. Future research should take this heterogeneity of facets into account, also when investigating the effects of attitudes on success, as only facets that are stable within individuals can be expected to have a stable effect on a person's behavior and success. For facets that change more strongly, it is much more important to investigate the role that the current status and its most recent development play for learning behavior and success. In addition, how quickly changes in attitudes lead to changes in behavior and success, or whether reverse causality can also be empirically proven, should be taken into account in future in-depth research on the effects of attitudes for learning statistics.

Study 9 concludes the studies conducted in this thesis by providing a look at how stable or unstable any relationships that may be found between non-cognitive factors, learning behavior, and learning success may be. Study 9 was not originally planned, but was started spontaneously at the beginning of the pandemic-related lockdowns at the beginning of 2020 in order to scientifically investigate the consequences of emergency remote teaching for learning. The study took advantage of the fact that extensive data, in particular digital behavioral data, had been collected for the testing of the software in Study 2 in the summer semester of 2019, which could now serve as a benchmark for 2020 to compare with. Study 9 interpreted the pandemic situation as a natural experiment and aimed to record the effects of emergency remote teaching on students' learning behavior and learning success, to investigate whether learning behavior in this situation is related to learning success in a different way and to ask what this learning situation does to students affectively.

A look at the learning behavior showed that the students were more engaged under the conditions of emergency remote teaching than in the comparison semester in almost all of the variables considered. Only participation in the tutorial sessions decreased. One explanation for this is that the students spaced their learning more over the week. However, this in turn may also be due to the fact that elements that structure the week, such as face-to-face events, have been removed and learning has therefore become less organized.

However, looking at the learning success showed that the students in 2020 were actually more successful than those in the comparison semester. The failure rate fell, the average number of points achieved and thus the average grade improved. A look at the correlations between learning behavior and learning success showed that learning success in 2020 was much more closely linked to the measured learning behavior than in the previous year.

The students were thus much more engaged under emergency remote teaching conditions than under normal conditions and, as an apparent consequence, were actually more successful. However, they also stated in a survey that they were less motivated, more isolated and more stressed. Learning during emergency remote teaching can therefore not be fully evaluated as a success. Even if it seemed to have worked to a certain extent for the urgent situation, the long-term feasibility and sustainability of emergency remote teaching is questionable due to the negative affective consequences.

It is furthermore interesting for the investigation of emergency remote teaching that the students felt a high level of stress, particularly due to a high level of perceived isolation, and at the same time participated less than before in the only remaining live and interactive sessions, the tutorials. For the investigation of the relationship between non-cognitive factors and learning behavior and success, it is interesting that the effects observed in this study run counter to the relationships in the other studies. While otherwise positive relationships were shown, the disruptive effect of emergency remote teaching appears to be stronger, so that learning behavior and learning success develop in the opposite direction to the non-cognitive factors considered. This finding shows that the positive relationship between non-cognitive factors and learning behavior does not have to remain stable in interventions, but that the effects on learning success, learning behavior and non-cognitive factors must all be considered, especially in future intervention studies.

11.2 Answers, Reflections and Conclusions to the Research Questions of This Thesis

After the results of the individual studies have now been summarized as to their relevance for this thesis, in the following, answers to the research questions of this thesis will be synthesized from these results. First, the relevant answers are summarized and discussed and a conclusion is drawn for each research question seperately. Then, an overall conclusion for the main research question completes the discussion. Reflections on the limitations of this work and follow-up topics for future research follow in subsequent subsections.

11.2.1 Discussion of Research Question 1

> *Which beliefs about statistics can be conceptualized? How can these concepts be measured?*
> *(RQ 1a: **Conceptualization of beliefs**, RQ 1b: **Measurement of beliefs**)*

This research question is addressed in this thesis mainly with Study 1. Within this study, this thesis identifies four types of beliefs about statistics: a rules-based conception, a confirmatory conception, a descriptive conception and an investigative conception.

If one compares these four types and their more detailed description in Study 1 with the other published typologies, the very close proximity to the work of Justice et al. (2020) is striking. After their qualitative analysis, Bond et al. (2012) considered their results to be so close to Reid and Petocz´s (2002) typology that they adopted Reid and Petocz´s (2002) typology in the following. If this thesis was only concerned about describing types of beliefs about statistics, a similar conclusion could be made in this thesis regarding the work of Justice et al. (2020). In this sense, this part of this thesis could be understood as a replication of Justice et al. (2020) in a different context and with a different methodological approach. This replication provides first great value of this thesis. Especially in the field of qualitative educational research, replications are rarely carried out. Having two qualitative studies conducted in such different ways with such similar results therefore indicates a particular strength of the results and must be seen as strong support for both typologies.

In the field of conceptualizations of beliefs about statistics, with Reid and Petocz (2002) and Justice et al. (2020) there are thus now two replicated typologies. In addition to being replicated, their quality can be measured by their theoretical foundation and their predictive validity, i.e. by the empirical finding that they are related to other relevant concepts. In terms of their theoretical foundation, Justice et al. (2020) provide a little more than all other conceptualizations by not only distinguishing novice conceptions from more advanced conceptions, but by locating them in two dimensions, one capturing the consideration of uncertainty and one the importance of the real world.

Empirical evidence on the relevance of beliefs about statistics has so far only been provided by Bond et al. (2012) using Reid and Petocz´s (2002) typology. In Study 6, however, this thesis reveals a relationship to attitudes for the conceptualization of beliefs about statistics developed here as well, both using qualitative and quantitative approaches. In addition to its replication, this work therefore provides further reinforcement for Justice et al.´s (2020) typology.

However, this thesis is not limited to the construction of a typology or its replication. Rather, it goes much further into a theoretical discourse than any previous work on the conceptualization of beliefs about statistics. This thesis conceptualizes the four types of beliefs found, such as Justice et al. (2020), in two dimensions, but is closely oriented towards scientific debates on the nature of statistics within the discipline. To this end, the metaphor of Bailyn (1977) and Grolemund and Wickham (2014) of understanding statistics as a communicative process between theory and data is adopted. The four types of beliefs about statistics are then classified along one dimension according to how uncommunicative, unconnected and static or communicative, connected and dynamic communication between theory and data is in that conception. The second dimension captures who the driving force is in the communication, i.e. whether the dialog is more data-driven or theory-driven. Applying these two dimensions, the four types of beliefs about statistics can be arranged in the form of a diamond, which earned the model the name Diamand Model of Statistics.

The four types described are the first to benefit from this conceptualization. Through their theoretical positioning and their theoretical relationship to each other, the types described gain in sharpness and explanatory power. This further strengthens the typology in a way that is not found in all other publications on beliefs about statistics.

In addition to an added value for the typology, both research on beliefs about statistics and the philosophical debate about the nature of statistics in the discipline benefit by linking the two. Thus, research on beliefs about statistics provides the field's self-reflection with a surface for comparison in which it can reflect on whether what students think about statistics is accurate or representative. This form of external perception can be used to derive either confirmation or a need to catch up. In addition, there may even be something in the students' statements that shows even experienced statisticians what statistics can also be.

Part of this reflection could be the realization that all conceptions are correct to a certain degree or in certain areas or situations, at least the descriptive conception, the confirmatory conception and the investigative conception. Although describing and testing are both legitimate and frequent concerns of statistics, they do not occur equally often in all areas. For example, the descriptive conception could be associated with the reporting work of official statistical offices, while statistics in clinical trials is strongly based on the testing logic of the confirmatory conception. It is also noticeable in the interviews in Study 1 that examples of the descriptive conception often come from the field of official statistics and opinion polling, while examples of the confirmatory conception mostly come from scientific theory testing studies (in the social sciences). Thus, the presented conceptualization of the Diamand model can broaden the discussion about facets and subdisciplines of statistics, which currently revolves strongly around the relationship to and role of data science.

From such a debate about facets of statistics expressed in the conceptions, conclusions can in turn be drawn for research on beliefs about statistics and for the teaching of statistics. If different facets are important in different disciplines and fields of application, statistics teaching should perhaps take this into account. While previous conceptions only distinguish between novice and advanced (preferable) beliefs, this adds an element of fit. The question of which beliefs students should have is then no longer simply answered with an ordinal hierarchy of beliefs. Instead, instructors and researchers must ask themselves which conception is more appropriate for the intended field of occupation of the students being taught. In some statistics courses for non-statisticians, this may mean that a descriptive or a confirmatory conception is favored or at least totally fine to have. In the training of statisticians, however, it probably means that both conceptions, and of course an investigative conception, must be fostered in order to prepare students for the polyvalence of their profession and to create an awareness for differences in perception of statistical work between fields of application.

In addition to these and other benefits from the typology created and its conceptualization for the discipline of statistics, for teaching statistics and for statistics education research, actors outside of statistics can also benefit. Conceptualizations of beliefs about other disciplines often also contain a dimension that describes them as novice / static versus advanced / dynamic. A second dimension, on the other hand, is often not included, and if it is, then it is not very well founded in subject-specific theory (see e.g. Grigutsch, Raatz, & Törner, 1995; Lederman et al., 2002). The given conceptualization can serve as an example of how discipline-unspecific dimensions (static vs. dynamic) and discipline-specific dimensions (data-driven vs. theory-driven) may interact. Regardless of whether statistics is considered a sub-discipline or a related discipline of mathematics, the question arises particularly for mathematics education research as to whether conceptualizations of beliefs about different sub-disciplines would be useful, as here for statistics, since surveys of teachers, for example, revealed differences in their perspectives on teaching different sub-disciplines (see e.g. Eichler, 2006; Meinke, 2016).

Besides describing a typology of beliefs about statistics and conceptualizing them in a comprehensive model and thus answering research question 1a, Study 1 also develops and tests a quantitative measurement instrument to answer research question 1b and publishes it within this thesis. The instrument was comprehensively tested qualitatively and psychometrically in Study 1 with data from a social science context in Germany and a life science context in the United States. The results were not outstanding, but sufficient and not inferior to the psychometric quality of other comparable measurement instruments. The scientific community is thus provided with an instrument with which it can work quantitatively on all the general research questions formulated at the beginning of this thesis. It is to be hoped that this will considerably accelerate scientific progress on beliefs about statistics, their causes, consequences and mechanisms, development and manipulation. The first steps in this direction will be taken below.

All in all, it can be said that this thesis enriches research question 1 with a comprehensive contribution that will hopefully stimulate future research on beliefs about statistics. However, the theoretical contribution and the instrument will certainly still have to prove themselves. Both authors of Study 1 therefore look forward to reading structure-confirming or critical validation studies on the instrument or studies applying it. Likewise, they hope for a lively debate on the theoretical conceptualization.

11.2.2 Discussion of Research Question 2

> *What consequences do the beliefs about statistics have for learning? What are the consequences of the beliefs for learning success? What are the consequences of the beliefs for the attitudes of the learner?*
> *(RQ 2a: **Consequences of beliefs for success**, RQ 2b: **Consequences of beliefs for attitudes**)*

This research question is adressed by multiple studies within this thesis, so that the results of these studies must first be synthesized. With the first part of Study 4 this thesis directly addresses research question 2a by looking for direct effects of beliefs about statistics on learning success. However, it cannot find such direct effects. Some links between beliefs about statistics and learning success are quite clearly not significant, other links miss significance less clearly, but also not particularly close. This does not disprove a link between beliefs about statistics and learning success, but a large effect seems unlikely given the power of the number of cases and the analysis used. Only whether there is a (very) small link or no link at all can still be considered open.

At this point, the first part of Study 6 of this thesis picks up and quantitatively finds relationships between beliefs about statistics and attitudes towards statistics. This fist of all provides a partial answer to research question 2b. However, since attitudes towards statistics have been shown in many studies and also in Study 5 of this thesis to be related to learning success, this implicitly also provides another answer to research question 2a. Beliefs about statistics are related to learning success in that this relationship is mediated by attitudes towards statistics. However, since both the link between beliefs about statistics and attitudes and their link to learning success are not particularly strong, the overall effect of beliefs about statistics on learning success is quite small.

With this finding, however, it should be noted that learning success here is measured by the course grade in the students' first statistics course. However, this target variable is subject to strict regulations and is more of a short-term perspective. Thus, all students must take this course and only have a little more than 14 weeks to let their beliefs about statistics affect their success. If the effects of beliefs were more in the sense that they influence which voluntary courses are chosen, how much extracurricular activities related to statistics are sought and which ones, and how quickly and how strongly belonging to statistics develops, then these effects could only be determined with a long-term observation. Such a long-term observation has not yet been carried out, neither in this thesis nor anywhere else, but Study 7 indicates in qualitative case studies that precisely these effects could exist and that such a long-term observation could therefore be worthwhile.

For the time being, however, it should be noted that the studies in this thesis, taken together, demonstrate a connection between beliefs about statistics and attitudes towards statistics and thereby also to learning success. This last connection is particularly important, as a relationship between beliefs about statistics and learning success is probably the only viable reason for many recipients to consider beliefs relevant at all. For this very reason, further research on the long-term effects on decision-making and performance would be helpful in order to empirically substantiate the evidence for such effects of beliefs about statistics found in this thesis within Study 7. However, a complete irrelevance of beliefs can already be rejected on the basis of this study. On the contrary, this thesis, particularly in Study 7, points to a large number of effects of beliefs about statistics and thus supplements research question 2b with some aspects that previous research has not yet considered.

11.2.3 Discussion of Research Question 3

> *Which mechanisms, in particular which changes in objectively observed learning behavior, help to explain why non-cognitive factors are related to learning success? (RQ 3a: **Mechanisms of beliefs**, RQ 3b: **Mechanisms of attitudes**, RQ 3c: **Mechanisms of statistics anxiety**)*

Direct answers to this main research question of this thesis are provided mainly within Study 4 and Study 5. In study 4, this thesis focuses on learning behavior as a mediator between beliefs about statistics and learning success, i.e. it aims to answer research question 3a. However, since, as already discussed for research question 2, no relationship between beliefs about statistics and learning success can be proven in study 4, it is hardly surprising that no relationship between beliefs about statistics and learning behavior can be found in this study either. Although, as just discussed for research question 2, effects of beliefs about statistics on learning can be assumed, the attempt to find causally interpretable mediators for this relationship fails due to the lack of detectability of the assumed relationship between beliefs about statistics and success. Research question 3a thus does not appear to be answerable in any meaningful way, as the mechanism of a non-existent effect would not be explainable. At the same time, however, the findings of this thesis from Study 6, which prove a relationship between beliefs about statistics and attitudes towards statistics, also apply here. In light of this, the final answer to research question 3a should at least be postponed. Understanding the relationship between attitudes towards statistics and learning success, on the other hand, is becoming increasingly important.

In Study 5, this thesis focuses on attitudes towards statistics and statistics anxiety and their relationships to learning success in order to determine whether variables of learning behavior can be considered as explanatory mediators as implied by research questions 3b and 3c. This thesis shows that both learning engagement and the distribution of learning, as measured by digital behavioral data, do indeed have a mediating function between attitudes towards statistics and learning success. In addition, however, a part of the relationship between attitudes towards statistics and learning success appears to be a spurious correlation, which can be explained by shared correlations with the high-school graduation average.

With this finding, this study can uncover a mechanism between attitudes towards statistics and learning success, which, at least according to Expectancy-Value Theory and Control-Value Theory, is a causal mechanism from attitudes towards statistics via learning behavior to learning success. This work thus provides empirical support for both theories and for their linkage as developed in Study 5. This can be seen as a first added value of the results of research question 3.

A second added value can be seen in the fact that the results of this study can be interpreted as causal, at least according to the theories. If the correlation between attitudes towards statistics and learning success were purely spurious, which is only the case to a small extent, it would be completely unclear or even highly doubtful whether an improvement in attitudes, for example through an intervention, could actually lead to an improvement in learning success. There are a large number of intervention studies that aim to improve motivational attitudes and often do so, both within statistics education (e.g. Paul & Cunnington, 2017; Posner, 2011; Smith, 2017; C. Xu et al., 2020) and even more beyond (for reviews and meta-analyses, see for example McBreen & Savage, 2021; Wigfield, Turci Faust, Cambria, & Eccles, 2019; Wigfield & Wentzel, 2007; J. Xu et al., 2021). Not all of these intervention studies measure the effects of their intervention on motivational attitudes and on learning success at the same time (see the meta-analyses and reviews of Lazowski & Hulleman, 2016 and Hulleman & Barron, 2016). If there is a causal chain from attitudes to learning success, the effectiveness of an intervention can be assumed (with a certain degree of caution) even if only the effect of the intervention on motivational attitudes has been empirically tested and confirmed. Even if a simultaneous examination of the effects of the intervention on learning success is still highly advisable, the causal relationship thus helps intervention research to make their points.

A third added value of identifying mediators between attitudes towards statistics and learning success lies in the identified mediators themselves, in particular in the mediator of the distribution of learning. Interventions to increase motivation have their limits. The mediator of distribution of learning can then be used to help achieve greater learning success where motivational attitudes cannot be changed or cannot be changed any further. There are numerous empirically tested interventions for improving self-regulation in learning (for reviews and meta-analyses, see for example Dignath, Buettner, & Langfeldt, 2008; Muijs & Bokhove, 2020; Pandey et al., 2018). As a kind of substitute for students with negative attitudes, these can promote learning success in other ways. In many cases, even transparency about the fact that good self-regulation significantly increases learning success even if the overall learning effort is not increased may have a good effect. This way, help can also be provided where motivational attitude interventions are not possible or have already been exhausted.

Now that it has been established that learning engagement and the distribution of learning are appropriate mediators between attitudes toward statistics and learning success, one can return to research question 3a. Using the argument that this thesis makes via Study 6, that beliefs about statistics are related to attitudes towards statistics, a mechanism can now also be built that links beliefs about statistics to learning success using a mediating role of learning behavior. As already discussed for research question 2, the effect sizes in this chain are often small, which makes overall effects across multiple chain elements even smaller. Nevertheless, this thesis shows that beliefs about statistics can be related to learning behavior and, via Study 7, provides reason to suspect that these effects may be larger in longer time periods than in the short term considered.

The significance of these findings discussed for attitudes towards statistics, particularly for interventions, can theoretically be transferred to beliefs about statistics. However, as little is known so far about interventions on beliefs about statistics and their possible effect size, these arguments remain theoretical for the moment.

For the mechanisms between statistics anxiety and learning success from research question 3c, no mediator can be identified in this thesis in study 5. One possible reason for this is the high correlation between statistics anxiety and expectancy. Since expectancy has significant effects, effects that might be significant in a model without attitudes towards statistics but with statistics anxiety might disappear. However, returning to such a model, although the high correlation with attitudes is known, suppresses the question of what behavioral differences could be expected on a theoretical level between students with statistical anxiety and students with negative attitudes towards statistics. Further work is needed on the conceptual differences between negative attitudes towards statistics and statistics anxiety before effects and mechanisms can be clearly separated.

One content-related reason for the lack of effects between statistics anxiety and learning behavior could lie in compensating effects. From the theory of statistics anxiety, it can be assumed that statistics anxiety has a negative effect on learning behavior, for example by increasing procrastination. At the same time, being aware of such anxiety, in particular the facet of test anxiety, in a compulsory course could ensure that learners prevent failure by learning with greater commitment. This way, both effects could cancel each other out either inter-individually or even intra-individually, so that no effect remains visible. Further studies that take a closer look at intended and unintended reactions to statistics anxiety are needed here in order to be able to break down mechanisms in more detail.

11.2.4 Discussion of Research Question 4

> *How do attitudes toward statistics change over time? What are the reasons and mechanisms driving this development?*
> *(RQ 4a: **Development of attitudes**, RQ 4b: **Mechanisms driving this development**)*

This research question is mainly addressed by Study 8. As a first step, this study finds that attitudes towards statistics develop over the duration of the course to a greater extent than a simple pre-post measurement can reveal. It could therefore be important to look at the development of attitudes, both to be able to better control these developments and to be able to better understand the effects of attitudes.

In the second part of the study, this thesis therefore traces the reasons for these developments. Firstly, it identifies factors that are generally shared and relatively stable, in particular the perceived value of statistics for further studies. Such factors can help teachers in statistics. Since the perceived value of statistics for further studies is widely regarded as positive, regular reminders of this attitude by teachers can help to make this facet a formative one that positively influences attitudes overall.

A second group of factors can be identified that do not reflect attitudes towards statistics, but are more specifically linked to the current learning content. In the context of the study, the most influential factor of this kind is that students find the introduction to inferential statistics particularly difficult, which has a negative impact on attitudes in this phase. This finding can serve as a reminder for teachers as to which content needs to be introduced particularly sensitively, and it can also inform statistics education research as to which content could benefit from further content didactical work.

A third group of factors shows itself to be very stable over time, but at the same time heterogeneous between the students. Examples include the perceived value of statistics for a career or for everyday life. This can be taken as an opportunity to develop interventions for those who have a negative attitude in these facets. Heterogeneity in students' attitudes towards statistics can thus be addressed precisely in the right facet. At the same time, the high temporal stability of these facets could ensure particularly sustainable effects.

In addition to the described impulses for teaching practice and statistics education research, which emanate from the answer to research question 4 in this thesis, the findings are also highly relevant for further reflections on the other research results of this thesis. This will be discussed below in a synthesizing conclusion for this thesis.

11.2.5 Conclusions from Synthesizing Across the Research Questions

This work has now answered four research questions. Each answer represents a separate contribution to statistics education research. In addition, some studies in this thesis have answered further research questions that were not directly research questions of this thesis. This has already generated some added value for educational research and practice. The logic of this thesis itself, however, is that not all of these contributions are equally important. As derived in chapter 1, research question 3 is the main research question of this thesis and is therefore clearly superior to the other research questions in the hierarchy within this thesis. Research question 1 and research question 2 served a more preparatory role in the logic of this thesis. Research question 1 enables to address research question 3 in the first place, at least for the area of beliefs about statistics, by conceptualizing them and making them measurable. Research question 2 to a certain extent legitimizes research question 3, at least part 3a on beliefs about statistics, by showing that these are related to learning success, even if this does not appear to be particularly pronounced at first glance. This can significantly strengthen the relevance of research question 3.

Research question 3 then addressed the core of this thesis by identifying learning engagement and the distribution of learning as mediators between attitudes towards statistics and learning success. This provided two significant contributions of this thesis: First, the empirical foundation of learning theories could be supported, here namely the very widely used Expectancy-Value Theory and Control-Value Theory. Secondly, a practical contribution was made by legitimizing interventions aimed at changing motivational attitudes. As such, this thesis provides good arguments for using such interventions and for developing and testing new ones.

Study 8, and thus research question 4, and study 9 then provide critical questions and a need for reassessment of these two contributions, which can also strengthen the contributions to research. The results of research question 4 tend to question the contribution to the empirical foundation of theory, while study 9 questions the legitimizing contribution to interventions.

The crucial point about research question 4 is that its results overturn an assumption that has been implicitly conveyed in all previous parts of this thesis. This assumption is that non-cognitive factors, in particular attitudes towards statistics, are stable over time. Weakening this assumption to some extent, it is at least expected that attitudes towards statistics change only slightly over the period under consideration. Without this assumption, it would be difficult to explain why the attitudes at the beginning of the statistics course should have an influence on learning behavior and learning success throughout the course, as it is difficult to argue for an effect of a long-past attitude on behavior at the moment.

However, research question 4 shows that attitudes change. These changes can be repeated in the same direction or an initial change can be reversed. These changes can be small, but they can also be somewhat larger. Research question 4 thus questions the extent to which the measurement of attitudes at the beginning of the course, which was used in all models for research question 3, can be suitable for capturing the effects of attitudes on learning behavior during the course. Rather, it would make sense to survey attitudes towards statistics very regularly in order to update the current status of students with regard to their attitudes and thus measure the effect of actual current attitudes on learning in the model.

There have already been several attempts to do this, but not yet in the field of statistics education research. In cross-legged panel models and reciprocal effects models, attempts are made to model multiple measures of non-cognitive factors and achievement together in order to identify effects of non-cognitive factors and achievement in t-1 on the respective statuses at time t, for example to show that non-cognitive factors and achievement influence each other (Marsh et al., 2005; Marsh, Guo, Pekrun, Lüdtke, & Núñez-Regueiro, 2024; Pekrun et al., 2017). This way it can also be ensured that at least a more recent measure of attitudes towards statistics is used. Using such modelling also for research question 3 of this thesis therefore seems reasonable and could be a future research question that is justified by the combined results of research questions 3 and 4. However, it must be considered that such models have so far generally only looked at relationships between non-cognitive factors and performance. Integrating learning behavior into the model would require, in particular, a large number of learners in order to estimate this model with sufficient statistical power. In the context of this study, it would not have been possible to achieve this number of cases, nor would the participating students have been willing to complete surveys on their attitudes sufficiently often.

However, instead of seeing the lack of cross-legged panel models and reciprocal effects models as a shortcoming, this can also be seen as a particular strength of this work. Ultimately, the criticism is that the effects of attitudes towards statistics on learning behavior and learning success are only imprecisely estimated and the effect can therefore only be inadequately recorded. The criticism is therefore that the approach chosen in this work may underestimate the effect of non-cognitive factors on learning behavior and learning success.

In this sense, it can be considered very positive that effects were found in the context of research question 3 with the given design. The criticism that may initially arise due to the instability of the measured attitudes can therefore be seen as further potential to investigate these effects further and to find the role of the theories in different granularities of temporal consideration.

After research question 4 was able to enrich the view of research question 3 and the temporal processes behind it, study 9 questions the usefulness of research question 3 for motivational interventions. The discussion of research question 3 had emphasized that the link found between non-cognitive factors and learning success via the mediator learning behavior leads to a causal interpretation that legitimizes motivational interventions, as the causal chain makes it seem possible to transfer the effect on motivational attitudes to learning success.

In study 9, however, emergency remote teaching appears as an intervention that seems to break this causal chain by improving learning behavior and learning success, while the non-cognitive factors decrease. This finding reminds of the fact that interventions often not only have one consequence, but can also be accompanied by (numerous) side effects. These side effects can in turn have subsequent effects that mask the positive relationships between the non-cognitive factors, learning behavior and learning success. The findings from Study 9 should therefore not be understood as a refutation of the positive relationship between non-cognitive factors and learning behavior and learning success. However, they are a reminder that in intervention studies, side effects should always be taken into account and therefore the effect of the intervention on learning success should always be considered, also in intervention studies that increase motivation.

By broadening the perspective on the basis of research question 4 and providing an admonition on the basis of study 9, this thesis thus provides a more comprehensive answer to its main research question than studies 4 and 5 alone would have been able to do.

11.3 Limitations of This Work

The limitations of this thesis are discussed in two categories below. In the first, limitations are discussed that deal with the validity of the statements made within the context in which the thesis was conducted. They therefore concern the internal validity of this thesis. The second part discusses limitations that relate to the generalizability of the results to other contexts, i.e. external validity.

11.3.1 Concerns Regarding Internal Validity

A first group of concerns can be raised with regard to the measurement of all constructs used in this thesis. For example, published measurement instruments were used to measure attitudes towards statistics and statistics anxiety. It has already been shown in Chapter 1 of this thesis that these instruments have been criticized and that they rarely achieve the highest psychometric standards. On the other hand, no better instruments exist (yet). For pragmatic reasons, it was therefore decided in this thesis to use these instruments nonetheless. Their use in this thesis, particularly in Study 5, confirms that the psychometric results are only mediocre. The measurements used could therefore be inaccurate, which is why effect sizes must be viewed with a certain degree of caution. At the same time, however, the measurement properties were not so poor that the measurement should be fundamentally questioned.

A separate measuring instrument was developed to measure beliefs about statistics. This means that there is no evaluation of the instrument by a body external to the developers. The development of the instrument in two different countries with different university cultures and in two different subject areas should go some way towards compensating for this lack of external review. Nevertheless, it remains to be seen how the instrument works when used by others. The quality of the measurement of beliefs about statistics is therefore still somewhat uncertain for the time being.

Learning behavior is measured using digital behavioral data from the learning platform developed in Study 2. However, such measurements come with some disadvantages. Two of these are particularly noteworthy: In terms of the extent of learning, it remains unclear whether all student learning is captured in this approach. This depends above all on the extent to which the digital learning platform in which the digital behavioral traces are created actually accompanies all learning. If, in addition to working with the system, students also learn with printed literature or in discussion with others, for example, this cannot be seen in the digital behavioral traces. However, this disadvantage must be countered by the fact that the data for the main research question was collected in the summer semester of 2020. Physical contact with fellow students and access to libraries, for example, was not possible during this time. These circumstances may have mitigated such background noise in the data. As Study 2 discusses, care was also taken in the development of the software to bring as many parts of learning into the system as possible. This does not eliminate the possibility of missing learning activity in the data, but hopefully greatly reduces it.

A second disadvantage is that the digital behavioral data can hardly provide any information about how cognitively activated or superficially a student has learned. The digital behavioral data only reflects the activity, not its quality. Attempts to better capture this in the future are still in their infancy (see e.g. Gasevic et al., 2017; Ouhaichi, Spikol, & Vogel, 2023). For the time being, therefore, we have to live with the vagueness of this data, which tends to increase the challenge of finding relationships to non-cognitive factors or learning success.

In this study, as in most studies of this kind, the learning success of the students is determined via the course examination. In order to make this somewhat more reliable than usual, the development of the examination was first preceded by a study of the literature on statistical literacy, statistical thinking and statistical reasoning. The examination was developed on this basis in cooperation and discussion with a colleague who had also taught the course involved on several occasions. The result is an exam that the authors are convinced is a high-quality test of statistical competencies at the end of an introductory statistics course. Nevertheless, the exam has not been tested and scaled in terms of measurement theory, nor has it been externally reviewed. Its measurement quality can therefore certainly be questioned. In particular, the creation of such an exam naturally means prioritization decisions that other people could have made differently. However, this would also have been the case with a more comprehensively evaluated instrument. Therefore, instead of developing an in-house test, using an already published test such as the Statistics Reasoning Assessment (Garfield, 2003) would have been possible. But this would have meant less direct connection to the course and therefore not being able to measure all learning outcomes, and in addition it would have meant separating the test and the exam, which would have turned the test into a low-stakes exam, which can also significantly impair the quality of measurement.

With regard to the measurements in the qualitative parts of this work, it should be noted in particular that the reliability of the coding is often not tested. In parts of Study 1 and in Studies 6 and 8, the coding was only carried out by one person, so that it was not possible to check the reliability at all. In other parts of Study 1 and in Studies 2 and 7, coding was carried out by several people, but they were all part of the research team, so that biases are still possible here.

In addition to these various measurement problems, a number of other concerns can be raised about other parts of the study implementation. One example for this are the samples that are used. Firstly, there is the common problem that not all students in the population are willing to participate in surveys or interviews on the selected topics. The selection of those who do take part does not necessarily have to be unbiased. Secondly, in this case it is not clear from the outset who belongs to the population and who does not, as there is no clear system for students to register for courses. Inclusion in the population is therefore defined in this thesis by registering in the learning software from Study 2. However, more students registered for the course in the university's learning management system, but without an account for the system from Study 2, they could not do much. At the same time, there are accounts in the system in which so little activity is recorded that it is questionable whether these people ever saw themselves as course participants. Thirdly, many analyses, especially the analyses for the main research question, are only based on the people who took part in the final exam, as the measurement of success was not defined for all others.

However, those who dropped out of the course during the course can be an interesting part of the population.

Reservations can also be raised about the statistical analyses. In particular, assumption violations should be mentioned here. Like many survey data, these are not always normally distributed. Within the narrow limits of a Likert scale, however, this may not be too much of a problem. The digital behavioral data is more difficult here. As many variables are the result of counting, they are necessarily truncated below zero and usually skewed to the right above it. Some variables additionally contain substantial outliers. In the sense of robustness checks of the statistical results, many models are therefore estimated using both maximum likelihood and robust maximum likelihood. In addition, many tests are performed both parametrically and non-parametrically. In each case, no substantial differences are found. Nevertheless, all models and especially effect sizes in this sense should be viewed with a certain degree of caution.

11.3.2 Concerns Regarding External Validity

With regard to the external validity of this work, special features of the context in which the work was carried out must be taken into account, which may be quite different in other contexts. This is particularly important because, with the exception of Study 7, the studies in this thesis are all conducted in the same context and the particularities of this context therefore have a very wide-ranging impact. A look at a second context as a point of comparison only is used in Study 1.

The first important characteristic of the context of this thesis is that the course under investigation is a compulsory course for about 99% of its participants. In voluntary courses, the relationships and mechanisms described could therefore work very differently, especially since, according to Study 7, non-cognitive factors could influence such choices of courses.

A second important characteristic is the learning and examination culture in the given context. Only a final examination, in this case a written exam, is considered a relevant examination. All achievements and activities along the way are voluntary for the students and are only there to promote the students' learning and to prepare them for these examinations. This gives students a lot of freedom, but also requires increased self-regulatory skills. For this study, this could mean that the effects found in this context are higher than elsewhere, as the high degree of freedom of the students enables attitudinal effects that would not have come to light in more strictly organized systems. At the same time, this learning and examination culture fuels the aforementioned limitation of course dropouts not being taken into account. The less than 300 exams with over 700 accounts in the learning platform that appear in this thesis are common in this culture. In other contexts, it would be necessary to examine whether the effects of non-cognitive factors are perhaps larger if course dropouts can be considered a failure and thus poor performance. In this context there is so much freedom that students can have many reasons to drop out. Counting every drop out as a failure would therefore missjudge many cases.

Another special characteristic of the context is that the participants are social science students. In this context, this primarily means students of sociology, political science, educational research and sports science, as well as some smaller subjects. In contrast to other institutions, psychology is not part of the social sciences. The students in the course are therefore generally not students who have chosen their degree program because of its statistics components. Another thing to say about the institution where the course takes place is that although it is a prestigious university in Germany, the relevant study programs have little or no competitive access policies.

It is essentially unclear what effects these contextual conditions have on the results of this thesis. In some cases, hypotheses can be made about the nature or direction of the effect, as is partly done here, but comparative studies in other contexts are necessary for a well-founded assessment of context dependency.

11.4 Perspectives for Future Research

Like most research projects, this work raises a variety of follow-up questions and provides starting points for future research. Subsequent research can, for example, make use of the presentation of the theoretical background of this thesis and its current state of research in order to take up research questions that are outlined but not considered a priority and are therefore not dealt with in this thesis. Subsequent research can also take the discussion of the results of this thesis as an opportunity to ask follow-up questions to the answers given or to take a closer or more comprehensive look at individual aspects. Likewise, subsequent research can address concerns based on the limitations of this thesis in order to counter them or to discuss their relevance. These different approaches to subsequent research open up so many possibilities that only a few options will be outlined below.

This outline begins with some approaches to addressing the limitations of this work in further research. One of these limitations was the restriction of many analyses, particularly in the studies on the main research question, to students who took part in the final course examination. However, the effects of non-cognitive factors on those others who apparently dropped out of the course would be particularly interesting in order to gain a more complete picture of the consequences of non-cognitive factors. Fortunately, very useful data to address this issue are available, but were excluded from the analyses so far based on the models in Studies 4 and 5. Using this data to investigate whether non-cognitive factors influence the likelihood of dropout and the timing and prior engagement of dropout is therefore an ongoing follow-up project.

Further work on the measurement issues described would also certainly be helpful in order to conduct future research in a valid and theory-based manner. A group around Unfried et al. (2022) is already working on new instruments for measuring attitudes using statistics. No such developments are known for statistics anxiety, although the measurement quality there is at least not better than for attitudes. Of particular importance for this work would of course be studies that deal with the conceptualization developed in Study 1 of this thesis and the associated measurement instrument for beliefs about statistics. This would both test the measurement quality in this thesis and enable many further projects on beliefs about statistics.

On the basis of a further validated instrument, many research questions could then be asked about beliefs about statistics, on which work has already been done for attitudes towards statistics and statistics anxiety. As the overview of the general research questions for beliefs about statistics shows, there are still large gaps here. We still know far too little about both the causes of the beliefs and about their effects. Quantitative as well as further qualitative projects would be necessary to obtain a more complete picture of such causal relationships and their effect sizes. However, the literature review also shows many possibilities for further research with regard to attitudes towards statistics and statistics anxiety. Just to give one example, the question of how large the effects of the various causes for attitudes and anxiety are in comparison to each other still is very open.

The answers given to research questions 2 and 4 of this paper also give cause for follow-up questions. With regard to the effects of beliefs about statistics, the long-term effects in particular need to be investigated. The breadth of the possible consequences under investigation also requires further work. The project started with Study 7 picks up here and aims to trace long-term causal relationships in case studies. To this end, a total of nine students have so far been interviewed at the start of their studies and then after each completed year of study. In the first interviews, the focus of these interviews is on the reasons for the beliefs about statistics that students bring into their studies; in the subsequent interviews, the focus is primarily on the consequences that these beliefs have, particularly for the process of appropriating statistics and for the development of a sense of belonging to statistics.

In response to research question 4, on the one hand an attempt could be made to take a closer look at the relationships of the very precisely broken down facets of attitudes described there. To this end, it would be a natural step to test the qualitatively described process in quantitative panel studies. However, this would first require the development of measurement instruments that adequately measure the individual facets described. Such a measuring instrument could then also be used to examine genesis and impact of these facets on the learning process in more detail.

On the other hand, seeing these developments of attitudes going on much could be done to get control over them. Interventions that specifically target one facet of attitudes could be more or differently effective than more generally targeted interventions. This should at least be examined in order to provide valuable tools for the practice of statistics education.

However, since research question 3 on the mechanisms behind the relationships between non-cognitive factors and learning success is the main research question of this thesis, the follow-up questions derived from its answers appear to be of particular relevance for this thesis. Firstly, there is the issue already mentioned in the conclusions that this thesis was only able to consider the non-cognitive factors in their status at the beginning of the course. A consideration of their development during the course would be a necessary next step in order to understand the role of non-cognitive factors in learning behavior. Cross-legged panel models and reciprocal effect models should be developed and tested, in order to relate non-cognitive factors and learning behavior to each other and to learning success over as many measurement points as possible.

A second important aim would be to expand the observation of learning behavior. In the models of this thesis, learning engagement and the distribution of learning show significant relationships, but no relevant operationalization has yet been developed for the quality of learning. Advances in the fields of educational data mining and multimodal learning analytics should be used here to make more profound layers of learning behavior and cognitive activation measurable in order to incorporate them into such models. This could provide statistics education research and educational research in general with valuable insights into the interplay of affective, organizational and cognitive factors in the learning process.

As a final aspect, it is also to be encouraged for the model proposed in this paper, as well as for all other proposed extensions, to keep the time period under consideration as long as possible in order to ideally see short-term and long-term effects simultaneously in the same model.

12 References

Abbiati, N. N., Del Fabrizio, M. C., López, M. V., Pérez, A., Plencovich, M. C., & Cueto, G. (2021). Attitudes Related to Students´ Performance in Statistics in University Programs in Argentina. *Statistics Education Research Journal*, *20*(2), 1–11. https://doi.org/10.52041/serj.v20i2.356

Abraham, J., & Barker, K. (2015). Exploring Gender Difference in Motivation, Engagement and Enrolment Behaviour of Senior Secondary Physics Students in New South Wales. *Research in Science Education*, *45*(1), 59–73. https://doi.org/10.1007/s11165-014-9413-2

Acee, T. W., & Weinstein, C. E. (2010). Effects of a Value-Reappraisal Intervention on Statistics Students' Motivation and Performance. *The Journal of Experimental Education*, *78*(4), 487–512. https://doi.org/10.1080/00220970903352753

Aiken, L. R. (1970). Attitudes Toward Mathematics. *Review of Educational Research*, *40*(4), 551–596. https://doi.org/10.3102/00346543040004551

Aiken, L. R. (1976). Update on Attitudes and Other Affective Variables in Learning Mathematics. *Review of Educational Research*, *46*(2), 293–311. https://doi.org/10.3102/00346543046002293

American Statistical Association Undergraduate Guidelines Workgroup (2014). *Curriculum Guidelines for Undergraduate Programs in Statistical Science*. Alexandria, VA: American Statistical Association (ASA). Retrieved from https://www.amstat.org/education/curriculum-guidelines-for-undergraduate-programs-in-statistical-science-

Andergassen, M., Mödritscher, F., & Neumann, G. (2014). Practice and Repetition during Exam Preparation in Blended Learning Courses: Correlations with Learning Results. *Journal of Learning Analytics*, *1*(1), 48–74. https://doi.org/10.18608/jla.2014.11.4

Andrews, C. M. (2021). *Improving Psychology Students' Attitudes and Anxiety about Statistics* (Dissertation). University of Wisconsin-Madison. Retrieved from https://www.proquest.com/openview/84d42cc4c0dd4d1b16e9ae51b6d4610a/1?pq-origsite=gscholar&cbl=18750&diss=y

Anke, J. M., & Schumann, M. [Matthias] (2018). E-Learning Modules for Sustainable Management: Potentials and Challenges. In *Proceedings of the 24th Americas Conference on Information Systems*, New Orleans, LA.

Aronson, E., Wilson, T. D., & Sommers, S. R. (2021). *Social psychology* (Tenth edition, global edition). Harlow, England: Pearson. Retrieved from https://ebookcentral.proquest.com/lib/kxp/detail.action?docID=6147495

Artino, A. R., & Jones, K. D. (2012). Exploring the complex relations between achievement emotions and self-regulated learning behaviors in online learning. *The Internet and Higher Education*, *15*(3), 170–175. https://doi.org/10.1016/j.iheduc.2012.01.006

Auzmendi, E. (1991). Factors Related to Attitudes toward Statistics: A Study with a Spanish Sample. *Paper Presented at the Annual Meeting of the American Educational Research Association*. Retrieved from https://eric.ed.gov/?id=ED333049

AXA Shared Services Spain S.A. (2019). axa-group/nlp.js. Retrieved from https://github.com/axa-group/nlp.js

Bailyn, L. (1977). Research as a cognitive process: Implications for data analysis. *Quality and Quantity*, *11*(2), 97–117. https://doi.org/10.1007/BF00151906

Baloglu, M. (1999). *A Comparison of Mathematics Anxiety and Statistics Anxiety in Relation to General Anxiety*. Retrieved from https://eric.ed.gov/?id=ED436703

Baloglu, M. (2004). Statistics Anxiety and Mathematics Anxiety: Some Interesting Differences I. *Educational Research Quarterly*, *27*(3), 38–48. Retrieved from https://eric.ed.gov/?id=EJ792861

Baloğlu, M., Deniz, M. E., & Kesici, Ş. (2011). A descriptive study of individual and cross-cultural differences in statistics anxiety. *Learning and Individual Differences*, *21*(4), 387–391. https://doi.org/10.1016/j.lindif.2011.03.003

Bandura, A. (1977). Self-efficacy: Toward a Unifying Theory of Behavioral Change. *Psychological Review*, *84*(2), 191–215. https://doi.org/10.1037//0033-295x.84.2.191

Bar-Tal, D. (1990). *Group Beliefs: A Conception for Analyzing Group Structure, Processes, and Behavior* (1st ed.). New York: Springer.

Barthakur, A., Kovanovic, V., Joksimovic, S., Siemens, G., Richey, M., & Dawson, S. (2021). Assessing program-level learning strategies in MOOCs. *Computers in Human Behavior*, *117*, 106674. https://doi.org/10.1016/j.chb.2020.106674

Baskerville, R., Baiyere, A., Gregor, S., Hevner, A., & Rossi, M. (2018). Design Science Research Contributions: Finding a Balance between Artifact and Theory. *Journal of the Association for Information Systems*, *19*(5), 358–376.

Batakei, L., Bolon, W., & Bond, M. E. (2018). A Framework and Survey for Measuring Instructors' Motivational Attitudes Toward Statistics. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)* (4J3). International Statistical Institute; International Association for Statistical Education. Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_4J3.pdf?1633129788

Bateiha, S., Marchionda, H., & Autin, M. (2020). Teaching Style and Attitudes: A Comparison of Two Collegiate Introductory Statistics Classes. *Journal of Statistics Education*, *28*(2), 154–164. https://doi.org/10.1080/10691898.2020.1765710

Bayer, T. (2016). *Effects of Guided Project-Based Learning Activities on Students' Attitudes Toward Statistics in an Introductory Statistics Course* (Dissertation). Old Dominion University, Norfolk (VA), United States. Retrieved from https://digitalcommons.odu.edu/stemps_etds/14 https://doi.org/10.25777/9g35-yy29

Bell, J. A. (2008). Statistics anxiety and business statistics: The international student. *Education*, *129*(2), 282–287. Retrieved from link.gale.com/apps/doc/A190331801/AONE?u=anon~42e85b90&sid=googleScholar&xid=2fa98fd0

Bell, P., & Linn, M. C. (2002). Beliefs About Science: How Does Science Instruction Contribute? In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 323–348). Mahwah, N.J.: Lawrence Erlbaum Associates. https://doi.org/10.4324/9781410604316-26

Berens, F. (2018). Attitudes towards Research as a source for negative Statistics Attitudes. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)* (3E1). International Statistical Institute; International Association for Statistical Education. Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3E1.pdf?1531364258

Berens, F. (2022). What Changes Students´ Attitudes? A Qualitative Panel Study on How and Why Attitudes Toward An Introductory Statistics Course Change. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering and Educating Today's Learners in Statistics: Proceedings of the Eleventh International Conference on Teaching Statistics (ICOTS11)*. International Association for Statistical Education. https://doi.org/10.52041/iase.icots11.T3G1

Berens, F., & Findley, K. (under review). The Diamond Model of Statistics: Framing and Measuring Students´ Conceptions about the Field.

Berens, F., Findley, K., & Hobert, S. (under review). Do Students´ Disciplinary Beliefs about Statistics Shape Their Learning Behavior? A Learning Analytics Approach.

Berens, F., Findley, K., & Hobert, S. (2023). How Students' Statistics Beliefs Influence Their Attitudes: A Quantitative and a Qualitative Approach. In G. F. Burrill, L. de Oliveria Souza, & E. Reston (Eds.), *Advances in mathematics education. Research on Reasoning with Data and Statistical Thinking: International Perspectives* (pp. 303–316). [S.l.]: Springer. https://doi.org/10.1007/978-3-031-29459-4_22

Berens, F., Findley, K., Justice, N., & Kinson, C. (2023). Disciplinary Appropriation at The Beginning of a Statistics Major. In E. M. Jones (Ed.), *Fostering Learning of Statistics and Data Science: Proceedings of the Satellite conference of the International Association for Statistical Education (IASE)*. The Hague, Netherlands: ISI/IACE. Retrieved from https://iase-web.org/Webinars.php?p=230711_2000

Berens, F., & Hobert, S. (under reviewa). Learning during COVID-19: (Too) Isolated yet Successful.

Berens, F., & Hobert, S. (under reviewb). The Motivated Are the Successful – But Why? Learning Analytics Shows Affects are Linked to Scaling, Spacing, and Success of Learning.

Berens, F., & Hobert, S. (under reviewc). Spacing Can Be More Effective Than Scaling in Inverted Classrooms: A Learning Analytics Field Study.

Berens, F., Reusch, N., & Hobert, S. (2021). Ein Einstieg in gute Sprechstunden: Leitfaden und Reflexionshilfen für Lehrende ohne Erfahrung. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (F 2.21). Berlin: DUZ Verlags- und Medienhaus GmbH. Retrieved from https://www.nhhl-bibliothek.de/de/handbuch/gliederung/?articleID=3383#/Beitragsdetailansicht/273/3383/Ein-Einstieg-in-gute-Sprechstunden-%25E2%2580%2593-Leitfaden-und-Reflexionshilfen-fuer-Lehrende-ohne-Erfahrung

Betz, N. E. (1978). Prevalence, distribution, and correlates of math anxiety in college students. *Journal of Counseling Psychology*, *25*(5), 441–448. https://doi.org/10.1037/0022-0167.25.5.441

Bieleke, M., Goetz, T., Yanagida, T., Botes, E., Frenzel, A. C., & Pekrun, R. (2023). Measuring emotions in mathematics: The Achievement Emotions Questionnaire-

Mathematics (AEQ-M). *ZDM*, *55*(2), 269–284. https://doi.org/10.1007/s11858-022-01425-8

Bjork, R. A. (1966). *Learning and short-term retention of paired associates in relation to specific sequences of interpresentation intervals* (Dissertation). Stanford University, Stanford (CA).

Bjork, R. A. (1970). Repetition and rehearsal mechanisms in models of short-term memory. In D. A. Norman (Ed.), *Models of Human Memory* (pp. 307–330). Saint Louis: Elsevier Science.

Bjork, R. A., & Allen, T. W. (1970). The spacing effect: Consolidation or differential encoding? *Journal of Verbal Learning and Verbal Behavior*, *9*(5), 567–572. https://doi.org/10.1016/S0022-5371(70)80103-7

Bond, M. E., Perkins, S. N., & Ramirez, C. (2012). Students' Perceptions of Statistics: An Exploration of Attitudes, Conceptualizations, and Content Knowledge of Statistics. *Statistics Education Research Journal*, *11*(2), 6–25.

Bozkurt, A., & Sharma, R. C. (2020). Emergency remote teaching in a time of global crisis due to CoronaVirus pandemic. *Asian Journal of Distance Education*, *15*(1), i–vi. https://doi.org/10.5281/ZENODO.3778083

Bransford, J. D., Brown, A. L., & Cocking, R. R. (2000). *How people learn: Brain, mind, experience, and school* (Expanded ed.,). Washington, DC: National Acad. Press. https://doi.org/10.17226/9853

Breckler, S. J. (1984). Empirical validation of affect, behavior, and cognition as distinct components of attitude. *Journal of Personality and Social Psychology*, *47*(6), 1191–1205. https://doi.org/10.1037//0022-3514.47.6.1191

Briñol, P., & Petty, R. E. (2012). A history of attitudes and persuasion research. In A. W. Kruglanski & W. Stroebe (Eds.), *Handbook of the history of social psychology* (pp. 285–320). New York: Psychology Press.

Budé, L., van de Wiel, M. W. J., Imbos, T., Candel, Math J. J. M., Broers, N. J., & Berger, M. P. F. (2007). Students' Achievements in a Statistics Course in Relation to Motivational Aspects and Study Behaviour. *Statistics Education Research Journal*, *6*(1), 5–21.

Bui, N. H., & Alearo, M. A. (2011). Statistics anxiety and science attitudes: Age, gender, and ethnicity factors. *College Student Journal*, *45*(3), 573. Retrieved from https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype

=crawler&jrnl=01463934&an=66893535&h=20ll3prmschvo9a44%2fic4yomhzy6pqvy3tgkm1mblug9d17wyyr9urx6mcsshgttsywbtuiqrg7q%2bkktu0xnkw%3d%3d&crl=c

Bulmer, M., & Haladyn, J. K. (2011). Life on an Island: A Simulated Population to Support Student Projects in Statistics. *Technology Innovations in Statistics Education*, *5*(1). https://doi.org/10.5070/T551000187

Bulmer, M., & Rolka, K. (2005). The "A4-Project": Statistical World Views expressed through Pictures. In H. L. Chick & J. L. Vincent (Eds.), *Proceedings of the 29th Conference of the International Group for the Psychology of Mathematics Education* (Vol. 2, pp. 193–200). PME.

Carlson, K. A., & Winquist, J. R. (2011). Evaluating an Active Learning Approach to Teaching Introductory Statistics: A Classroom Workbook Approach. *Journal of Statistics Education*, *19*(1). https://doi.org/10.1080/10691898.2011.11889596

Carmona, J., Martínez, R. J., & Sánchez, M. (2005). Mathematical background and attitudes toward statistics in a sample of Spanish college students. *Psychological Reports*, *97*(1), 53–62. https://doi.org/10.2466/pr0.97.1.53-62

Carnell, L. J. (2008). The Effect of a Student-Designed Data Collection Project on Attitudes Toward Statistics. *Journal of Statistics Education*, *16*(1). https://doi.org/10.1080/10691898.2008.11889551

Cashin, S. E., & Elmore, P. B. (2005). The Survey of Attitudes Toward Statistics Scale: A Construct Validity Study. *Educational and Psychological Measurement*, *65*(3), 509–524. https://doi.org/10.1177/0013164404272488

Cepeda, N. J., Pashler, H., Vul, E., Wixted, J. T., & Rohrer, D. (2006). Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin*, *132*(3), 354–380. https://doi.org/10.1037/0033-2909.132.3.354

Chaiken, S., & Stangor, C. (1987). Attitudes and Attitude Change. *Annual Review of Psychology*, *38*(1), 575–630. https://doi.org/10.1146/annurev.ps.38.020187.003043

Chambers, D. W. (1983). Stereotypic images of the scientist: The draw-a-scientist test. *Science Education*, *67*(2), 255–265. https://doi.org/10.1002/sce.3730670213

Chance, B., Wong, J. [Jimmy], & Tintle, N. (2016). Student Performance in Curricula Centered on Simulation-Based Inference: A Preliminary Report. *Journal of Statistics Education*, *24*(3), 114–126. https://doi.org/10.1080/10691898.2016.1223529

Charmaz, K. (2014). Grounded Theory in Global Perspective. *Qualitative Inquiry, 20*(9), 1074–1084. https://doi.org/10.1177/1077800414545235

Chatziralli, I., Ventura, C. V., Touhami, S., Reynolds, R., Nassisi, M., Weinberg, T., . . . Loewenstein, A. (2021). Transforming ophthalmic education into virtual learning during COVID-19 pandemic: A global perspective. *Eye*, *35*(5), 1459–1466. https://doi.org/10.1038/s41433-020-1080-0

Chen, C.-M., & Wu, C.-H. (2015). Effects of different video lecture types on sustained attention, emotion, cognitive load, and learning performance. *Computers & Education, 80*, 108–121. https://doi.org/10.1016/j.compedu.2014.08.015

Chen, T., Tang, W., Lu, Y., & Tu, X. (2014). Rank regression: An alternative regression approach for data with outliers. *Shanghai Archives of Psychiatry, 26*(5), 310–315. https://doi.org/10.11919/j.issn.1002-0829.214148

Cheng, S.-L., & Xie, K. (2021). Why college students procrastinate in online courses: A self-regulated learning perspective. *The Internet and Higher Education*, *50.* https://doi.org/10.1016/j.iheduc.2021.100807

Cheng, S., Xie, K., & Collier, J. (2023). Motivational beliefs moderate the relation between academic delay and academic achievement in online learning environments. *Computers & Education*, *195*, 104724. https://doi.org/10.1016/j.compedu.2023.104724

Chew, P. K. H., & Dillon, D. B. (2014). Statistics Anxiety Update: Refining the Construct and Recommendations for a New Research Agenda. *Perspectives on Psychological Science*, *9*(2), 196–208. https://doi.org/10.1177/1745691613518077

Chew, P. K. H., Dillon, D. B., & Swinbourne, A. L. (2018). An examination of the internal consistency and structure of the Statistical Anxiety Rating Scale (STARS). *PLoS ONE, 13*(3), e0194195. https://doi.org/10.1371/journal.pone.0194195

Chi, M. T. H., & Wylie, R. (2014). The ICAP Framework: Linking Cognitive Engagement to Active Learning Outcomes. *Educational Psychologist*, *49*(4), 219–243. https://doi.org/10.1080/00461520.2014.965823

Chiesi, F., & Bruno, F. (2021). Mean differences and individual changes in nursing students' attitudes toward statistics: The role of math background and personality traits. *Nurse Education in Practice*, *52*, 103043. https://doi.org/10.1016/j.nepr.2021.103043

Chiesi, F., & Primi, C. (2009). Assessing statistics attitudes among college students: Psychometric properties of the Italian version of the Survey of Attitudes toward Statistics

(SATS). *Learning and Individual Differences*, *19*(2), 309–313. https://doi.org/10.1016/j.lindif.2008.10.008

Chiesi, F., & Primi, C. (2010). Cognitive and Non-Cognitive Factors Related to Students' Statistics Achievement. *Statistics Education Research Journal*, *9*(1), 6–26.

Chiesi, F., & Primi, C. (2015). Italian University Students' Achievement in Statistics: The Role of Attitudes and Approaches to Learning. In *ISI World Statistics Congress.* Symposium conducted at the meeting of International Statistical Institute, Rio de Janeiro, Brazil. Retrieved from https://flore.unifi.it/bitstream/2158/1011763/1/Chiesi%20%26%20Primi%20ISI2015_IPS043%20Paper.pdf

Chiesi, F., & Primi, C. (2017). Attitudes and learning approaches: modelling a pathway to achievement in statistics. In *ISI World Statistics Congress.* Symposium conducted at the meeting of International Statistical Institute, Marrakech, Morocco. Retrieved from https://hdl.handle.net/2158/1142296

Chiesi, F., & Primi, C. (2018). What happens when Attitudes toward Statistics Change (Increase vs Descease) during the Course. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)* (7D1). International Statistical Institute; International Association for Statistical Education. Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_7D1.pdf?1531364286

Chiesi, F., Primi, C., Bilgin, A. A., Lopez, M. V., & Del Fabrizio, M. C. (2014). Measuring university students' approaches to learning statistics: a cross-cultural and multilingual version of the ASSIST. In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in Statistics Education: Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)* (4E1). International Association for Statistical Education. Retrieved from https://iase-web.org/icots/9/proceedings/pdfs/ICOTS9_4E1_PRIMI.pdf?1405041634

Cladera, M., Rejón-Guardia, F., Vich-I-Martorell, G. À., & Juaneda, C. (2021). Evolution of undergraduate tourism students' attitudes towards Statistics: pre and post evaluation. *Journal of Teaching in Travel & Tourism*, 1–22.

Coetzee, S., & van der Merwe, P. (2010). Industrial psychology students' attitudes towards statistics. *SA Journal of Industrial Psychology*, *36*(1), 8 pages. Retrieved from http://www.scielo.org.za/scielo.php?pid=s2071-07632010000100009&script=sci_arttext

Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd ed.). Hillsdale (NJ), United States: Erlbaum. Retrieved from https://scholar.google.de/citations?user=_sk1pbeaaaaj&hl=en&oi=sra https://doi.org/10.4324/9780203771587

Cojean, S., & Jamet, E. (2018). The role of scaffolding in improving information seeking in videos. *Journal of Computer Assisted Learning*, *34*(6), 960–969. https://doi.org/10.1111/jcal.12303

Corbin, J. M., & Strauss, A. L. (2008). *Basics of qualitative research: Techniques and procedures for developing grounded theory* (3. ed.). Los Angeles, Calif.: Sage Publ. https://doi.org/10.4135/9781452230153

Corte, E. de, Op’t Eynde, P., & Verschaffel, L. (2002). “Knowing What to Believe”: The Relevance of Students’ Mathematical Beliefs for Mathematics Education. In B. K. Hofer & P. R. Pintrich (Eds.), *Personal epistemology: The psychology of beliefs about knowledge and knowing* (pp. 299–322). Mahwah, N.J.: Lawrence Erlbaum Associates.

Corwyn, R. F., & McGarry, P. P. (2020). An Expectancy Value Theory Predicts Achievement in Undergraduate Statistics Through Academic Delay of Gratification. *Statistics Education Research Journal*, *19*(2), 42–56.

Crawford, J., Butler-Henderson, K., Rudolph, J., Malkawi, B., Glowatz, M., Burton, R., . . . Lam, S. (2020). COVID-19: 20 countries’ higher education intra-period digital pedagogy responses. *Journal of Applied Learning & Teaching*, *3*(1), 9–28. https://doi.org/10.37074/jalt.2020.3.1.7

Crawford, K., Gordon, S., Nicholas, J., & Prosser, M. (1994). Conceptions of mathematics and how it is learned: The perspectives of students entering university. *Learning and Instruction*, *4*(4), 331–345. https://doi.org/10.1016/0959-4752(94)90005-1

Crawford, K., Gordon, S., Nicholas, J., & Prosser, M. (1998). Qualitatively different experiences of learning mathematics at university. *Learning and Instruction*, *8*(5), 455–468. https://doi.org/10.1016/S0959-4752(98)00005-X

Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (4th edition). Thousand Oaks, California: SAGE Publications.

Cruise, R. J., Cash, R. W., & Bolton, D. L. (1985). Development and validation of an instrument to measure statistical anxiety. *Paper Presented at the Meeting of the American Educational Research Association*.

Cui, S., Zhang, J., Guan, D., Zhao, X., & Si, J. (2019). Antecedents of statistics anxiety: An integrated account. *Personality and Individual Differences, 144*, 79–87. https://doi.org/10.1016/j.paid.2019.02.036

Dabas, C. S., Muljana, P. S., & Luo, T. (2021). Female Students in Quantitative Courses: An Exploration of their Motivational Sources, Learning Strategies, Learning Behaviors, and Course Achievement. *Technology, Knowledge and Learning*, 1–29. https://doi.org/10.1007/s10758-021-09552-z

Daley, C. E., & Onwuegbuzie, A. J. (1997). The Role of Multiple Intelligences in Statistics Anxiety. In *Annual Meeting of the Mid-South Educational Research Association*, Memphis (TN), United States. Retrieved from https://eric.ed.gov/?id=ED415272

Daniels, L. M., & Stupnisky, R. H. (2012). Not that different in theory: Discussing the control-value theory of emotions in online learning environments. *The Internet and Higher Education*, *15*(3), 222–226. https://doi.org/10.1016/j.iheduc.2012.04.002

Davidian, M., & Louis, T. A. (2012). Why statistics? *Science*, *336*(6077), 12. https://doi.org/10.1126/science.1218685

Dayupay, J. P., Velos, S. P., Go, M. B., Cababat, F. G., & Golbin, R. A., JR. (2022). College Students’ Motivational Strategies, Study Skills, and Mathematics Performance: A Structural Model. *Journal of Positive School Psychology*, *6*(5), 2246 – 2264. Retrieved from https://journalppw.com/index.php/jpsp/article/view/6290

DeBoer, J., Ho, A. D., Stump, G. S., & Breslow, L. (2014). Changing “Course”: Reconceptualizing Educational Variables for Massive Open Online Courses. *Educational Researcher*, *43*(2), 74–84. https://doi.org/10.3102/0013189X14523038

Deci, E. L., & Ryan, R. M. (1985). *Intrinsic Motivation and Self-Determination in Human Behavior*. *Perspectives in social psychology*. New York: Springer Science+Business Media.

Dempster, M., & McCorry, N. K. (2009). The Role of Previous Experience and Attitudes Toward Statistics in Statistics Assessment Outcomes among Undergraduate Psychology Students. *Journal of Statistics Education*, *17*(2).

Desvars-Larrive, A., Dervic, E., Haug, N., Niederkrotenthaler, T., Chen, J., Di Natale, A., . . . Thurner, S. (2020). A structured open dataset of government interventions in response to COVID-19. *Scientific Data*, *7*(1), 285. https://doi.org/10.1038/s41597-020-00609-9

DeVaney, T. A. (2010). Anxiety and Attitude of Graduate Students in On-Campus vs. Online Statistics Courses. *Journal of Statistics Education*, *18*(1), 20–34.

DeVeaux, R. D., & Velleman, P. F. (2008). Math is Music; Statistics is Literature: or Why are there no six year old Novelists? *Amstat News*, *375*, 54–58. Retrieved from https://www.researchgate.net/profile/Paul-Velleman-2/publication/280114347_Math_is_Music_Statistics_is_Literature/links/55abdd3108aea3d08685e1c5/Math-is-Music-Statistics-is-Literature.pdf

Dew, K. H., Galassi, J. P., & Galassi, M. D. (1984). Math anxiety: Relation with situational test anxiety, performance, physiological arousal, and math avoidance behavior. *Journal of Counseling Psychology*, *31*(4), 580–583. https://doi.org/10.1037/0022-0167.31.4.580

Diamond, A. H., & Stylianides, A. J. (2017). Personal Epistemologies of Statisticians in Academia: An Exploratory Study. *Statistics Education Research Journal*, *16*(2), 335–361. https://doi.org/10.52041/serj.v16i2.195

Dignath, C., Buettner, G., & Langfeldt, H.-P. (2008). How can primary school students learn self-regulated learning strategies most effectively? A meta-analysis on self-regulation training programmes. *Educational Research Review*, *3*(2), 101–129. https://doi.org/10.1016/j.edurev.2008.02.003

Dreger, R. M., & Aiken, L. R. (1957). The identification of number anxiety in a college population. *Journal of Educational Psychology*, *48*(6), 344–351. https://doi.org/10.1037/h0045894

Durak, İ., & Karagöz, Y. (2021). Adaptation of Statistics Anxiety Scale to Turkish: Validity and Reliability Study. *International Journal of Assessment Tools in Education*, *8*(3), 667–683. https://doi.org/10.21449/ijate.863225

Dutton, W. H. (1951). Attitudes of Prospective Teachers toward Arithmetic. *The Elementary School Journal*, *52*(2), 84–90. https://doi.org/10.1086/459310

Earp, M. (2007). *Development and Validation of the Statistics Anxiety Measure* (Dissertation). University of Denver, Denver (CO), United States. Retrieved from http://www.iase-web.org/documents/dissertations/07.Earp.Dissertation.pdf

Eccles, J. S., Adler, T. F., Futterman, R., Goff, S. B., Kaczala, C. M., Meerce, J. L., & Midgley, C. (1983). Expectancies, values and academic behaviors. In J. T. Spence (Ed.), *Achievement and Achievement Motives: Psychological and Sociological Approaches* (pp. 75–146). San Francisco: W. H. Freeman. Retrieved from http://education-webfiles.s3-website-us-west-2.amazonaws.com/arp/garp/articles/ecclesparsons83b.pdf

Eccles, J. S., & Wigfield, A. (2020). From expectancy-value theory to situated expectancy-value theory: A developmental, social cognitive, and sociocultural perspective on motivation. *Contemporary Educational Psychology*, *61*, 101859. https://doi.org/10.1016/j.cedpsych.2020.101859

Eichler, A. (2006). Individuelle Stochastikcurricula von Lehrerinnen und Lehrern. *Journal Für Mathematik-Didaktik*, *27*(2), 140–162. https://doi.org/10.1007/BF03339033

Elder, A. (2023). Statistics Attitudes after using Guided Guided Project-based Learnings as an Andragogical Strategy in a Graduate Statistics Course. *Statistics Education Research Journal*, *22*(3), 4. https://doi.org/10.52041/serj.v22i3.436

Emmioğlu, E. (2011). *A structural equation model examining the relationships among mathematics achievement, attitudes toward statistics, and statistics outcomes* (Dissertation). Middle East Technical University, Ankara, Turkey. Retrieved from https://open.metu.edu.tr/handle/11511/21169

Emmioğlu, E., & Çapa Aydın, Y. (2011). A Meta-Analysis on Students' Attitudes toward Statistics. In *World Statistics Congress.* Symposium conducted at the meeting of International Statistical Institute, Dublin, Ireland.

Emmioğlu, E., & Çapa Aydın, Y. (2012). Attitudes and Achievement in Statistics: A Meta-Analysis Study. *Statistics Education Research Journal*, *11*(2), 95–102. https://doi.org/10.52041/serj.v11i2.332

Entwistle, N. J., & Ramsden, P. (1983). *Understanding student learning*. London: Croom Helm.

Esnard, T. R., Alladin, F. M., & Samlal, K. C. (2021). Prior Mathematics Performance, Statistics Anxiety, Self-Efficacy and Expectations for Performance in Statistics: A Survey of Social Sciences Students in a Caribbean Institution of Higher Education. *Statistics Education Research Journal*, *20*(1), 4. https://doi.org/10.52041/serj.v20i1.98

Estrada, A., & Batanero, C. (2008). Explaining teachers’ attitudes towards statistics. In C. Batanero, G. F. Burrill, C. Reading, & A. Rossman (Eds.), *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics: Challenges for Teaching and Teacher Education: Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference*. International Congress on Mathematical Education; International Association for Statistical Education. https://doi.org/10.52041/srap.08304

Evans, B. R. (2005). *Student attitudes, conceptions, and achievement in introductory undergraduate college statistics* (Dissertation). Temple University, Philadelphia (PA), United States. Retrieved from https://www.proquest.com/openview/d29c34f282f95722b0d9b750a549377d/1?pq-origsite=gscholar&cbl=18750&diss=y

Faber, G., & Drexler, H. (2019). Predicting Education Science Students' Statistics Anxiety: The Role of Prior Experiences within a Framework of Domain-Specific Motivation Constructs. *Higher Learning Research Communications*, *9*(1). Retrieved from https://eric.ed.gov/?id=EJ1221253

Feierabend, R. L. (1960). Review of Research on Psychological Problems in Mathematics Education. In U.S. Department of Health, Education, and Welfare, Office of Education (Ed.), *Research Problems in Mathematics Education: Reports from a Conference on Psychological Problems and Research Methods* (pp. 3–46). U.S. Department of Health, Education, and Welfare, Office of Education.

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2019). A Taxonomy of Social Cues for Conversational Agents. *International Journal of Human-Computer Studies*, *132*, 138–161. https://doi.org/10.1016/j.ijhcs.2019.07.009

Feine, J., Gnewuch, U., Morana, S., & Maedche, A. (2020). Gender Bias in Chatbot Design. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Lecture Notes in Computer Science. Chatbot Research and Design* (Vol. 11970, pp. 79–93). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-39540-7_6

Findley, K. (2022). Navigating a Disciplinary Chasm: The Statistical Perspectives of Graduate Teaching Assistants. *Statistics Education Research Journal*, *21*(1), 12. https://doi.org/10.52041/serj.v21i1.14

Findley, K., & Berens, F. (2020). Assessing the Disciplinary Perspectives of Introductory Statistics Students. In S. S. Karunakaran, Z. Reed, & A. Higgins (Chairs), *Proceedings of the 23rd Annual Conference on Research in Undergraduate Mathematics Education.* Symposium conducted at the meeting of Special Interest Group of the Mathematical Association of America, Boston (MA). Retrieved from https://www.researchgate.net/publication/339712352_Assessing_the_Disciplinary_Perspectives_of_Introductory_Statistics_Students

Findley, K., Justice, N., & Berens, F. (2022). Lois Lane, Superman, and Iron Man: How Perspectives of Statistics Relate to Students' Identities and Career Pursuits. In S. A. Peters, L. Zapata-Cardona, F. Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering and Educating Today's Learners in Statistics: Proceedings of the Eleventh International*

Conference on Teaching Statistics (ICOTS11). International Association for Statistical Education. https://doi.org/10.52041/iase.icots11.t3f1

Findley, K., Justice, N., Kinson, C., & Berens, F. (2025). Why Swipe Right? The Career Interests and Expectations of Incoming Statistics Majors. *Journal of Statistics and Data Science Education*, *33*(1). https://doi.org/10.1080/26939169.2024.2430244

Findley, K., & Kaplan, J. J. (2018). What is Statistics? Examining the Disciplinary Beliefs of Incoming Statistics TAs. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)* (C175). International Statistical Institute; International Association for Statistical Education. Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_C175.pdf?1531364321

Finney, S. J., & Schraw, G. (2003). Self-efficacy beliefs in college statistics courses. *Contemporary Educational Psychology*, *28*(2), 161–186. https://doi.org/10.1016/S0361-476X(02)00015-2

Fishbein, M., & Ajzen, I. (2009). *Predicting and changing behavior: The reasoned action approach*. New York: Psychology Press. https://doi.org/10.4324/9780203838020

Francisco, J. M. (2013). The mathematical beliefs and behavior of high school students: Insights from a longitudinal study. *The Journal of Mathematical Behavior*, *32*(3), 481–493. https://doi.org/10.1016/j.jmathb.2013.02.012

GAISE College Report ASA Revision Committee (2016). *Guidelines for Assessment and Instruction in Statistics Education (GAISE) College Report 2016*. American Statistical Association (ASA). Retrieved from http://www.amstat.org/education/gaise

Gal, I., & Ginsburg, L. (1994). The Role of Beliefs and Attitudes in Learning Statistics: Towards an Assessment Framework. *Journal of Statistics Education*, *2*(2), 7–23. Retrieved from http://ww2.amstat.org/publications/jse/v2n2/gal.html

Gal, I., Ginsburg, L., & Schau, C. (1997). Monitoring attitudes and beliefs in statistics education. In I. Gal & J. B. Garfield (Eds.), *The assessment challenge in statistics education* (pp. 37–51). Amsterdam, Washington, DC: IOS Press.

Garfield, J. B. (2003). Assessing Satistical Reasoning. *Statistics Education Research Journal*, *2*(1), 22–38. https://doi.org/10.52041/serj.v2i1.557

Gasevic, D., Jovanovic, J., Pardo, A., & Dawson, S. (2017). Detecting Learning Strategies with Analytics: Links with Self-reported Measures and Academic Performance. *Journal of Learning Analytics*, *4*(2), 113–128-113–128. https://doi.org/10.18608/jla.2017.42.10

Geisler, S., & Rolka, K. (2021). “That Wasn’t the Math I Wanted to do!”: Students’ Beliefs During the Transition from School to University Mathematics. *International Journal of Science and Mathematics Education*, *19*(3), 599–618. https://doi.org/10.1007/s10763-020-10072-y

Giannakos, M. N. (2013). Exploring the video-based learning research: A review of the literature. *British Journal of Educational Technology*, *44*(6), 191–195. https://doi.org/10.1111/bjet.12070

Gibeau, R.-M., Maloney, E. A., Béland, S., Lalande, D., Cantinotti, M., Williot, A., . . . Cousineau, D. (2023). The correlates of statistics anxiety: Relationships with spatial anxiety, mathematics anxiety and gender. *Journal of Numerical Cognition*, *9*(1), 16–43. https://doi.org/10.5964/jnc.8199

Gijsen, M., Catrysse, L., Maeyer, S. de, & Gijbels, D. (2024). Mapping cognitive processes in video-based learning by combining trace and think-aloud data. *Learning and Instruction*, *90*, 101851. https://doi.org/10.1016/j.learninstruc.2023.101851

Github (2020). ColorlibHQ/AdminLTE. Retrieved from https://github.com/ColorlibHQ/AdminLTE

Glaser, B. G., & Strauss, A. L. (1967). *The Discovery of Grounded Theory: Strategies for Qualitative Research*. *Observations (Chicago, Ill.)*. Aldine. Retrieved from https://books.google.de/books?id=oUxEAQAAIAAJ

Gluckman, M., Vlach, H. A., & Sandhofer, C. M. (2014). Spacing Simultaneously Promotes Multiple Forms of Learning in Children's Science Curriculum. *Applied Cognitive Psychology*, *28*(2), 266–273. https://doi.org/10.1002/acp.2997

González, A., Rodríguez, Y., Faílde, J. M., & Carrera, M. V. (2016). Anxiety in the statistics class: Structural relations with self-concept, intrinsic value, and engagement in two samples of undergraduates. *Learning and Individual Differences*, *45*, 214–221. https://doi.org/10.1016/j.lindif.2015.12.019

Gordon, S. (2004). Understanding Students´ Experience of Statistics in a Service Course. *Statistics Education Research Journal*, *3*(1), 40–59. https://doi.org/10.52041/serj.v3i1.541

Gould, R. (2010). Statistics and the Modern Student. *International Statistical Review*, *78*(2), 297–315. https://doi.org/10.1111/j.1751-5823.2010.00117.x

Graesser, A. C., Cai, Z., Morgan, B., & Wang, L. (2017). Assessment with computer agents that engage in conversational dialogues and trialogues with learners. *Computers in Human Behavior*, *76*, 607–616. https://doi.org/10.1016/j.chb.2017.03.041

Graesser, A. C., Person, N. K., & Magliano, J. P. (1995). Collaborative dialogue patterns in naturalistic one-to-one tutoring. *Applied Cognitive Psychology, 9*(6), 495–522. https://doi.org/10.1002/acp.2350090604

Gravett, K., & Ajjawi, R. (2022). Belonging as situated practice. *Studies in Higher Education, 47*(7), 1386–1396. https://doi.org/10.1080/03075079.2021.1894118

Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS Quarterly*, *37*(2), 337–356.

Gregor, S., & Jones, D. (2007). The Anatomy of a Design Theory. *Journal of the Association for Information Systems*, *8*(5), 312–335.

Griffith, J. D., Adams, L. T., Gu, L. L., Hart, C. L., & Nichols-Whitehead, P. (2012). Students' Attitudes Toward Statistics Across the Disciplines: A Mixed-Methods Approach. *Statistics Education Research Journal*, *11*(2), 45–56.

Grigutsch, S. (1996). *Mathematische Weltbilder von Schülern: Struktur, Entwicklung, Einflußfaktoren.*

Grigutsch, S., Raatz, U., & Törner, G. (1995). *Mathematische Weltbilder bei Lehrern*. *Schriftenreihe des Fachbereichs Mathematik : SM - DU / Universität Duisburg*. Duisburg: FB Mathematik, Univ.

Grolemund, G., & Wickham, H. (2014). A Cognitive Interpretation of Data Analysis. *International Statistical Review*, *82*(2), 184–204. https://doi.org/10.1111/insr.12028

Grossman, J., Lin, Z., Sheng, H., Wei, J. T.-Z., Williams, J. J., & Goel, S. (2019). *MathBot: Transforming Online Resources for Learning Math into Conversational Interactions*. Retrieved from http://logical.ai/story-enabled-intelligence/papers/mathbot.pdf

Gundlach, E., Richards, K. A. R., Nelson, D., & Levesque-Bristol, C. (2015). A Comparison of Student Attitudes, Statistical Reasoning, Performance, and Perceptions for Web-augmented Traditional, Fully Online, and Flipped Sections of a Statistical Literacy Class. *Journal of Statistics Education*, *23*(1), 49–81.

Hammer, D. (1994). Epistemological Beliefs in Introductory Physics. *Cognition and Instruction*, *12*(2), 151–183. https://doi.org/10.1207/s1532690xci1202_4

Hanna, D., & Dempster, M. (2009). The Effect of Statistics Anxiety on Students' Predicted and Actual Test Scores. *The Irish Journal of Psychology, 30*(3-4), 201–209. https://doi.org/10.1080/03033910.2009.10446310

Hanna, D., Shevlin, M., & Dempster, M. (2008). The structure of the statistics anxiety rating scale: A confirmatory factor analysis using UK psychology students. *Personality and Individual Defferences, 45*(1), 65–74. https://doi.org/10.1016/j.paid.2008.02.021

Hannigan, A., Gill, O., & Leavy, A. M. (2013). An investigation of prospective secondary mathematics teachers' conceptual knowledge of and attitudes towards statistics. *Journal of Mathematics Teacher Education*, *16*(6), 427–449. https://doi.org/10.1007/s10857-013-9246-3

Hannula, M. S. (2004). *Affect in mathematical thinking and learning* (Dissertation). University of Turku.

Hannula, M. S. (2006). Affect in Mathematical Thinking and Learning: Towards Integration Of Emotion, Motivation, and Cognition. In J. Maasz (Ed.), *New Mathematics Education Research and Practice* (pp. 209–232). Boston: Brill. https://doi.org/10.1163/9789087903510_019

Hannula, M. S. (2015). Emotions in Problem Solving. In S. J. Cho (Ed.), *Selected Regular Lectures from the 12th International Congress on Mathematical Education* (pp. 269–288). Springer.

Hattie, J. (2015). The applicability of Visible Learning to higher education. *Scholarship of Teaching and Learning in Psychology*, *1*(1), 79–91. https://doi.org/10.1037/stl0000021

Hedges, S. (2017). Statistics Student Performance and Anxiety: Comparisons in Course Delivery and Student Characteristics. *Statistics Education Research Journal*, *16*(1), 320–336.

Hedges, S., & Harkness, S. S. (2017). Is GAISE Evident? College Students' Perceptions of Statistics Classes as "Almost not Math". *Statistics Education Research Journal*, *16*(1), 337–356.

Heinzman, E. (2022). "I Love Math Only if It's Coding": A Case Study of Student Experiences in an Introduction to Data Science Course. *Statistics Education Research Journal*, *21*(2), 5. https://doi.org/10.52041/serj.v21i2.43

Hendel, D. D. (1980). Experiential and Affective Correlates of Math Anxiety in Adult Women. *Psychology of Women Quarterly*, *5*(2), 219–230. https://doi.org/10.1111/j.1471-6402.1980.tb00958.x

Henrie, C. R., Halverson, L. R., & Graham, C. R. (2015). Measuring student engagement in technology-mediated learning: A review. *Computers & Education*, *90*, 36–53. https://doi.org/10.1016/j.compedu.2015.09.005

Herman, J., & Kerby-Helm, A. (2022). Question of the Week: Can a Low-Stakes Assignment Improve Students' Attitudes? *Journal of Statistics and Data Science Education*, *30*(1), 39–44. https://doi.org/10.1080/26939169.2021.2020697

Hevner, A. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems*, *19*(2), 87–92.

Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Systems Research. *Management Information Systems Quarterly*, *28*(1), 75–105.

Hien, H. T., Cuong, P.-N., Le Nam, N. H., Nhung, Ho Le Thi Kim, & Le Thang, D. (2018). Intelligent Assistants in Higher-Education Environments: The FIT-EBot, a Chatbot for Administrative and Learning Support. In *Proceedings of the Ninth International Symposium on Information and Communication Technology* (pp. 69–76). New York: ACM. https://doi.org/10.1145/3287921.3287937

Hobert, S. (2019). Say Hello to 'Coding Tutor'! Design and Evaluation of a Chatbot-based Learning System Supporting Students to Learn to Program. In *ICIS 2019 Proceedings*, Munich, Germany.

Hobert, S., & Berens, F. (2019). Fostering Students' Motivation to Learn Daily on a Voluntary Basis: A Gamified Mobile Learning Approach for Formal Learning Settings. In *International Conference on Information Systems*, Munich, Germany. Retrieved from https://aisel.aisnet.org/icis2019/learning_environ/learning_environ/11

Hobert, S., & Berens, F. (2020). Small Talk Conversations and the Long-Term Use of Chatbots in Educational Settings: Experiences from a Field Study. In A. Følstad, T. Araujo, S. Papadopoulos, E. Lai-Chong Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Lecture Notes in Computer Science: Vol. 11970, Chatbot Research and Design: Third International Workshop, CONVERSATIONS 2019* (pp. 260–272). Springer, Cham. https://doi.org/10.1007/978-3-030-39540-7_18

Hobert, S., & Berens, F. (2021a). Ein digitaler Tutor für teilnehmendenstarke Lehrveranstaltungen. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer, & J. Wildt (Eds.), *Neues Handbuch Hochschullehre* (D 2.15). Berlin: DUZ Verlags- und Medienhaus GmbH. Retrieved from https://www.nhhl-bibliothek.de/de/handbuch/gliederung/?articleID=3313#/Beitragsdetailansicht/255/3313/Ein-digitaler-Tutor-fuer-teilnehmendenstarke-Lehrveranstaltungen

Hobert, S., & Berens, F. (2021b). Learning Analytics for Students: A Field Study on Giving Students Access to Their Own Data. In M. Sahin & D. Ifenthaler (Eds.), *Advances in*

Analytics for Learning and Teaching. Visualizations and Dashboards for Learning Analytics (pp. 213–231). Springer, Cham. https://doi.org/10.1007/978-3-030-81222-5_10

Hobert, S., & Berens, F. (2023). Play to get Instant Feedback: Using Game-Based Apps to Overcome the Feedback Gap in Lectures and Online Teaching. *Communications of the Association for Information Systems*, *53*(1), 491–507. https://doi.org/10.17705/1CAIS.05321

Hobert, S., & Berens, F. (2024). Developing a digital tutor as an intermediary between students, teaching assistants, and lecturers. *Educational Technology Research and Development*, *72*(2), 797–818. https://doi.org/10.1007/s11423-023-10293-2

Hobert, S., & Meyer von Wolff, R. [R.] (2019). Say Hello to Your New Automated Tutor: A Structured Literature Review on Pedagogical Conversational Agents. In *Proceedings of the 14th International Conference on Wirtschaftsinformatik.*

Hoegler, S., & Nelson, M. (2018). The Influence of Anxiety and Self-Efficacy on Statistics Performance: A Path Analysis. *Psi Chi Journal of Psychological Research*, *23*(5), 364–375. https://doi.org/10.24839/2325-7342.JN23.5.364

Hofer, B. K. (2006). Domain specificity of personal epistemology: Resolved questions, persistent issues, new models. *International Journal of Educational Research*, *45*(1-2), 85–95. https://doi.org/10.1016/j.ijer.2006.08.006

Hommik, C., & Luik, P. (2017). Adapting the Survey of Attitudes Towards Statistics (SATS-36) for Estonian Secondary School Students. *Statistics Education Research Journal*, *16*(1), 228–239.

Hong, K., Chew, P., & Dillon, D. B. (2014). Statistics anxiety and the Big Five personality factors. *Procedia - Social and Behavioral Sciences*, *112*, 1177–1186.

Hood, M., Creed, P. A., & Neumann, D. L. (2012). Using the Expectancy Value Model of Motivation to Understand the Relationship Between Student Attitudes and Achievement in Statistics. *Statistics Education Research Journal*, *11*(2), 72–85.

Hooper, D., Coughlan, J., & Mullen, M. R. (2008). Structural Equation Modelling: Guidelines for Determining Model Fit. *Electronic Journal of Business Research Methods*, *6*(1), pp53-60-pp53-60. Retrieved from https://academic-publishing.org/index.php/ejbrm/article/view/1224

Horton, N. J., & Hardin, J. S. (2015). Teaching the Next Generation of Statistics Students to “Think With Data”: Special Issue on Statistics and the Undergraduate Curriculum. *The American Statistician*, *69*(4), 259–265. https://doi.org/10.1080/00031305.2015.1094283

Hu, L. [Li-tze], & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, *6*(1), 1–55. https://doi.org/10.1080/10705519909540118

Hulleman, C. S., & Barron, K. E. (2016). Motivation Interventions in Education: Bridging Theory, Research, and Practice. In L. Corno & E. M. Anderman (Eds.), *Handbook of educational psychology* (3rd ed., pp. 174–185). New York, London: Routledge. https://doi.org/10.4324/9781315688244-22

Huynh, M. (2018). *Bringing the Islands to schools: Improving student attitudes towards statistics* (Dissertation). Royal Melbourne Institute of Technology. Retrieved from https://researchrepository.rmit.edu.au/view/pdfCoverPage?instCode=61RMIT_INST&filePid=13248363650001341&download=true

Intepe, G., & Shearman, D. (2020). Developing statistical understanding and overcoming anxiety via drop-in consultations. *Statistics Education Research Journal*, *19*(1), 149–166.

Johnson, M., & Kuennen, E. (2006). Basic Math Skills and Performance in an Introductory Statistics Course. *Journal of Statistics Education*, *14*(2).

Jones, B. D. (2009). Motivating Students to Engage in Learning: The MUSIC Model of Academic Motivation. *International Journal of Teaching and Learning in Higher Education*, *21*(2), 272–285. Retrieved from https://eric.ed.gov/?id=EJ899315

Jones, B. D., & Carter, D. (2019). Relationships between students' course perceptions, engagement, and learning. *Social Psychology of Education*, *22*(4), 819–839. https://doi.org/10.1007/s11218-019-09500-x

Jones, R. C. (2019). How Do We Change Statistical and Critical Thinking Attitudes in Young People? *Numeracy*, *12*(2). https://doi.org/10.5038/1936-4660.12.2.9

Jovanović, J., Gašević, D., Dawson, S., Pardo, A., & Mirriahi, N. (2017). Learning analytics to unveil learning strategies in a flipped classroom. *The Internet and Higher Education*, *33*, 74–85. https://doi.org/10.1016/j.iheduc.2017.02.001

Justice, N., Morris, S., Henry, V., & Brondos Fry, E. (2020). Paint-By-Number or Picasso? A Grounded Theory Phenomenographical Study of Students' Conceptions of Statistics. *Statistics Education Research Journal*, *19*(2), 76–102. https://doi.org/10.52041/serj.v19i2.111

Kaldo, I., & Hannula, M. S. (2012). Structure of university students' view of mathematics in Estonia. *Nordic Studies in Mathematics Education*, *17*(2), 5–26. Retrieved from http://ncm.gu.se/wp-content/uploads/2020/06/17_2_005026_kaldo.pdf

Kapitanoff, S., & Pandey, C. (2018). Collaborative Testing in Statistics: Group Interaction, Anxiety, and Class Performance. *Statistics Education Research Journal*, *17*(2), 51–67.

Katz, D. (1960). The Functional Approach to the Study of Attitudes. *Public Opinion Quarterly*, *24*(2), 163–204. https://doi.org/10.1086/266945

Keeley, J., Zayac, R., & Correia, C. (2008). Curvilinear Relationships Between Statistics Anxiety and Performance Among Undergraduate Students: Evidence for Optimal Anxiety. *Statistics Education Research Journal*, *7*(1), 4–15.

Kelly, W. P., & Tomhave, W. K. (1985). A Study of Math Anxiety/Math Avoidance in Preservice Elementary Teachers. *The Arithmetic Teacher*, *32*(5), 51–53. https://doi.org/10.5951/AT.32.5.0051

Kerby, A. T., & Wroughton, J. R. (2017). When do Students' Attitudes Change? Investigating Student Attitudes at Midterm. *Statistics Education Research Journal*, *16*(2), 476–486.

Kesici, Ş., Baloğlu, M., & Deniz, M. E. (2011). Self-regulated learning strategies in relation with statistics anxiety. *Learning and Individual Differences*, *21*(4), 472–477. https://doi.org/10.1016/j.lindif.2011.02.006

Kiekkas, P., Panagiotarou, A., Malja, A., Tahirai, D., Zykai, R., Bakalis, N., & Stefanopoulos, N. (2015). Nursing students' attitudes toward statistics: Effect of a biostatistics course and association with examination performance. *Nurse Education Today*, *35*(12), 1283–1288. https://doi.org/10.1016/j.nedt.2015.07.005

Kim, K. R., & Seo, E. H. (2015). The relationship between procrastination and academic performance: A meta-analysis. *Personality and Individual Differences*, *82*, 26–33. https://doi.org/10.1016/j.paid.2015.02.038

Klee, H. L., Buehl, M. M., & Miller, A. D. (2022). Strategies for alleviating students' math anxiety: Control-value theory in practice. *Theory into Practice*, *61*(1), 49–61. https://doi.org/10.1080/00405841.2021.1932157

Kloosterman, P. (2002). Beliefs About Mathematics and Mathematics Learning in the Secondary School: Measurement and Implications for Motivation. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 247–269). Springer, Dordrecht. https://doi.org/10.1007/0-306-47958-3_15

Köller, O. (2001). Mathematical world views and achievement in advanced mathematics in Germany: findings from TIMSS population 3. *Studies in Educational Evaluation*, *27*(1), 65–78. https://doi.org/10.1016/S0191-491X(01)00014-1

Köller, O., Baumert, J., & Neubrand, J. (2000). Epistemologische Überzeugungen und Fachverständnis im Mathematik- und Physikunterricht. In J. Baumert, W. Bos, & R. Lehmann (Eds.), *TIMSS/III Dritte Internationale Mathematik-und Naturwissenschaftsstudie: Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn: Band 2: Mathematische und physikalische Kompetenzen am Ende der gymnasialen Oberstufe*. Opladen: Leske & Budrich. Retrieved from https://pure.mpg.de/pubman/faces/viewitemoverviewpage.jsp?itemid=item_2102613

Kowski, L. E. (2013). Does High School Performance Predict College Math Placement? *Community College Journal of Research and Practice*, *37*(7), 514–527. https://doi.org/10.1080/10668926.2012.754730

Krosnick, J. A., & Petty, R. E. (1995). Attitude Strength: An Overview. In R. E. Petty & J. A. Krosnick (Eds.), *Attitude Strength: Antecedents and Consequences* (pp. 1–24). Psychology Press.

La Hernández de Hera, J. M., Morales-Rodríguez, F. M., Rodríguez-Gobiet, J. P., & Martínez-Ramón, J. P. (2023). Attitudes toward mathematics/statistics, anxiety, self-efficacy and academic performance: An artificial neural network. *Frontiers in Psychology*, *14*, 1214892. https://doi.org/10.3389/fpsyg.2023.1214892

Lalonde, R. N., & Gardner, R. C. (1993). Statistics as a second language? A model for predicting performance in psychology students. *Canadian Journal of Behavioural Science / Revue Canadienne Des Sciences Du Comportement*, *25*(1), 108–125. https://doi.org/10.1037/h0078792

Lang, D., Chen, G., Mirzaei, K., & Paepcke, A. (2020). Is Faster Better? A Study of Video Playback Speed. In C. Rensing & H. Drachsler (Eds.), *Proceedings of the Tenth International Conference on Learning Analytics & Knowledge* (pp. 260–269). Association for Computing Machinery.

Lassig, C. (2022). Attributes of Rigorous Grounded Theory Research and Reporting: Illustrations From a Grounded Theory of Adolescent Creativity. *Educational Researcher*, *51*(2), 98–108. https://doi.org/10.3102/0013189X211069571

Lave, J., & Wenger, E. (1991). *Situated learning: Legitimate peripheral participation* (Reprint). *Learning in doing*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511815355

Lavidas, K., Barkatsas, T., Manesis, D. [Dionisios], & Gialamas, V. (2020). A Structural Equation Model Investigating the Impact of Tertiary Students’ Attitudes Toward Statistics, Perceived Competence at Mathematics, and Engagement on Statistics Performance. *Statistics Education Research Journal*, *19*(2), 27–41.

Lavidas, K., Manesis, D. [Dionysios], & Gialamas, V. (2021). Investigation of the Statistical Anxiety Rating Scale Psychometric Properties with a Sample of Greek Students. *International Journal of Educational Psychology*, *10*(2), 116. https://doi.org/10.17583/ijep.2021.6032

Lazowski, R. A., & Hulleman, C. S. (2016). Motivation Interventions in Education. *Review of Educational Research*, *86*(2), 602–640. https://doi.org/10.3102/0034654315617832

Leavy, A. M., Hannigan, A., & Fitzmaurice, O. (2013). If you're doubting yourself then, what's the fun in that? An exploration of why prospective secondary mathematics teachers perceive statistics as difficult. *Journal of Statistics Education*, *21*(3).

Lederman, N. G., Abd-El-Khalick, F., Bell, R. L., & Schwartz, R. S. (2002). Views of nature of science questionnaire: Toward valid and meaningful assessment of learners' conceptions of nature of science. *Journal of Research in Science Teaching*, *39*(6), 497–521. https://doi.org/10.1002/tea.10034

Lehmann, K., Söllner, M. [Matthias], & Leimeister, J. M. (2016). Increasing Learner Interaction in Large-Scale Lectures by Using a Mobile Learning Application. In D. Parsons (Ed.), *Advances in Mobile and Distance Learning. Mobile and Blended Learning Innovations for Improved Learning Outcomes* (Vol. 49, pp. 102–120). Hershey, PA: IGI Global. https://doi.org/10.4018/978-1-5225-0359-0.ch006

Lenzner, T., Neuert, C., & Otto, W. (2016). *Kognitives Pretesting*. SDM-Survey Guidelines (GESIS Leibniz Institute for the Social Sciences). https://doi.org/10.15465/GESIS-SG_EN_010

Lesser, L. M., & Reyes III, R. (2015). Student reactions to the integration of fun material in a high-anxiety subject: A case study in the teaching of College Introductory Statistics. *Teaching & Learning Journal*, *8*(1), 1–19. Retrieved from https://journals.kpu.ca/index.php/td/article/download/1171/629

Levrini, O., Fantini, P., Tasquier, G., Pecori, B., & Levin, M. (2015). Defining and Operationalizing Appropriation for Science Learning. *Journal of the Learning Sciences, 24*(1), 93–136. https://doi.org/10.1080/10508406.2014.928215

Liao, C.-H., & Wu, J.-Y. (2023). Learning analytics on video-viewing engagement in a flipped statistics course: Relating external video-viewing patterns to internal motivational dynamics and performance. *Computers & Education*, *197*, 104754. https://doi.org/10.1016/j.compedu.2023.104754

Liebendörfer, M., Göller, R., Gildehaus, L., Kortemeyer, J., Biehler, R., Hochmuth, R., . . . Schaper, N. (2022). The role of learning strategies for performance in mathematics courses for engineers. *International Journal of Mathematical Education in Science and Technology*, *53*(5), 1133–1152. https://doi.org/10.1080/0020739X.2021.2023772

Liebendörfer, M., & Schukajlow, S. (2017). Interest development during the first year at university: Do mathematical beliefs predict interest in mathematics? *ZDM – Mathematics Education*, *49*(3), 355–366. https://doi.org/10.1007/s11858-016-0827-3

Liebrecht, C., & van Hooijdonk, C. (2020). Creating Humanlike Chatbots: What Chatbot Developers Could Learn from Webcare Employees in Adopting a Conversational Human Voice. In A. Følstad, T. Araujo, S. Papadopoulos, E. L.-C. Law, O.-C. Granmo, E. Luger, & P. B. Brandtzaeg (Eds.), *Lecture Notes in Computer Science. Chatbot Research and Design* (Vol. 11970, pp. 51–64). Cham: Springer International Publishing. https://doi.org/10.1007/978-3-030-39540-7_4

Liljedahl, P., Oesterle, S., & Bernèche, C. (2012). Stability of beliefs in mathematics education: A critical analysis. *Nordic Studies in Mathematics Education*, *17*(3-4), 101–118.

Liu, S., Onwuegbuzie, A. J., & Meng, L. (2011). Examination of the Score Reliability and Validity of the Statistics Anxiety Rating Scale. *The Journal of Educational Enquiry*, *11*(1). Retrieved from https://ojs.unisa.edu.au/index.php/EDEQ/article/view/662

Lyle, K. B., Bego, C. R., Ralston, P. A. S., & Immekus, J. C. (2022). Spaced Retrieval Practice Imposes Desirable Difficulty in Calculus Learning. *Educational Psychology Review*, *34*(3), 1799–1812. https://doi.org/10.1007/s10648-022-09677-2

Ma, X., & Kishor, N. (1997). Assessing the Relationship between Attitude toward Mathematics and Achievement in Mathematics: A Meta-Analysis. *Journal for Research in Mathematics Education*, *28*(1), 26. https://doi.org/10.2307/749662

MacCallum, R. C., Browne, M. W., & Sugawara, H. M. (1996). Power analysis and determination of sample size for covariance structure modeling. *Psychological Methods*, *1*(2), 130–149. https://doi.org/10.1037/1082-989X.1.2.130

Macfadyen, L. P., & Dawson, S. (2010). Mining LMS data to develop an “early warning system” for educators: A proof of concept. *Computers & Education*, *54*(2), 588–599. https://doi.org/10.1016/j.compedu.2009.09.008

Macher, D., Paechter, M., Papousek, I., Ruggeri, K., Freudenthaler, H. H., & Arendasy, M. (2013). Statistics anxiety, state anxiety during an examination, and academic achievement. *The British Journal of Educational Psychology*, *83*(Pt 4), 535–549. https://doi.org/10.1111/j.2044-8279.2012.02081.x

Maedche, A., Legner, C., Benlian, A., Berger, B., Gimpel, H., Hess, T., . . . Söllner, M. [Matthias] (2019). AI-Based Digital Assistants. *Business & Information Systems Engineering*, *61*(4), 535–544. https://doi.org/10.1007/s12599-019-00600-8

Marsh, H. W., Gerlach, E., Trautwein, U., Lüdtke, O., & Brettschneider, W.-D. (2007). Longitudinal study of preadolescent sport self-concept and performance: Reciprocal effects and causal ordering. *Child Development*, *78*(6), 1640–1656. https://doi.org/10.1111/j.1467-8624.2007.01094.x

Marsh, H. W., Guo, J., Pekrun, R., Lüdtke, O., & Núñez-Regueiro, F. (2024). Cracking Chicken-Egg Conundrums: Juxtaposing Contemporaneous and Lagged Reciprocal Effects Models of Academic Self-Concept and Achievement's Directional Ordering. *Educational Psychology Review*, *36*(2), 1–55. https://doi.org/10.1007/s10648-024-09887-w

Marsh, H. W., Hau, K.-T., & Wen, Z. (2004). In Search of Golden Rules: Comment on Hypothesis-Testing Approaches to Setting Cutoff Values for Fit Indexes and Dangers in Overgeneralizing Hu and Bentler's (1999) Findings. *Structural Equation Modeling: A Multidisciplinary Journal*, *11*(3), 320–341. https://doi.org/10.1207/s15328007sem1103_2

Marsh, H. W., & Martin, A. J. (2011). Academic self-concept and academic achievement: Relations and causal ordering. *The British Journal of Educational Psychology*, *81*(Pt 1), 59–77. https://doi.org/10.1348/000709910X503501

Marsh, H. W., Trautwein, U., Lüdtke, O., Köller, O., & Baumert, J. (2005). Academic self-concept, interest, grades, and standardized test scores: Reciprocal effects models of causal ordering. *Child Development*, *76*(2), 397–416. https://doi.org/10.1111/j.1467-8624.2005.00853.x

Martin, N., Hughes, J., & Fugelsang, J. (2017). The Roles of Experience, Gender, and Individual Differences in Statistical Reasoning. *Statistics Education Research Journal*, *16*(2), 454–475.

Marton, F., & Säljö, R. (1976). On qualitative differences in learning: Outcome and process. *British Journal of Educational Psychology*, *46*(1), 4–11. https://doi.org/10.1111/j.2044-8279.1976.tb02980.x

Mason, G. S., Shuman, T. R., & Cook, K. E. (2013). Comparing the Effectiveness of an Inverted Classroom to a Traditional Classroom in an Upper-Division Engineering Course. *IEEE Transactions on Education*, *56*(4), 430–435. https://doi.org/10.1109/TE.2013.2249066

Mason, L., & Scrivani, L. (2004). Enhancing students' mathematical beliefs: An intervention study. *Learning and Instruction*, *14*(2), 153–176. https://doi.org/10.1016/j.learninstruc.2004.01.002

Mayring, P. (2015). Qualitative Content Analysis: Theoretical Background and Procedures. In A. Bikner-Ahsbahs, C. Knipping, & N. Presmeg (Eds.), *Advances in mathematics education. Approaches to Qualitative Research in Mathematics Education: Examples of Methodology and Methods* (pp. 365–380). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-94-017-9181-6_13

McBreen, M., & Savage, R. (2021). The Impact of Motivational Reading Instruction on the Reading Achievement and Motivation of Students: A Systematic Review and Meta-Analysis. *Educational Psychology Review*, *33*(3), 1125–1163. https://doi.org/10.1007/s10648-020-09584-4

McIntee, S.-E., Goulet-Pelletier, J.-C., Williot, A., Deck-Léger, E., Lalande, D., Cantinotti, M., & Cousineau, D. (2022). (Mal)Adaptive Cognitions as Predictors of Statistics Anxiety. *Statistics Education Research Journal*, *21*(1), 5. https://doi.org/10.52041/serj.v21i1.364

McLeod, D. B. (1992). Research on affect in mathematics education: A reconceptualization. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics* (pp. 575–596). Reston, Va.: National Council of Teachers of Mathematics. Retrieved from https://peterliljedahl.com/wp-content/uploads/affect-mcleod.pdf

McLeod, D. B., & McLeod, S. H. (2002). Synthesis — Beliefs and Mathematics Education: Implications for Learning, Teaching, and Research. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 115–123). Springer, Dordrecht. https://doi.org/10.1007/0-306-47958-3_7

McNeish, D., & Wolf, M. G. (2023). Dynamic fit index cutoffs for confirmatory factor analysis models. *Psychological Methods*, *28*(1), 61–88. https://doi.org/10.1037/met0000425

Meinke, J. (2016). *Individuelle Curricula von Lehrkräften in der Algebra* (Dissertation). Georg-August-Universität Göttingen. Retrieved from https://ediss.uni-goettingen.de/handle/11858/00-1735-0000-0028-87EB-6?show=full

Merkel, A. (2020). An address to the nation by Federal Chancellor Merkel. Retrieved from https://www.bundesregierung.de/breg-en/search/statement-chancellor-1732296

Meyer von Wolff, R. [Raphael], Hobert, S., & Schumann, M. [M.] (2019). How May I Help You? – State of the Art and Open Research Questions for Chatbots at the Digital Workplace. In *Proceedings of the 52th Hawaii International Conference on System Sciences*.

Mikic, F. A., Burguillo, J. C., Llamas, M., Rodriguez, D. A., & Rodriguez, E. (2009). CHARLIE: An AIML-based chatterbot which works as an interface among INES and humans. In *2009 EAEEIE Annual Conference*.

Miles, M. B., Huberman, A. M., & Saldaña, J. (2020). *Qualitative data analysis: A methods sourcebook* (Fourth edition, International student edition). Los Angeles, London, New Delhi, Singapore, Washington DC, Melbourne: SAGE.

Millar, A. M., & White, B. J. G. (2014). How do attitudes change from one stats course to the next? In K. Makar, B. de Sousa, & R. Gould (Eds.), *Sustainability in Statistics Education: Proceedings of the Ninth International Conference on Teaching Statistics (ICOTS9)* (1F2). International Association for Statistical Education. Retrieved from http://icots.info/icots/9/proceedings/pdfs/ICOTS9_1F2_MILLAR.pdf

Mills, J. D. (2004). Students' attitudes toward statistics: Implications for the future. *College Student Journal*, *38*(3), 349–362. Retrieved from https://go.gale.com/ps/i.do?id=GALE%7CA123321892&sid=googleScholar&v=2.1&it=r&linkaccess=abs&issn=01463934&p=AONE&sw=w&userGroupName=anon%7E91b8c4dd&aty=open-web-entry

Miyamoto, Y. R., Coleman, C., Williams, J. J., Whitehill, J., Nesterko, S., & Reich, J. (2015). Beyond Time-on-Task: The Relationship Between Spaced Study and Certification in MOOCs. *Journal of Learning Analytics*, *2*(2), 47–69. https://doi.org/10.18608/jla.2015.22.5

Mji, A. (2009). Differences in university students' attitudes and anxiety about statistics. *Psychological Reports*, *104*(3), 737–744. https://doi.org/10.2466/PR0.104.3.737-744

Mödritscher, F., Andergassen, M., & Neumann, G. (2013). Dependencies between E-Learning Usage Patterns and Learning Results. In S. Lindstaedt & M. Granitzer (Eds.), *Proceedings* *of the 13th International Conference on Knowledge Management and Knowledge Technologies* (pp. 1–8). New York, NY: ACM. https://doi.org/10.1145/2494188.2494206

Molenaar, I., & Wise, A. F. (2022). Temporal Aspects of Learning Analytics: Grounding Analyses in Concepts of Time. In C. Lang, G. Siemens, A. F. Wise, D. Gašević, & A. Merceron (Eds.), *Handbook of Learning Analytics* (2nd ed., pp. 66–76). Society for Learning Analytics Research. https://doi.org/10.18608/hla22.007

Muijs, D., & Bokhove, C. (2020). *Metacognition and Self-Regulation: Evidence Review*. London: Education Endowment Foundation. Retrieved from https://eric.ed.gov/?id=ED612286

Muis, K. R. (2004). Personal Epistemology and Mathematics: A Critical Review and Synthesis of Research. *Review of Educational Research*, *74*(3), 317–377. https://doi.org/10.3102/00346543074003317

Muis, K. R., Bendixen, L. D., & Haerle, F. C. (2006). Domain-Generality and Domain-Specificity in Personal Epistemology Research: Philosophical and Empirical Reflections in the Development of a Theoretical Framework. *Educational Psychology Review*, *18*(1), 3–54. https://doi.org/10.1007/s10648-006-9003-6

Muis, K. R., Franco, G. M., & Gierus, B. (2011). Examining epistemic beliefs across conceptual and procedural knowledge in statistics. *ZDM – Mathematics Education*, *43*(4), 507–519. https://doi.org/10.1007/s11858-011-0337-2

Murphy, D. H., Bjork, R. A., & Bjork, E. L. (2022). Going beyond the spacing effect: Does it matter how time on a task is distributed? *Quarterly Journal of Experimental Psychology*, 17470218221113933. https://doi.org/10.1177/17470218221113933

Mvududu, N. (2003). A Cross-Cultural Study of the Connection between Students' Attitudes toward Statistics and the Use of Constructivist Strategies in the Course. *Journal of Statistics Education*, *11*(3).

Nakata, T., & Suzuki, Y. (2019). Effects of Massing and Spacing on the Learning of Semantically Related and Unrelated Words. *Studies in Second Language Acquisition*, *41*(2), 287–311. https://doi.org/10.1017/S0272263118000219

Nass, C., & Moon, Y. (2000). Machines and Mindlessness: Social Responses to Computers. *Journal of Social Issues*, *56*(1), 81–103. https://doi.org/10.1111/0022-4537.00153

Nass, C., Moon, Y., Fogg, B. J., Reeves, B., & Dryer, D. (1995). Can computer personalities be human personalities? *International Journal of Human-Computer Studies*, *43*(2), 223–239. https://doi.org/10.1006/ijhc.1995.1042

Nasser, F. M. (2004). Structural Model of the Effects of Cognitive and Affective Factors on the Achievement of Arabic- Speaking Pre-service Teachers in Introductory Statistics. *Journal of Statistics Education*, *12*(1), 38–56.

National Center for Science and Engineering Statistics (2021). *Women, minorities, and persons with disabilities in science and engineering: 2021 (Special Report NSF 21-321)*. National Science Foundation. Retrieved from https://ncses.nsf.gov/wmpd

Niculescu, A. C., Tempelaar, D., Leppink, J., Dailey, A., Segers, M., & Gijselaers, W. (2015). Feelings and performance in the first year at university: Learning-related emotions as predictors of achievement outcomes in mathematics and statistics. *Electronic Journal of Research in Educational Psychology*, *13*(3), 432–462. https://doi.org/10.14204/ejrep.37.15012

Nolan, M. M., Beran, T., & Hecker, K. G. (2012). Surveys Assessing Students' Attitudes Toward Statistics: A Systematic Review of Validity and Reliability. *Statistics Education Research Journal*, *11*(2), 103–123.

Nye, B. D., Graesser, A. C., & Hu, X. (2014). AutoTutor and Family: A Review of 17 Years of Natural Language Tutoring. *International Journal of Artificial Intelligence in Education*, *24*(4), 427–469. https://doi.org/10.1007/s40593-014-0029-5

O'Bryant, M., Natesan Batley, P., & Onwuegbuzie, A. J. (2021). Validation of an Adapted Version of the Statistical Anxiety Scale in English and Its Relationship to Attitudes Toward Statistics. *SAGE Open*, *11*(1), 215824402110013. https://doi.org/10.1177/21582440211001378

Ober, T. M., Hong, M. R., Rebouças-Ju, D. A., Carter, M. F., Liu, C., & Cheng, Y. (2021). Linking self-report and process data to performance as measured by different assessment types. *Computers & Education*, *167*, 104188. https://doi.org/10.1016/j.compedu.2021.104188

OECD (2018). *Education at a Glance 2018: OECD Indicators*. Paris: OECD Publishing. https://doi.org/10.1787/eag-2018-en

OECD (2019a). *PISA 2018 Assessment and Analytical Framework*. Paris: OECD Publishing. https://doi.org/10.1787/b25efab8-en

OECD (2019b). Review education policies - Education GPS - OECD: Class size & Student-teacher ratio. Retrieved from https://gpseducation.oecd.org/revieweducationpolicies/#!node=41720&filter=all

Ojennus, D. D. (2016). Assessment of learning gains in a flipped biochemistry classroom. *Biochemistry and Molecular Biology Education*, *44*(1), 20–27. https://doi.org/10.1002/bmb.20926

Onwuegbuzie, A. J. (1999). Statistics Anxiety among African American Graduate Students: An Affective Filter? *Journal of Black Psychology*, *25*(2), 189–209. https://doi.org/10.1177/0095798499025002004

Onwuegbuzie, A. J. (2004). Academic procrastination and statistics anxiety. *Assessment & Evaluation in Higher Education*, *29*(1), 3–19. https://doi.org/10.1080/0260293042000160384

Onwuegbuzie, A. J., Da Ros, D., & Ryan, J. M. (1997). The Components of Statistics Anxiety: A Phenomenological Study. *Focus on Learning Problems in Mathematics*, *19*(4), 11–35. Retrieved from https://eric.ed.gov/?id=ej558838

Onwuegbuzie, A. J., & Wilson, V. A. (2003). Statistics Anxiety: Nature, etiology, antecedents, effects, and treatments: a comprehensive review of the literature. *Teaching in Higher Education*, *8*(2), 195–209. https://doi.org/10.1080/1356251032000052447

Op't Eynde, P., Corte, E. de, & Verschaffel, L. (2002). Framing Students' Mathematics-Related Beliefs: A Quest For Conceptual Clarity And A Comprehensive Categorization. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 13–37). Springer, Dordrecht. https://doi.org/10.1007/0-306-47958-3_2

Osborne, J., Simon, S., & Collins, S. (2003). Attitudes towards science: A review of the literature and its implications. *International Journal of Science Education*, *25*(9), 1049–1079. https://doi.org/10.1080/0950069032000032199

Ouhaichi, H., Spikol, D., & Vogel, B. (2023). Research trends in multimodal learning analytics: A systematic mapping study. *Computers and Education: Artificial Intelligence*, *4*, 100136. https://doi.org/10.1016/j.caeai.2023.100136

Özmen, Z. M., & Baki, A. (2018). The views of undergraduate students about their introductory statistics course process. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)* (3E3). International Statistical Institute; International Association for Statistical Education. Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_3E3.pdf?1531364258

Ozturk, T., & Guven, B. (2016). Evaluating Students' Beliefs in Problem Solving Process: A Case Study. *Eurasia Journal of Mathematics, Science and Technology Education*, *12*(3), 411–429. https://doi.org/10.12973/eurasia.2016.1208a

Paechter, M., Macher, D., Martskvishvili, K., Wimmer, S., & Papousek, I. (2017). Mathematics Anxiety and Statistics Anxiety. Shared but Also Unshared Components and Antagonistic Contributions to Performance in Statistics. *Frontiers in Psychology*, *8*, 1196. https://doi.org/10.3389/fpsyg.2017.01196

Pajares, M. F. (1992). Teachers' Beliefs and Educational Research: Cleaning Up a Messy Construct. *Review of Educational Research*, *62*(3), 307–332. https://doi.org/10.3102/00346543062003307

Pandey, A., Hale, D., Das, S., Goddings, A.-L., Blakemore, S.-J., & Viner, R. M. (2018). Effectiveness of Universal Self-regulation-Based Interventions in Children and Adolescents: A Systematic Review and Meta-analysis. *JAMA Pediatrics*, *172*(6), 566–575. https://doi.org/10.1001/jamapediatrics.2018.0232

Papousek, I., Ruggeri, K., Macher, D., Paechter, M., Heene, M., Weiss, E. M., . . . Freudenthaler, H. H. (2012). Psychometric evaluation and experimental validation of the statistics anxiety rating scale. *Journal of Personality Assessment*, *94*(1), 82–91. https://doi.org/10.1080/00223891.2011.627959

Patulny, R., Bellocchi, A., Olson, R. E., Khorana, S., McKenzie, J., & Peterie, M. (2019). Introduction. In R. Patulny, A. Bellocchi, R. E. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in Late Modernity* (pp. 1–7). Routledge.

Patulny, R., & Olson, R. E. (2019). Emotions in late modernity. In R. Patulny, A. Bellocchi, R. E. Olson, S. Khorana, J. McKenzie, & M. Peterie (Eds.), *Emotions in Late Modernity* (pp. 8–24). Routledge.

Paul, W., & Cunnington, R. C. (2017). An Exploration of Student Attitudes and Satisfaction in a GAISE-Influenced Introductory Statistics Course. *Statistics Education Research Journal*, *16*(2), 487–510.

Pehkonen, E., & Törner, G. (Eds.) (1999). *Mathematical Beliefs and their Impact on Teaching and Learning of Mathematics*. Retrieved from https://www.researchgate.net/profile/Helen-Forgasz/publication/226784267_Measuring_Methematical_Beliefs_and_Their_Impact_on_the_Learning_of_Mathematics_A_New_Approach/links/0046351d6394ed076d000000/Measuring-Methematical-Beliefs-and-Their-Impact-on-the-Learning-of-Mathematics-A-New-Approach.pdf#page=24

Pekrun, R. (2006). The Control-Value Theory of Achievement Emotions: Assumptions, Corollaries, and Implications for Educational Research and Practice. *Educational Psychology Review*, *18*(4), 315–341. https://doi.org/10.1007/s10648-006-9029-9

Pekrun, R. (2021). Self-appraisals and emotions: A generalized control-value approach. In T. Dicke, F. Guay, H. W. Marsh, R. G. Craven, & D. M. McInerney (Eds.), *Self: A Multidisciplinary Concept* (pp. 1–30). Information Age Publishing.

Pekrun, R., Frenzel, A. C., Goetz, T., & Perry, R. P. (2007). The Control-Value Theory of Achievement Emotions: An Integrative Approach to Emotions in Education. In P. A. Schutz & R. Pekrun (Eds.), *Emotion in Education : Educational Psychology* (pp. 13–36). Burlington: Academic Press. https://doi.org/10.1016/B978-012372545-5/50003-4

Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., & Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary Educational Psychology*, *36*(1), 36–48.

Pekrun, R., Goetz, T., Titz, W., & Perry, R. P. (2002). Academic Emotions in Students' Self-Regulated Learning and Achievement: A Program of Qualitative and Quantitative Research. *Educational Psychologist*, *37*(2), 91–105. https://doi.org/10.1207/S15326985EP3702_4

Pekrun, R., Lichtenfeld, S., Marsh, H. W., Murayama, K., & Goetz, T. (2017). Achievement Emotions and Academic Performance: Longitudinal Models of Reciprocal Effects. *Child Development*, *88*(5), 1653–1670. https://doi.org/10.1111/cdev.12704

Pekrun, R., & Stephens, E. J. (2010). Achievement Emotions: A Control-Value Approach. *Social and Personality Psychology Compass*, *4*(4), 238–255. https://doi.org/10.1111/j.1751-9004.2010.00259.x

Peterson, D. J. (2016). The Flipped Classroom Improves Student Achievement and Course Satisfaction in a Statistics Course: A Quasi-Experimental Study. *Teaching of Psychology*, *43*(1), 10–15. https://doi.org/10.1177/0098628315620063

Petty, R. E., Briñol, P., Fabrigar, L. R., & Wegener, D. T. (2019). Attitude Structure and Change. In Finkel, Eli J., Baumeister, Roy F. (Ed.), *Advanced Social Psychology: The State of the Science* (2nd ed., pp. 117–156). Oxford: Oxford University Press Incorporated.

Philipp, R. A. (2007). Mathematics Teachers´ Beliefs and Affect. In F. K. Lester (Ed.), *Second handbook of research on mathematics teaching and learning: a project of the National Council of Teachers of Mathematics* (2nd ed., pp. 257–315). Charlotte, NC: Information Age Pub.

Pi, Z., Hong, J., & Yang, J. (2017). Does instructor's image size in video lectures affect learning outcomes? *Journal of Computer Assisted Learning*, *33*(4), 347–354. https://doi.org/10.1111/jcal.12183

Pierce, R. L. (2006). SATS at Ball State University: Approaches and attitudes. In A. Rossman & B. Chance (Eds.), *Proceedings of the 7th International Conference on Teaching Statistics (ICOTS7)*. International Association for Statistical Education; International Statistical Institute. Retrieved from https://www.stat.auckland.ac.nz/~iase/publications/17/c116.pdf

Posner, M. A. (2011). The Impact of a Proficiency-Based Assessment and Reassessment of Learning Outcomes System on Student Achievement and Attitudes. *Statistics Education Research Journal*, *10*(1), 3–14.

Presmeg, N. (2002). Beliefs About the Nature of Mathematics in the Bridging of Everyday and School Mathematical Practices. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 293–312). Springer, Dordrecht. https://doi.org/10.1007/0-306-47958-3_17

Pretorius, T. B., & Norman, A. M. (1992). Psychometric Data on the Statistics Anxiety Scale for a Sample of South African Students. *Educational and Psychological Measurement*, *52*(4), 933–937. https://doi.org/10.1177/0013164492052004015

Putwain, D. W., & Wood, P. (2022). Anxiety in the mathematics classroom: Reciprocal relations with control and value, and relations with subsequent achievement. *ZDM – Mathematics Education*, 1–14. https://doi.org/10.1007/s11858-022-01390-2

Rabin, L. A., Krishnan, A., Bergdoll, R., & Fogel, J. (2021). Correlates of Exam Performance in an Introductory Statistics Course: Basic Math Skills along with Self-Reported Psychological/Behavioral and Demographic Variables. *Statistics Education Research Journal*, *20*(1), 3. https://doi.org/10.52041/serj.v20i1.97

Ralston, K. (2020). 'Sociologists Shouldn't Have to Study Statistics': Epistemology and Anxiety of Statistics in Sociology Students. *Sociological Research Online*, *25*(2), 219–235. https://doi.org/10.1177/1360780419888927

Ralston, K., Gorton, V., MacInnes, J., Gayle, V., & Crow, G. (2021). Anxious women or complacent men? Anxiety of statistics in a sample of UK sociology undergraduates. *International Journal of Social Research Methodology*, *24*(1), 79–91. https://doi.org/10.1080/13645579.2020.1761186

Ramirez, C., Schau, C., & Emmioglu, E. (2012). The Importance of Attitudes in Statistics Education. *Statistics Education Research Journal*, *11*(2), 57–71.

Reid, A., & Petocz, P. (2002). Students' Conceptions of Statistics: A Phenomenographic Study. *Journal of Statistics Education*, *10*(2). https://doi.org/10.1080/10691898.2002.11910662

Reyes, M. R., Brackett, M. A., Rivers, S. E., White, M., & Salovey, P. (2012). Classroom emotional climate, student engagement, and academic achievement. *Journal of Educational Psychology*, *104*(3), 700–712. https://doi.org/10.1037/a0027268

Richardson, F. C., & Suinn, R. M. (1972). The Mathematics Anxiety Rating Scale: Psychometric data. *Journal of Counseling Psychology*, *19*(6), 551–554. https://doi.org/10.1037/h0033456

Richardson, V. (1996). The Role of Attitudes and Beliefs in Learning to Teach. In J. Sikula (Ed.), *Handbook of research on teacher education: A project of the Association of Teacher Educators* (2nd ed., pp. 102–119). New York, NY u.a: Macmillan. Retrieved from https://www.researchgate.net/profile/virginia-richardson/publication/239666513_the_role_of_attitudes_and_beliefs_in_learning_to_teach/links/572cdb6f08aeb1c73d11b2e2/the-role-of-attitudes-and-beliefs-in-learning-to-teach.pdf

Riel, J., Lawless, K. A., & Brown, S. W. (2018). Timing Matters: Approaches for Measuring and Visualizing Behaviours of Timing and Spacing of Work in Self-Paced Online Teacher Professional Development Courses. *Journal of Learning Analytics*, *5*(1), 25–40-25–40. https://doi.org/10.18608/jla.2018.51.3

Roberts, D. M., & Bilderback, E. W. (1980). Reliability and Validity of a Statistics Attitude Survey. *Educational and Psychological Measurement*, *40*(1), 235–238. https://doi.org/10.1177/001316448004000138

Roberts, D. M., & Saxe, J. E. (1982). Validity of a Statistics Attitude Survey: A Follow-Up Study. *Educational and Psychological Measurement*, *42*(3), 907–912. https://doi.org/10.1177/001316448204200326

Rodriguez, F., Kataoka, S., Janet Rivas, M., Kadandale, P., Nili, A., & Warschauer, M. (2021). Do spacing and self-testing predict learning outcomes? *Active Learning in Higher Education*, *22*(1), 77–91. https://doi.org/10.1177/1469787418774185

Roesken, B., Hannula, M. S., & Pehkonen, E. (2011). Dimensions of students' views of themselves as learners of mathematics. *ZDM – Mathematics Education*, *43*(4), 497–506. https://doi.org/10.1007/s11858-011-0315-8

Rokeach, M. (1972). *Beliefs, attitudes and values: A theory of organization and change* (1. ed., 9. print). San Francisco, Calif.: Jossey-Bass. Retrieved from https://eduq.info/xmlui/handle/11515/11480

Rolka, K., & Bulmer, M. (2005). Picturing Student Beliefs in Statistics. *ZDM – Mathematics Education*, *37*(5), 412-417.

Rosen, J. A., Glennie, E. J., Dalton, B. W., Lennon, J. M., & Bozick, R. N. (2010). *Noncognitive Skills in the Classroom: New Perspectives on Educational Research*. *RTI Press Publication: BK-0004-1009*. Research Triangle Park, NC: RTI Press/RTI International. https://doi.org/10.3768/rtipress.2010.bk.0004.1009

Rosenberg, M. J., & Hovland, C. I. (Eds.) (1960). *Attitude Organization and Change: An Analysis of Consistency among Attitude Components*. New Haven: Yale University Press.

Rosseel, Y. (2012). lavaan: An R Package for Structural Equation Modeling. *Journal of Statistical Software*, *48*(2). https://doi.org/10.18637/jss.v048.i02

Rozgonjuk, D., Kraav, T., Mikkor, K., Orav-Puurand, K., & Täht, K. (2020). Mathematics anxiety among STEM and social sciences students: The roles of mathematics self-efficacy, and deep and surface approach to learning. *International Journal of STEM Education*, *7*(1), 1–11. https://doi.org/10.1186/s40594-020-00246-z

Ruan, S., Jiang, L., Xu, J. [Justin], Tham, B. J.-K., Qiu, Z., Zhu, Y., . . . Landay, J. A. (2019). QuizBot. In S. Brewster, G. Fitzpatrick, A. Cox, V. Kostakos, & A. L. Cox (Eds.), *CHI 2019: Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems : May 4-9, 2019, Glasgow, Scotland, UK* (pp. 1–13). New York, New York: The Association for Computing Machinery. https://doi.org/10.1145/3290605.3300587

Rubin, V. L., Chen, Y., & Thorimbert, L. M. (2010). Artificially intelligent conversational agents in libraries. *Library Hi Tech*, *28*(4), 496–522. https://doi.org/10.1108/07378831011096196

Ruggeri, K., Diaz, C., Kelley, K., Papousek, I., Dempster, M., & Hanna, D. (2008). International Issues in Education. *Psychology Teaching Review*, *14*(2), 65–74. Retrieved from https://eric.ed.gov/?id=EJ876503

Ryan, R. M., & Deci, E. L. (2020). Intrinsic and extrinsic motivation from a self-determination theory perspective: Definitions, theory, practices, and future directions.

Contemporary Educational Psychology, *61*, 101860. https://doi.org/10.1016/j.cedpsych.2020.101860

Saavedra, J. (2020). Educational challenges and opportunities of the Coronavirus (COVID-19) pandemic. Retrieved from https://blogs.worldbank.org/education/educational-challenges-and-opportunities-covid-19-pandemic

Saidi, S. S., & Siew, N. M. (2022). Assessing Secondary School Students' Statistical Reasoning, Attitude towards Statistics, and Statistics Anxiety. *Statistics Education Research Journal*, *21*(1), 1–19. https://doi.org/10.52041/serj.v21i1.67

Saint, J., Fan, Y., Gašević, D., & Pardo, A. (2022). Temporally-focused analytics of self-regulated learning: A systematic review of literature. *Computers and Education: Artificial Intelligence*, *3*, 100060. https://doi.org/10.1016/j.caeai.2022.100060

Saldaña, J. (2003). *Longitudinal qualitative research: Analyzing change through time*. AltaMira Press.

Sanchez-Gordon, S., & Luján-Mora, S. (2018). Technological Innovations in Large-Scale Teaching: Five Roots of Massive Open Online Courses. *Journal of Educational Computing Research*, *56*(5), 623–644. https://doi.org/10.1177/0735633117727597

Saqr, M., & López-Pernas, S. (2021). The longitudinal trajectories of online engagement over a full program. *Computers & Education*, *175*, 104325. https://doi.org/10.1016/j.compedu.2021.104325

Schau, C. (2003). *Survey of Attitudes Toward Statistics (SATS-36)*. http://evaluationandstatistics.com/.

Schau, C., & Emmioğlu, E. (2012). Do Introductory Statistics Courses in the United States Improve Students' Attitudes? *Statistics Education Research Journal*, *11*(2), 86–94.

Schau, C., Stevens, J., Dauphinee, T. L., & Del Vecchio, A. (1995). The Development and Validation of the Survey of Antitudes toward Statistics. *Educational and Psychological Measurement*, *55*(5), 868–875. https://doi.org/10.1177/0013164495055005022

Schibeci, R. A. (1984). Attitudes to Science: An Update. *Studies in Science Education*, *11*(1), 26–59. https://doi.org/10.1080/03057268408559913

Schiefele, U., Krapp, A., & Winteler, A. (1992). Interest as a predictor of academic achievement: A meta-analysis of research. In K. A. Renninger, S. Hidi, & A. Krapp (Eds.), *The Role of interest in Learning and Development* (pp. 183–212). Hoboken: Taylor and Francis.

Schoenfeld, A. H. (1985). *Mathematical problem solving*. Orlando, San Diego, New York, London, Toronto, Montreal, Sydney, Tokyo: Academic Press Inc. https://doi.org/10.1016/C2013-0-05012-8

Schoenfeld, A. H. (1992). Learning to Think Mathematically: Problem Solving, Metacognition, and Sense Making in Mathematics. In D. Grouws (Ed.), *Handbook of Research on Mathematics Teaching and Learning: A Project of the National Council of Teachers of Mathematics* (pp. 334–370). Reston, Va.: National Council of Teachers of Mathematics.

Schoenfeld, A. H. (2015). What counts, when? Reflection on beliefs, affect, attitude, orientations, habits of mind, grain size, time scale, context, theory, and method. In B. Pepin & B. Roesken-Winter (Eds.), *From beliefs to dynamic affect systems in mathematics education: Exploring a mosaic of relationships and interactions* (pp. 395–404). Springer.

Schubert, S., Pekrun, R., & Ufer, S. (2023). The role of epistemic emotions in undergraduate students' proof construction. *ZDM – Mathematics Education*, *55*(2), 299–314. https://doi.org/10.1007/s11858-022-01413-y

Schukajlow, S., Rakoczy, K., & Pekrun, R. (2023). Emotions and motivation in mathematics education: Where we are today and where we need to go. *ZDM – Mathematics Education*, 1–19. https://doi.org/10.1007/s11858-022-01463-2

Scott, J. S. (2001). *Modeling aspects of students' attitudes and performance in an undergraduate introductory statistics course* (Dissertation). University of Georgia, Athens (GA), United States. Retrieved from https://www.proquest.com/openview/8cf8d00263ff1d25d5c9f1f4249355f3/1?pq-origsite=gscholar&cbl=18750&diss=y

Sesé, A., Jiménez, R., Montaño, J. J., & Palmer, A. (2015). Can Attitudes toward Statistics and Statistics Anxiety Explain Students' Performance? *Revista De Psicodidáctica*, *20*(2), 285–304. Retrieved from https://ojs.ehu.eus/index.php/psicodidactica/article/view/13080

Siemens, G., & Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE Review*, *46*(5), 30–40.

Singh, K., Granville, M., & Dika, S. (2002). Mathematics and Science Achievement: Effects of Motivation, Interest, and Academic Engagement. *The Journal of Educational Research*, *95*(6), 323–332. https://doi.org/10.1080/00220670209596607

Sizemore, O. J., & Lewandowski, G. W. (2009). Learning Might Not Equal Liking: Research Methods Course Changes Knowledge but Not Attitudes. *Teaching of Psychology*, *36*(2), 90–95. https://doi.org/10.1080/00986280902739727

Smith, T. (2017). Gamified Modules for an Introductory Statistics Course and Their Impact on Attitudes and Learning. *Simulation & Gaming*, *48*(6), 832–854. https://doi.org/10.1177/1046878117731888

Sommerfeld, A. (2011). Recasting Non-Cognitive Factors in College Readiness as What They Truly Are: Non-Academic Factors. *Journal of College Admission*. Retrieved from https://eric.ed.gov/?id=EJ962511

Song, D., Oh, E. Y., & Rice, M. (2017). Interacting with a conversational agent system for educational purposes in online courses. In *2017 10th International Conference on Human System Interactions (HSI),* Ulsan, South Korea.

Songsore, E., & White, B. J. G. (2018). Students' Perceptions of the Future Relevance of Statistics After Completing an Online Introductory Statistics Course. *Statistics Education Research Journal*, *17*(2), 120–140.

Southerland, S. A., Sinatra, G. M., & Matthews, M. R. (2001). Belief, Knowledge, and Science Education. *Educational Psychology Review*, *13*(4), 325–351. Retrieved from http://www.jstor.org/stable/23363496

Spencer, D., Griffith, E., Briska, K., Post, J., & Willis, C. (2023). The Role of Non-Cognitive Factors in the Introductory Statistics Classroom. *Statistics Education Research Journal*, *22*(1), 4. https://doi.org/10.52041/serj.v22i1.77

Stanisavljevic, D., Trajkovic, G., Marinkovic, J., Bukumiric, Z., Cirkovic, A., & Milic, N. (2014). Assessing attitudes towards statistics among medical students: Psychometric properties of the Serbian version of the Survey of Attitudes Towards Statistics (SATS). *PLoS ONE*, *9*(11), e112567. https://doi.org/10.1371/journal.pone.0112567

Steel, P. (2007). The nature of procrastination: A meta-analytic and theoretical review of quintessential self-regulatory failure. *Psychological Bulletin*, *133*(1), 65–94. https://doi.org/10.1037/0033-2909.133.1.65

Steiger, J. H. (2007). Understanding the limitations of global fit assessment in structural equation modeling. *Personality and Individual Differences*, *42*(5), 893–898. https://doi.org/10.1016/j.paid.2006.09.017

Steinberger, P. (2020). Assessing the Statistical Anxiety Rating Scale as applied to prospective teachers in an Israeli Teacher-Training College. *Studies in Educational Evaluation*, *64*, 100829. https://doi.org/10.1016/j.stueduc.2019.100829

Suanpang, P., Petocz, P., & Kalceff, W. (2004). Student attitudes to learning business statistics: Comparison of online and traditional methods. *Journal of Educational Technology & Society*, *7*(3), 9–20. Retrieved from https://www.jstor.org/stable/10.2307/jeductechsoci.7.3.9

Sun, L. [Lihui], Hu, L. [Linlin], & Zhou, D. (2022). Programming attitudes predict computational thinking: Analysis of differences in gender and programming experience. *Computers & Education*, *181*, 104457. https://doi.org/10.1016/j.compedu.2022.104457

Sun, L. [Litao], Tang, Y., & Zuo, W. (2020). Coronavirus pushes education online. *Nature Materials*, *19*(6), 687. https://doi.org/10.1038/s41563-020-0678-8

Taber, K. S. (2018). The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Research in Science Education*, *48*(6), 1273–1296. https://doi.org/10.1007/s11165-016-9602-2

Tait, H., Entwistle, N. J., & McCune, V. (1998). ASSIST: A reconceptualisation of the Approaches to Studying Inventory. In C. Rust (Ed.), *Improving students as learners* (pp. 262–271). Oxford: Oxford Brookes University, The Oxford Centre for Staff and Learning Development.

Tegos, S., Demetriadis, S., & Karakostas, A. (2011). MentorChat: Introducing a Configurable Conversational Agent as a Tool for Adaptive Online Collaboration Support. In *2011 15th Panhellenic Conference on Informatics*.

Tempelaar, D. T. (2023). Incorporating Time in Dispositional Learning Analytics Models. In *Open and Inclusive Educational Practice in the Digital World* (pp. 29–45). Springer, Cham. https://doi.org/10.1007/978-3-031-18512-0_3

Tempelaar, D. T., Gijselaers, W. H., & van der Loeff, S. S. (2006). Puzzles in Statistical Reasoning. *Journal of Statistics Education*, *14*(1). https://doi.org/10.1080/10691898.2006.11910576

Tempelaar, D. T., Nguyen, Q., & Rienties, B. (2020). Learning Analytics and the Measurement of Learning Engagement. In *Adoption of Data Analytics in Higher Education Learning and Teaching* (pp. 159–176). Springer, Cham. https://doi.org/10.1007/978-3-030-47392-1_9

Tempelaar, D. T., & Nijhuis, J. (2007). Commonalities in Attitudes and Beliefs Toward Different Academic Subjects. In M. K. McCuddy, H. van den Bosch, W. B. Martz, A. V. Matveev, & K. O. Morse (Eds.), *The Challenges of Educating People to Lead in a Challenging World* (pp. 225–249). Springer, Dordrecht. https://doi.org/10.1007/978-1-4020-5612-3_11

Tempelaar, D. T., Rienties, B., & Nguyen, Q. (2020). Subjective data, objective data and the role of bias in predictive modelling: Lessons from a dispositional learning analytics application. *PLoS ONE, 15*(6), e0233977. https://doi.org/10.1371/journal.pone.0233977

Tempelaar, D. T., van der Loeff, S. S., & Gijselaers, W. (2007). A Structural Equation Model Analyzing the Relationship of Students' Attitudes Toward Statistics, Prior Reasoning Abilities and Course Performance. *Statistics Education Research Journal*, *6*(2), 78–102.

Törner, G. (2002). Mathematical Beliefs — A Search for a Common Ground: Some Theoretical Considerations on Structuring Beliefs, Some Research Questions, and Some Phenomenological Observations. In G. C. Leder, E. Pehkonen, & G. Törner (Eds.), *Beliefs: A Hidden Variable in Mathematics Education?* (pp. 73–94). Springer, Dordrecht. https://doi.org/10.1007/0-306-47958-3_5

Törner, G., & Grigutsch, S. (1994). „Mathematische Weltbilder" bei Studienanfängern: eine Erhebung. *Journal Für Mathematik-Didaktik*, *15*(3-4), 211–251. https://doi.org/10.1007/BF03338808

Törner, G., & Pehkonen, E. (1996). On the structure of mathematical belief systems. *ZDM – Mathematics Education*, *28*(4), 109–112. Retrieved from https://www.emis.de/journals/ZDM/zdm964a.html#peh2

Trassi, A. P., Leonard, S. J., Rodrigues, L. D., Rodas, J. A., & Santos, F. H. (2022). Mediating factors of statistics anxiety in university students: A systematic review and meta-analysis. *Annals of the New York Academy of Sciences*, *1512*(1), 76–97. https://doi.org/10.1111/nyas.14746

Tsai, C.-C., Jessie Ho, H. N., Liang, J.-C., & Lin, H.-M. (2011). Scientific epistemic beliefs, conceptions of learning science and self-efficacy of learning science among high school students. *Learning and Instruction*, *21*(6), 757–769. https://doi.org/10.1016/j.learninstruc.2011.05.002

Unfried, A., Whitaker, D., Batakei, L., Bolon, W., Bond, M. E., Kerby-Helm, A., & Posner, M. A. (2022). The big picture: A family of instruments for understanding university-level statistics and data science attitudes. In S. A. Peters, L. Zapata-Cardona, F.

Bonafini, & A. Fan (Eds.), *Bridging the Gap: Empowering and Educating Today's Learners in Statistics: Proceedings of the Eleventh International Conference on Teaching Statistics (ICOTS11)*. International Association for Statistical Education. Retrieved from http://iase-web.org/icots/11/proceedings/pdfs/ICOTS11_263_UNFRIED.pdf?1669865546

Vahedi, S., Farrokhi, F., Gahramani, F., & Issazadegan, A. (2012). The relationship between procrastination, learning strategies and statistics anxiety among Iranian college students: A canonical correlation analysis. *Iranian Journal of Psychiatry and Behavioral Sciences*, *6*(1), 40–46. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3939948/

Valle, N., Antonenko, P., Valle, D., Dawson, K., Huggins-Manley, A. C., & Baiser, B. (2021). The influence of task-value scaffolding in a predictive learning analytics dashboard on learners' statistics anxiety, motivation, and performance. *Computers & Education*, *173*, 104288. https://doi.org/10.1016/j.compedu.2021.104288

Van Es, C., & Weaver, M. M. (2018). Race, Sex, and their Influences on Introductory Statistics Education. *Journal of Statistics Education*, *26*(1), 48–54. https://doi.org/10.1080/10691898.2018.1434426

Van Griethuijsen, R. A. L. F., van Eijck, M. W., Haste, H., Brok, P. J. den, Skinner, N. C., Mansour, N., . . . BouJaoude, S. (2015). Global Patterns in Students' Views of Science and Interest in Science. *Research in Science Education*, *45*(4), 581–603. https://doi.org/10.1007/s11165-014-9438-6

Vanhoof, S. (2010). *Statistics attitudes in university students: Structure, stability and relationship with achievement* (Dissertation). Katholieke Universiteit Leuven, Leuven, Belgien. Retrieved from https://lirias.kuleuven.be/bitstream/123456789/289258/1/Doctoral+dissertation+Stijn+Vanhoof.pdf

Vanhoof, S., Castro Sotos, A. E., Onghena, P., Verschaffel, L., van Dooren, W., & van den Noortgate, W. (2006). Attitudes Toward Statistics and Their Relationship with Short- and Long-Term Exam Results. *Journal of Statistics Education*, *14*(3).

Vanhoof, S., Kuppens, S., Castro Sotos, A. E., Verschaffel, L., & Onghena, P. (2011). Measuring Statistics Attitudes: Structure of the Survey of Attitudes Toward Statistics (SATS-36). *Statistics Education Research Journal*, *10*(1), 35–51.

Verhoeven, P. S. (2009). *Quality in Statistics Education: Determinants of Student Outcomes in Methods & Statistics Education at Universities and Colleges* (Dissertation). Utrecht

University, Utrecht, Netherlands. Retrieved from https://www.academia.edu/download/77031496/09.Verhoeven.Dissertation.pdf

Vigil-Colet, A., Lorenzo-Seva, U., & Condon, L. (2008). Development and validation of the Statistical Anxiety Scale. *Psicothema*, *20*(1), 174–180. Retrieved from https://www.redalyc.org/pdf/727/72720124.pdf

Vlach, H. A., & Sandhofer, C. M. (2012). Distributing learning over time: The spacing effect in children's acquisition and generalization of science concepts. *Child Development*, *83*(4), 1137–1144. https://doi.org/10.1111/j.1467-8624.2012.01781.x

Wallace, R. S. (2009). The Anatomy of A.L.I.C.E. In R. Epstein, G. Beber, & G. Roberts (Eds.), *Parsing the Turing Test: Philosophical and Methodological Issues in the Quest for the Thinking Computer* (Vol. 7, pp. 181–210). Dordrecht: Springer Netherlands. https://doi.org/10.1007/978-1-4020-6710-5_13

Watson, F. S., Lang, Thomas, R., & Kromrey, J. D. (2002). Breaking Ground for EncStat: A Statistics Anxiety Intervention Program. In *Annual Meeting of the Florida Educational Research Association.* Symposium conducted at the meeting of Florida Educational Research Association, Gainesville (FL), United States. Retrieved from https://eric.ed.gov/?id=ED477936

Watson, F. S., Lang, Thomas, R., Kromrey, J. D., Ferron, J. M., Hess, M. R., & Hogarty, K. Y. (2003). An Assessment Blueprint for EncStat: A Statistics Anxiety Intervention Program. In *Annual Meeting of the American Educational Research Association.* Symposium conducted at the meeting of American Educational Research Association, Chicago (IL), United States. Retrieved from https://eric.ed.gov/?id=ED477939

Weeden, K. A., & Cornwell, B. (2020). The Small-World Network of College Classes: Implications for Epidemic Spread on a University Campus. *Sociological Science*, *7*(9), 222–241. https://doi.org/10.15195/v7.a9

Wegener, S., Wang, H.-C., Beyersmann, E., Nation, K., Colenbrander, D., & Castles, A. (2022). The effects of spacing and massing on children's orthographic learning. *Journal of Experimental Child Psychology*, *214*, 105309. https://doi.org/10.1016/j.jecp.2021.105309

Weizenbaum, J. (1966). ELIZA - a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

Whitaker, D. (2020). Epistemic and Mathematical Beliefs of Exemplary Statistics Teachers. *International Journal for Mathematics Teaching and Learning*, *21*(2), 119–142. Retrieved from https://www.cimt.org.uk/ijmtl/index.php/IJMTL/article/view/253

Whitaker, D., Unfried, A., & Batakei, L. (2018). A Framework and Survey for Measuring Students' Motivational Attitudes Toward Statistics. In M. A. Sorto, A. White, & L. Guyot (Eds.), *Looking back, looking forward: Proceedings of the Tenth International Conference on Teaching Statistics (ICOTS10)* (C200). International Statistical Institute; International Association for Statistical Education. Retrieved from https://iase-web.org/icots/10/proceedings/pdfs/ICOTS10_C200.pdf?1531364322

Whitaker, D., Unfried, A., & Bond, M. E. (2022). Challenges Associated with Measuring Attitudes Using the SATS Family of Instruments. *Statistics Education Research Journal*, *21*(1), 1–23. https://doi.org/10.52041/serj.v21i1.88

Wiberg, M. (2009). Teaching Statistics in Integration with Psychology. *Journal of Statistics Education*, *17*(1). https://doi.org/10.1080/10691898.2009.11889509

Wigfield, A., & Eccles, J. S. (2000). Expectancy-Value Theory of Achievement Motivation. *Contemporary Educational Psychology*, *25*(1), 68–81. https://doi.org/10.1006/ceps.1999.1015

Wigfield, A., Eccles, J. S., Fredricks, J. A., Simpkins, S., Roeser, R. W., & Schiefele, U. (2015). Development of Achievement Motivation and Engagement. In M. E. Lamb & R. M. Lerner (Eds.), *Handbook of Child Psychology and Developmental Science* (7th ed., Vol. 3, pp. 657–700). John Wiley & Sons, Ltd. https://doi.org/10.1002/9781118963418.childpsy316

Wigfield, A., Turci Faust, L., Cambria, J., & Eccles, J. S. (2019). Motivation in Education. In R. M. Ryan (Ed.), *Oxford handbooks online. The Oxford handbook of human motivation* (pp. 443–462). New York: Oxford University Press.

Wigfield, A., & Wentzel, K. R. (2007). Introduction to Motivation at School: Interventions That Work. *Educational Psychologist*, *42*(4), 191–196. https://doi.org/10.1080/00461520701621038

Wikipedia Contributors (2020). COVID-646 19 pandemic in Germany. Retrieved from https://en.wikipedia.org/w/index.php?title=COVID-2019_pandemic_in_Germany&oldid=980237864

Wild, C. J. (2017). Statistical Literacy as the Earth Moves. *Statistics Education Research Journal*, *16*(1), 31–37. https://doi.org/10.52041/serj.v16i1.211

Wild, C. J., & Pfannkuch, M. (1999). Statistical Thinking in Empirical Enquiry. *International Statistical Review*, *67*(3), 223–248. https://doi.org/10.1111/j.1751-5823.1999.tb00442.x

Wild, C. J., Utts, J. M., & Horton, N. J. (2018). What Is Statistics? In D. Ben-Zvi, K. Makar, & J. B. Garfield (Eds.), *International Handbook of Research in Statistics Education* (pp. 5–36). Springer, Cham. https://doi.org/10.1007/978-3-319-66195-7_1

Wilensky, U. (1997). What is Normal Anyway? Therapy for Epistemological Anxiety. *Educational Studies in Mathematics*, *33*(2), 171–202. https://doi.org/10.1023/A:1002935313957

Williams, A. S. (2010). Statistics Anxiety and Instructor Immediacy. *Journal of Statistics Education*, *18*(2).

Williams, A. S. (2013). Worry, Intolerance of Uncertainty, and Statistics Anxiety. *Statistics Education Research Journal*, *12*(1), 48–59.

Williams, A. S. (2015). Statistics Anxiety and Worry: The Roles of Worry Beliefs, Negative Problem Orientation, and Cognitive Avoidance. *Statistics Education Research Journal*, *14*(2), 53–75.

Winkler, R. [R.], & Söllner, M. [M.] (2018). Unleashing the Potential of Chatbots in Education: A State-Of-The-Art Analysis. In *Academy of Management Annual Meeting (AOM)*. Chicago, USA.

Winkler, R. [R.], & Söllner, M. [M.] (2020). Towards Empowering Educators to Create their own Smart Personal Assistants. In *Proceedings of the 53rd Hawaii International Conference on System Sciences* (pp. 22–31).

Winkler, R. [Rainer], Hobert, S., Salovaara, A., Söllner, M. [Matthias], & Leimeister, J. M. (2020). Sara, the Lecturer: Improving Learning in Online Education with a Scaffolding-Based Conversational Agent. In R. Bernhaupt (Ed.), *ACM Conferences, Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (pp. 1–14). New York,NY,United States: Association for Computing Machinery. https://doi.org/10.1145/3313831.3376781

Wise, S. L. (1985). The Development and Validation of a Scale Measuring Attitudes toward Statistics. *Educational and Psychological Measurement*, *45*(2), 401–405. https://doi.org/10.1177/001316448504500226

Wolfe, C. R., & Cedillos, E. M. (2015). E-Communications Platforms and E-Learning. In J. D. Wright (Ed.), *International Encyclopedia of the Social & Behavioral Sciences* (pp. 895–902). Elsevier. https://doi.org/10.1016/b978-0-08-097086-8.41022-6

Wong, J. [Jacqueline], Baars, M., Davis, D., van der Zee, T., Houben, G.-J., & Paas, F. (2019). Supporting Self-Regulated Learning in Online Learning Environments and MOOCs: A Systematic Review. *International Journal of Human–Computer Interaction*, *35*(4-5), 356–373. https://doi.org/10.1080/10447318.2018.1543084

World Health Organization (2020). Coronavirus disease 2019 (COVID-19): Situation Report – 72. Retrieved from https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200401-sitrep-72-covid-19.pdf?sfvrsn=3dd8971b_2

Wu, H., Guo, Y., Yang, Y., Le Zhao, & Guo, C. (2021). A Meta-analysis of the Longitudinal Relationship Between Academic Self-Concept and Academic Achievement. *Educational Psychology Review*, *33*(4), 1749–1778. https://doi.org/10.1007/s10648-021-09600-1

Wu, J.-Y. (2021). Learning analytics on structured and unstructured heterogeneous data sources: Perspectives from procrastination, help-seeking, and machine-learning defined cognitive engagement. *Computers & Education*, *163*, 104066. https://doi.org/10.1016/j.compedu.2020.104066

Xu, C., Peters, M., & Brown, S. (2020). Instructor and Institutional Effects on Students' Statistics Attitudes. *Statistics Education Research Journal*, *19*(2), 7–26. https://doi.org/10.52041/serj.v19i2.107

Xu, J. [Joy], Lio, A., Dhaliwal, H., Andrei, S., Balakrishnan, S., Nagani, U., & Samadder, S. (2021). Psychological interventions of virtual gamification within academic intrinsic motivation: A systematic review. *Journal of Affective Disorders*, *293*, 444–465. https://doi.org/10.1016/j.jad.2021.06.070

Ye, D., & Pennisi, S. (2022). Using trace data to enhance Students' self-regulation: A learning analytics perspective. *The Internet and Higher Education*, *54*, 100855. https://doi.org/10.1016/j.iheduc.2022.100855

Yeckehzaare, I., Mulligan, V., Ramstad, G., & Resnick, P. (2022). Semester-level Spacing but Not Procrastination Affected Student Exam Performance. In A. F. Wise (Ed.), *ACM Digital Library, Lak22: 12th International Learning Analytics and Knowledge Conference* (pp. 304–314). New York,NY,United States: Association for Computing Machinery. https://doi.org/10.1145/3506860.3506907

Yousef, A. M. F., Chatti, M. A., & Schroeder, U. (2014). The state of video-based learning: A review and future perspectives. *International Journal on Advances in Life Sciences*, *6*(3&4), 122–135. Retrieved from https://www.researchgate.net/profile/ahmed-mohamed-fahmy-yousef/publication/278702595_the_state_of_video-based_learning_a_review_and_future_perspectives/links/55841e5608aefa35fe33d131/the-state-of-video-based-learning-a-review-and-future-perspectives.pdf

Yusuf, Y., Suyitno, H., Sukestiyarno, Y. L., & Isnarto (2019). The Influence of Statistical Anxiety on Statistic Reasoning of Pre-service Mathematics Teachers. *Bolema: Boletim De Educação Matemática*, *33*(64), 694–706. https://doi.org/10.1590/1980-4415v33n64a12

Zanakis, S. H., & Valenzi, E. R. (1997). Student Anxiety and Attitudes in Business Statistics. *Journal of Education for Business*, *73*(1), 10–16. https://doi.org/10.1080/08832329709601608

Zeidner, M. (1991). Statistics and mathematics anxiety in social science students: Some interesting parallels. *British Journal of Educational Psychology*, *61*(3), 319–328. https://doi.org/10.1111/j.2044-8279.1991.tb00989.x

Zhang, Y. [Yingbin], Paquette, L., Baker, R. S., Bosch, N., Ocumpaugh, J., & Biswas, G. (2022). How are feelings of difficulty and familiarity linked to learning behaviors and gains in a complex science learning task? *European Journal of Psychology of Education*, 1–24. https://doi.org/10.1007/s10212-022-00616-x

Zhang, Y. [Yuhai], Shang, L., Wang, R., Zhao, Q., Li, C., Xu, Y., & Su, H. (2012). Attitudes toward statistics in medical postgraduates: Measuring, evaluating and monitoring. *BMC Medical Education*, *12*(1), 117. https://doi.org/10.1186/1472-6920-12-117

Zhou, M., & Winne, P. H. (2012). Modeling academic achievement by self-reported versus traced goal orientation. *Learning and Instruction*, *22*(6), 413–419. https://doi.org/10.1016/j.learninstruc.2012.03.004

Zieffler, A., Garfield, J., Alt, S., Dupuis, D., Holleque, K., & Chang, B. (2008). What Does Research Suggest About the Teaching and Learning of Introductory Statistics at the College Level? A Review of the Literature. *Journal of Statistics Education*, *16*(2). Retrieved from https://doi.org/10.1080/10691898.2008.11889566

Zimmerman, W. A., & Austin, S. R. (2018). Using Attitudes and Anxieties to Predict End-of-Course Outcomes in Online and Face-to-Face Introductory Statistics Courses. *Statistics Education Research Journal*, *17*(2), 68–81.

Zimprich, D. (2012). Attitudes Toward Statistics Among Swiss Psychology Students. *Swiss Journal of Psychology*, *71*(3), 149–155. https://doi.org/10.1024/1421-0185/a000082

13 Appendix

13.1 Declaration (in German)

Ich erkläre:

1. Die Gelegenheit zum vorliegenden Promotionsvorhaben ist mir nicht kommerziell vermittelt worden. Insbesondere habe ich keine Organisation eingeschaltet, die gegen Entgelt Betreuerinnen und Betreuer für die Anfertigung von Dissertationen sucht oder die mir obliegenden Pflichten hinsichtlich der Prüfungsleistungen für mich ganz oder teilweise erledigt.
2. Ich versichere, dass ich die eingereichte Dissertation „How Students´ Disciplinary Attitudes and Beliefs Affect Learning in Introductory Statistics Courses“ selbstständig und ohne unerlaubte Hilfsmittel verfasst habe; fremde Hilfe habe ich dazu weder unentgeltlich noch entgeltlich entgegengenommen und werde dies auch zukünftig so halten. Anderer als der von mir angegebenen Hilfsmittel und Schriften habe ich mich nicht bedient. Alle wörtlich oder sinngemäß den Schriften anderer Autoren entnommenen Stellen habe ich kenntlich gemacht.
3. Die eingereichte Dissertation habe ich nicht bereits in einem anderen Prüfungsverfahren vorgelegt.
4. Des Weiteren ist mir bekannt, dass Unwahrhaftigkeiten hinsichtlich der vorstehenden Erklärung die Zulassung zur Promotion ausschließen bzw. später zum Verfahrensabbruch oder zur Rücknahme des erlangten Titels berechtigen.

13.2 Curriculum Vitae

Florian Berens

born July 03, 1989 in Nordhorn

German citizenship

Mail: florian.berens@uni-tuebingen.de

13.2.1 *Professional Experience*

Since Mar. 2023	Postdoctoral researcher at Eberhard Karls University of Tübingen at the Hector Research Institute of Education Sciences and Psychology (Director: Prof. Dr. Ulrich Trautwein), Hub Affiliation: Tübingen Center for Randomized Controlled Field Trials (Director: Prof. Dr. Benjamin Nagengast)
Oct. 2016 - Feb. 2023	Research associate at the Georg-August-University of Göttingen at the Center of Methods in the Social Sciences (since Oct. 2022: Institute of Methods and Methodological Principles in the Social Sciences), Chair of Quantitative Methods and Statistics (Prof. Dr. Steffen Kühnel, since Oct. 2022: Prof. Dr. Tobias Stubbe) (parental leave from May - Sep. 2022)
Feb. 2020 - Mar. 2020	Visiting Research Fellow at the Department for Statistics, University of Illinois at Urbana-Champaign (7 weeks)
Apr. 2015 - Sep. 2016	Student Assistant at the Chair of Quantitative Methods and Statistics (Prof. Dr. Steffen Kühnel) and Tutor for Statistics for the Social Sciences I, II & III, University of Göttingen
Sep. 2011 - Sep. 2015	Student Assistant at the Chair of Mathematics and its Teaching (Prof. Dr. Stefan Halverscheid) at the Institute of Mathematics and Tutor e.g. for Mathematics Teaching and Multidimensional Analysis, University of Göttingen

13.2.2 Education

Since Apr. 2020	Ph.D. Program Applied Statistics and Empirical Methods at the Center for Statistics of the University of Göttingen as a postgraduate program within the doctoral program in Social Sciences
Since Apr. 2017	Doctoral studies in social sciences at the University of Göttingen; Title of the dissertation: „How Students´ Disciplinary Attitudes and Beliefs Affect Learning in Introductory Statistics Courses“ Supervisors: Prof. Dr. Steffen-M. Kühnel (Chair of quantitative Methods in the Social Sciences, University of Göttingen) Prof. Dr. Stefan Halverscheid (Chair of Mathematics and its Teaching, University of Göttingen) Prof. Dr. Thomas Kneib (Chair of Statistics, University of Göttingen)
Oct. 2012 - Nov. 2016	Master of Education for high school teaching (Gymnasium) with the subjects mathematics & politics/economics, University of Göttingen (final grade: 1.3)
Oct. 2012 - Oct. 2016	Master of Arts in Political Science, University of Göttingen (final grade: 1.3, with distinction)
Oct. 2009 - Aug. 2012	Bachelor of Mathematics and Political Science in the profile of teacher training, University of Göttingen (final grade: 1.8)

13.2.3 Awards

Oct. 2018	Tandem Fellowship "Excellence in University Teaching" of the Stifterverband (with Sebastian Hobert); in addition Funding of the project "Interactive Learning on Demand", https://www.stifterverband.org/ lehrfellowships/2018/hobert_berens