



Bewertungskompetenz für den Biologieunterricht – Vom Modell zur empirischen Überprüfung

Dissertation zur Erlangung des Doktorgrades
der Mathematisch-Naturwissenschaftlichen Fakultäten
der Georg-August-Universität Göttingen

vorgelegt von
Sabina Eggert
aus Bad Gandersheim

Göttingen 2008

D7

Referent: Prof. Dr. Susanne Bögeholz

Korreferent: Prof. Dr. Marcus Hasselhorn

Tag der mündlichen Prüfung: 27. 06. 2008

Danksagung

In den letzten drei Jahren meiner Promotionszeit haben mich sehr viele Menschen begleitet und unterstützt, ohne die diese Arbeit nie möglich gewesen wäre. Diesen Menschen möchte ich danken.

Zuerst möchte ich mich bei meiner Betreuerin, Frau Prof. Dr. Susanne Bögeholz bedanken. Sie stand mir in allen Phasen meiner Arbeit mit fachlichem und freundschaftlichem Rat zur Seite und hat mir bei meinem Forschungsvorhaben Freiheit und Vertrauen geschenkt. Außerdem möchte ich mich bei meinem Zweitbetreuer Herrn Prof. Dr. Marcus Hasselhorn für sein stets offenes Ohr und seine wertvollen Ideen und Ratschläge bedanken.

Ein besonderer Dank gilt Herrn Prof. Dr. Jürgen Rost für seine unschätzbare Unterstützung in allen methodischen Fragen meiner Arbeit. Ein herzlicher Dank geht auch an Herrn Prof. Dr. Rainer Watermann, der ebenfalls in allen methodischen Fragen immer ein offenes Ohr hatte.

Außerdem möchte ich mich bei den Mitgliedern meiner Prüfungskommission: Herrn Prof. Dr. Gradstein, Herrn Prof. Dr. Feußner, Herrn Prof. Dr. Leuschner sowie Herrn Prof. Dr. Schulz-Hardt sehr herzlich bedanken.

Ein großer Dank gilt ebenfalls allen Lehrerinnen und Lehrern sowie allen Schülerinnen und Schülern, die mich bei der Durchführung meiner Studien sowie durch das Ausfüllen von Fragebögen tatkräftig unterstützt haben. Außerdem möchte ich mich bei allen Beteiligten des Projekts Biologie im Kontext, in dessen Rahmen ich meine Dissertation anfertigen konnte, bedanken. Vielen Dank an alle!

Darüber hinaus möchte ich mich bei allen Kolleginnen der Abteilung der Didaktik der Biologie für ihre Unterstützung bedanken. Dabei geht ein besonderer Dank an Janin Schatz, mit der ich nur zusammen die Entwicklung des Fragebogens realisieren konnte und ein Dank an Simone, die mich die gesamte Promotionszeit über begleitet hat.

Ein derartiges Promotionsvorhaben kann natürlich nicht ohne Freunde und Familie gelingen. Aus diesem Grund möchte ich mich ganz besonders bei meinen Freunden Susanne und Rapha, Konrad und Christina, Inken und Wilfried, Ellen und Ulli, Jim und Jan bedanken. Danke für eure Unterstützung, für eure Geduld und dafür, dass Ihr da seid!

Am meisten möchte ich mich natürlich bei meiner Familie bedanken, die immer für mich da ist und ohne die ich das hier nicht geschafft hätte!

Vielen Dank an Olaf, an Livi, und besonders an Mama und Papa!

Danke!

Göttingen, Mai 2008

Zusammenfassung

Das deutsche Bildungssystem befindet sich seit dem unterdurchschnittlichen Abschneiden deutscher Schülerinnen und Schüler in den PISA Studien in einem tiefgreifenden Umbruch. Eine bildungspolitische Reaktion auf die PISA Studien war die Entwicklung und Verabschiedung outputorientierter Bildungsstandards für den mittleren Schulabschluss. Seit Verabschiedung der Bildungsstandards für den Biologieunterricht (KMK, 2004) stellt Bewertungskompetenz einen von vier fest vorgeschriebenen Kompetenzbereichen dar. Die Integration von Bewertungskompetenz in den Biologieunterricht ist aber vielfach noch schwierig. Schwierigkeiten liegen einerseits in einem unterschiedlichen Verständnis von Bewertungskompetenz begründet. Andererseits stellt der Bereich Bewertungskompetenz im Gegensatz zu den in den naturwissenschaftlichen Fächern etablierten Kompetenzbereichen Fachwissen und Erkenntnisgewinnung einen relativ neuartigen Kompetenzbereich dar. Daraus folgt, dass eine Integration von Bewertungskompetenz eher zögerlich erfolgt. Darüber hinaus gibt es nur wenige kompetenzdiagnostische Ansätze, die sich der Operationalisierung und Messung von Bewertungskompetenz widmen.

Ziel der vorliegenden Arbeit ist es, a) zentrale Strukturen von Bewertungskompetenz in einem Modell theoretisch herzuleiten, und damit einen Beitrag zur Präzisierung des Bewertungskompetenzbegriffs zu liefern, und b) ein quantitatives Testinstrument für zunächst eine Dimension von Bewertungskompetenz zu entwickeln. Im ersten, theoretischen Schritt wurden auf Basis nationaler und internationaler Forschungsarbeiten zentrale Teilkompetenzen von Bewertungskompetenz identifiziert. Eine Teilkompetenz fasst dabei die Fähigkeit von Schülerinnen und Schülern mehrere Handlungsalternativen in Entscheidungssituationen unter Anwendung einer Entscheidungsstrategie miteinander vergleichen zu können, um zu einer informierten Entscheidung zu gelangen. Für diese Teilkompetenz wurde ein Fragebogen entwickelt, der die Anwendung von Entscheidungsstrategien in derartigen Entscheidungssituationen analysiert. In einer ersten explorativen Studie konnten mehrere verschiedene Entscheidungsstrategien identifiziert werden, die sich v.a. in der Intensität der Informationsverarbeitung unterschieden. Auf Basis dieser Ergebnisse wurde das Testinstrument speziell für Entscheidungssituationen einer Nachhaltigen Entwicklung unserer Erde weiterentwickelt. Das Antwortverhalten in Hinblick auf das optimierte Testinstrument kann durch ein eindimensionales polytomes Raschmodell (Partial Credit Modell) modelliert werden. Das Testinstrument ist dabei hinreichend reliabel und valide. Des Weiteren kann es über Jahrgangsstufen hinweg eingesetzt werden.

Das im ersten Schritt entwickelte theoretische Modell liefert einen Orientierungsrahmen für eine Integration und systematische Förderung von Bewertungskompetenz im Biologieunterricht. Das entwickelte Testinstrument zeigt Möglichkeiten einer quantitativen Messung von Bewertungskompetenz auf.

Abstract

Germany's educational system has experienced a radical change over the last few years. One major reason for this change were high school students' mediocre results in the PISA studies. One reaction to these results was the development and implementation of National Educational Standards for most school subjects. For biology education, four central competence areas were defined: scientific knowledge, scientific inquiry, communication and decision making competence with respect to issues of applied science (KMK, 2004).

The implementation of decision making competence into the biology classroom still faces many difficulties. Difficulties mainly stem from different views on what really constitutes decision making competence as well as teachers' reluctance to integrate this relatively new competence area into their teaching practices. Concerning research on decision making competence, especially quantitative assessment procedures are rare.

Thus, two major foci were identified for this study: a) the development of a theoretical model that identifies central dimensions of decision making competence and b) the development of a quantitative test instrument that can adequately measure one central dimension. This dimension captures students' ability to compare different possible options in a decision making situation by applying some sort of decision making strategy. The developed test instrument analysed students' answers with respect to the use of different decision making strategies.

In a first exploratory study different types of decision making strategies could be found. These strategies differed above all in their degree of information processing. The test instrument was then optimized on the basis of these findings and focused on decision making situations about the sustainable development of our planet. Students' responses on the test items could be modelled with a polytomous Rasch model (Partial Credit model). The test instrument meets requirements with respect to reliability and validity. Besides, it can be applied across different age groups.

Putting it into a nutshell, the developed theoretical model can provide a guideline for teachers to integrate and to foster students' decision making competence in the biology classroom. The developed test instrument provides one approach for a quantitative assessment of students' decision making competence.

Inhaltsverzeichnis

Danksagung.....	i
Zusammenfassung / Abstract.....	iii
Inhaltsverzeichnis.....	v
Abkürzungsverzeichnis.....	viii
Abbildungsverzeichnis.....	ix
Tabellenverzeichnis.....	x
Verzeichnis Kästen.....	x
1 Einleitung.....	1
2 Kompetenzen.....	3
2.1 Kompetenzen in der kompetenzorientierten Bildungsforschung.....	3
2.2 Bewertungskompetenz - Ansätze aus fachdidaktischer und psychologischer Sicht.....	6
3 Modellierung von Kompetenzen.....	15
3.1 Operationalisierung und Messung von Kompetenzen.....	15
3.2 Modellierung mit quantifizierenden eindimensionalen IRT-Modellen.....	18
3.3 Modellierung mit IRT-Modellen jenseits der Eindimensionalität.....	21
3.4 Modellgeltungstests für IRT-Modelle.....	24
4 Ableitung der Forschungsstrategie.....	27
4.1 Forschungsschwerpunkt I: Theoretische Entwicklung eines Kompetenzstrukturmodells für Bewertung.....	27
4.2 Forschungsschwerpunkt II: Operationalisierung und Messung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“	28

5 Göttinger Modell der Bewertungskompetenz - Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung.....	35
5.1 Zusammenfassung.....	35
5.2 Einleitung.....	36
5.3 Theoretischer Hintergrund: Prozessmodelle zu Bewerten und Entscheiden.....	37
5.4 Bewerten und Entscheiden im Kontext Nachhaltiger Entwicklung.....	42
5.5 Bewertungskompetenz im Biologieunterricht.....	45
5.6 Elemente von Kompetenzmodellen.....	46
5.7 Das Göttinger Modell der Bewertungskompetenz – konkretisiert für Gestaltungsaufgaben Nachhaltiger Entwicklung.....	52
5.8 Zusammenfassung und Ausblick.....	61
6 Identifying students’ decision making strategies in everyday life contexts and contexts of sustainable development.....	63
6.1 Abstract.....	63
6.2 Introduction.....	64
6.3 Decision making competence and decision theory.....	65
6.4 Methods.....	67
6.5 Results and Discussion.....	71
6.6 General Discussion.....	79
6.7 Conclusion.....	81
6.8 Appendix A/B.....	83
6.9 Weiterführende Erörterungen.....	85

7 Students' Use of Decision Making Strategies with regard to Socio-Scientific Issues – An Application of the Partial Credit Model.....	87
7.1 Abstract.....	87
7.2 Introduction.....	88
7.3 Theoretical background – Decision making competence and socio-scientific issues.....	90
7.4 Methods.....	94
7.5 Results.....	102
7.6 Summary and Discussion.....	109
7.7 Conclusion.....	114
7.8 Appendix A-C.....	116
7.9 Ergänzende Ausführungen.....	120
8 Zusammenfassung und Diskussion.....	121
8.1 Forschungsschwerpunkt I.....	121
8.2 Forschungsschwerpunkt II.....	122
8.3 Kritische Betrachtung des theoretischen Modells auf Basis der empirischen Ergebnisse.....	128
9 Fazit und Ausblick.....	131
10 Literaturverzeichnis.....	137
11 Anhang.....	147
11.1 Scoring Guide zur Kodierung von Schülerantworten.....	148
11.2 Ausführliche Beschreibung zum Scoring von Schülerantworten	151
11.3 Lebenslauf.....	163

Abkürzungsverzeichnis

BIC	Bayes Information Criterion
bik	Biologie im Kontext
BNE	Bildung für Nachhaltige Entwicklung
CAIC	Consistent Akaike Information Criterion
CCC	Characteristic Category Curve
DESI	Deutsch-Englisch-Schülerleistungen International
ICC	Item Characteristic Curve
IQB	Institut für die Qualitätsentwicklung im Bildungswesen
IRT	Item Response Theory
KMK	Kultusministerkonferenz
LCA	Latent-Class-Analyse
OECD	Organisation for Economic Co-Operation and Development
PISA	Programme for International Student Assessment
PCM	Partial Credit Modell
MRM	Mixed Rasch Modell
NOS	Nature of Science
RM	Raschmodell
SEPUP	Science Education for Public Understanding Project
SSI	Socio-Scientific Issues
TIMSS	Third International Mathematics and Science Study
VERA	Vergleichsarbeiten in der Grundschule

Abbildungsverzeichnis

Abb. 2.1: Phasenmodell der Entscheidungsfindung (verändert nach Betsch & Haberstroh, 2005, S. 263).....	12
Abb. 3.1: Item-Characteristic-Curve: Zusammenhang zwischen der Personenfähigkeit (θ) und der Wahrscheinlichkeit das Item i zu lösen (verändert nach Rost, 2004, S. 120).....	19
Abb. 3.2: Category-Characteristic-Curves für ein polytomes Item im Partial Credit Modell (verändert nach Rost, 2004, S. 204).....	20
Abb. 3.3: Unterschiede in der Modellierung von Mehrdimensionalität (verändert und übersetzt nach Wu et al., 1998, S. 115).	22
Abb. 5.1: Göttinger Modell der Bewertungskompetenz.....	53
Abb. 5.2: Beispielaufgabe zur Überprüfung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“	60
Fig. 6.1: Mean item response probabilities for all items across all three decision making tasks.....	74
Fig. 7.1: Assumed increase in decision making competence with respect to the use of decision making strategies and the ability to reflect on such processes for decision making tasks with equally legitimate options.....	92
Fig. 7.2: Person-Item map (adapted from Wilson, 2005, p. 96).....	100
Fig. 7.3: Person-Item map: Partial Credit Model.....	104
Fig. 7.4: Influence of years of education on individuals' performance on the decision making items.....	108

Tabellenverzeichnis

Tab. 3.1: Überblick der für die Studie relevanten IRT-Modelle.....	23
Tab. 5.1: Prozessmodell der Entscheidungsfindung und Konkretisierung für Gestaltungsaufgaben Nachhaltiger Entwicklung.....	40
Tab. 5.2: Scientific Literacy am Beispiel von zwei Prozessen und deren Niveaustufen (vereinfacht nach: Deutsches PISA-Konsortium, 2001, S. 204).....	47
Tab. 5.3: Scoring Guide für die SEPUP Variable “Using Evidence to Make Trade-offs” (übersetzt, leicht modifiziert nach Wilson & Sloane, 2000, S. 193).....	51
Tab. 5.4: Kompetenzniveaus der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“.....	57
Tab. 5.5: Kontextmatrix für eine systematische Aufgabenentwicklung zur Umsetzung von Bildung für eine Nachhaltige Entwicklung (verändert nach Hammann, 2006a, 86).....	61
Tab. 6.1: Final Scoring Guide used to analyze individuals’ responses to the three decision making tasks.....	69
Tab. 6.2: Fit statistics for all applied models.....	72
Tab. 6.3: Mean probabilities of class membership for all three classes in the hybrid model.....	73
Tab. 6.4: Q-indices for both Rasch classes.....	73
Tab. 6.5: Example responses for class 1: “mainly non compensatory”.....	76
Tab. 6.6: Example responses for class 2: „mainly weighing“.....	77
Tab. 6.7: Example responses for class 3: „compensatory“.....	79
Tab. 7.1: Scoring Guide for booklet 1.....	98
Tab. 7.2: Scoring Guide for booklet 2.....	99

Verzeichnis Kästen

Kasten 5.1: Beispiele von ausgewählten Entscheidungsstrategien (in Anlehnung an Jungermann et al., 1998, S. 123).....	41
Kasten 5.2: Das Leitbild der Nachhaltigen Entwicklung – Herausforderung für den Biologieunterricht.....	44
Kasten 5.3: Science Education for Public Understanding Program (SEPUP).....	51
Kasten 5.4: SEPUP Beispielaufgabe „Verpackungsmaterial“.....	52

1 Einleitung

Was verstehen Sie unter dem Begriff Bewertungskompetenz? Oftmals ist die Antwort in Gesprächen, nicht nur mit Biologielehrerinnen und -lehrern, darauf: „Bewertungskompetenz ist die Fähigkeit, (Schüler-)leistungen objektiv, und damit gerecht, bewerten und benoten zu können.“

In der vorliegenden Dissertation „Bewertungskompetenz für den Biologieunterricht - Vom Modell zur empirischen Überprüfung“, die im Rahmen des nationalen Projekts „Biologie im Kontext“ entstanden ist, wird unter Bewertungskompetenz jedoch ein anderes Konzept als das der Leistungsbeurteilung verstanden. „Biologie im Kontext“ folgt dabei den im Jahr 2004 verabschiedeten Bildungsstandards für den mittleren Schulabschluss in den naturwissenschaftlichen Fächern, wo Bewertungskompetenz als die Fähigkeit von Schülerinnen und Schülern (und somit nicht von Lehrkräften!) verstanden wird, Themen angewandter Naturwissenschaften verstehen, analysieren und bewerten zu können, um anschließend am gesellschaftlich geführten Diskurs kompetent teilhaben zu können (KMK, 2004, S. 14). Charakteristische Themen angewandter Naturwissenschaften sind Themen des Umweltschutzes und einer zukunftsfähigen, nachhaltigen Entwicklung unserer Erde, aber auch Themen im Bereich der Bio- oder Medizinethik. Allen diesen Themen ist gemeinsam, dass sie nicht allein auf Basis naturwissenschaftlichen Wissens gelöst werden können, sondern vielmehr durch Integration vielfältiger gesellschaftlicher Perspektiven bearbeitet werden müssen.

„Biologie im Kontext“ greift die Forderungen der Bildungsstandards auf und hat als einen Schwerpunkt die Förderung von Schülerkompetenzen an innovativen, schüler- und gesellschaftlich relevanten Kontexten. Das Projekt ist auch Ergebnis der vielfachen Reaktionen, die nach dem mäßigen Abschneiden deutscher Schülerinnen und Schüler in den internationalen PISA Vergleichstudien einsetzten. Die Verabschiedung der Bildungsstandards stellte 2004 eine Reaktion auf die nach PISA geforderten Veränderungen schulischer Curricula von einer Inputorientierung hin zu einer Outputorientierung dar (Klieme et al., 2003). Die Bildungsstandards sind dabei als Regelstandards formuliert und beschreiben, was Schülerinnen und Schüler aller Schulformen am Ende der Klasse 10 können sollen. Aus bildungspolitischer Sicht stellt sich u.a. die Frage, wie die Kompetenzen

der Bildungsstandards überprüfbar und somit messbar gemacht werden können. Zur Klärung dieser Frage wurde das Institut für die Qualitätssicherung im Bildungswesen (IQB) gegründet. Für die naturwissenschaftlichen Fächer liegt der Fokus des IQB bei dem für 2012 anvisierten Assessment auf der Überprüfung der Kompetenzbereiche Fachwissen und Erkenntnisgewinnung. Damit stehen die für den Unterricht in den naturwissenschaftlichen Fächern zentralen Themen mit langer Tradition im Zentrum. Die für die naturwissenschaftlichen Fächer eher neuen Kompetenzbereiche Bewertung und Kommunikation werden zunächst einmal nicht berücksichtigt. Gerade aus diesem Grund ist es notwendig, Bewertungskompetenz, welche für die Bearbeitung von Themen angewandter Naturwissenschaften und damit gesellschaftlich relevanter Fragestellungen wichtig ist, mehr Aufmerksamkeit zu widmen. Ein Forschungsschwerpunkt der vorliegenden Dissertation, und somit auch des Göttinger Teilprojekts im Rahmen von „Biologie im Kontext“, ist die theoretische Präzisierung von Bewertungskompetenz im Rahmen eines Kompetenzstrukturmodells. Die für den Kompetenzbereich Bewertung formulierten Teilstandards der Bildungsstandards stellen einen sehr breiten, an der Schulpraxis orientierten Ansatz dar.

Für eine empirische Überprüfung von Bewertungskompetenz ist zunächst eine Präzisierung übergreifender, theoretisch begründbarer und voneinander zu trennender Strukturen erforderlich. Auf Basis des entwickelten theoretischen Modells wird im zweiten Teil der Arbeit eine der identifizierten Teilkompetenzen von Bewertungskompetenz empirisch überprüft. Dazu wurde ein Testinstrument entwickelt, welches die Anwendung von Bewertungsstrategien bei Themen des Umweltschutzes bzw. einer nachhaltigen Entwicklung unserer Erde von Schülerinnen und Schülern über Jahrgangsstufen hinweg messen kann. Die Ergebnisse der Dissertation lieferten zudem Ansätze für die Förderung von Bewertungskompetenz im Biologieunterricht in Form von Unterrichtskonzepten, die gemeinsam mit Biologielehrkräften in „Biologie im Kontext“ entwickelt wurden.

2 Kompetenzen

In diesem Kapitel wird zunächst auf die Genese der Begriffe Kompetenz und Bewertungskompetenz eingegangen. Im dritten Kapitel werden notwendige kompetenzdiagnostische Voraussetzungen und zum Einsatz kommende Messmodelle für eine Beschreibung von Kompetenzstrukturen und deren anschließender empirischer Überprüfung, beschrieben. Im vierten Kapitel werden die in der Dissertation vorgenommene Forschungsstrategie sowie die daraus abgeleiteten Forschungsfragen präsentiert.

2.1 Kompetenzen in der kompetenzorientierten Bildungsforschung

Der Kompetenzbegriff hat in Deutschland seit Verabschiedung der Bildungsstandards für den mittleren Schulabschluss neue Beliebtheit erlangt und ist geradezu zu einem Modewort nicht nur in der schulpolitischen Diskussion avanciert. Auslöser der Formulierung von Bildungsstandards waren die Ergebnisse der TIMSS (vgl. u.a. Baumert, Bos & Lehmann, 2000) und der PISA Studien (vgl. u.a. Deutsches PISA Konsortium, 2001), die das unterdurchschnittliche Abschneiden deutscher Schülerinnen und Schüler im internationalen Vergleich dokumentieren und damit die Qualität des deutschen Schulsystems in Frage stellen. Der Grundstein für die Bildungsstandards wurde durch die Expertise zur Entwicklung nationaler Bildungsstandards gelegt (Klieme et al., 2003). Eine zentrale Forderung dieser Expertise betrifft die Neuorientierung des Bildungssystems von einer Input- zu einer Outputorientierung (ebd.). Unter Output eines Bildungssystems wird dabei

„neben der Vergabe von Zertifikaten im Wesentlichen der Aufbau von Kompetenzen, Qualifikationen, Wissensstrukturen, Einstellungen, Überzeugungen und Werthaltungen verstanden“ (Klieme et al., 2003, S. 12).

Trotz seiner ungebremsten Popularität ist der Kompetenzbegriff ein schwieriger. Dies liegt v.a. an seiner Mehrdeutigkeit und den z.T. unterschiedlichen Funktionen, die ihm in Politik,

Schule, Wissenschaft und im alltäglichen Leben zukommen. Weinert selbst, der den Kompetenzbegriff maßgeblich geprägt hat und dessen Definition der Expertise zugrunde gelegt wurde, beleuchtet den Gebrauch des Begriffs und die Entwicklung seiner Beliebtheit kritisch, indem er von einer „Inflation von Kompetenzbegriffen“ in den letzten Jahrzehnten spricht, die es nahezu unmöglich macht, einen Kern des Kompetenzbegriffs theoretisch zu formulieren (Weinert, 2001). Die Probleme in der Definition des Begriffs Kompetenz spiegeln sich auch in der Weinertschen Definition wider:

„Kompetenzen sind die bei Individuen verfügbaren oder durch sie erlernbaren kognitiven Fähigkeiten und Fertigkeiten, um bestimmte Probleme zu lösen, sowie die damit verbundenen motivationalen, volitionalen und sozialen Bereitschaften und Fähigkeiten, die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können“ (Weinert, 2001, S. 27-28).

Diese Definition stellt einen maximal integrierenden Kompetenzbegriff dar und versucht, mit den Worten von Rost, nicht zu trennen, sondern zu verbinden:

„Der Kompetenzbegriff versucht nicht zu zerlegen, was zusammengehört. Eine, oder besser jede Kompetenz umfasst Wissen, Verstehen, Fähigkeiten, Können, Erfahrung, Handeln und Motivation“ (Rost, 2006, S. 5).

Beide Definitionen von Kompetenz sind plausibel. Sollte man ein Individuum beschreiben, welches in einem bestimmten Bereich kompetent ist, dann würde man eine vergleichbare Beschreibung erwarten. Eine derart umfassende Beschreibung stellt jedoch Herausforderungen v.a. an eine theoretische, fachdidaktische Beschreibung von Kompetenzen in Form von Kompetenzmodellen, auf dessen Basis empirische Überprüfungen geplant und durchgeführt werden sollen. Weinert selbst und andere Autoren (vgl. u.a. Hartig & Klieme, 2006) thematisieren drei zentrale Probleme bei der Beschreibung von Kompetenzen: Inwiefern sind Kompetenzen erlernbar? Wie konkret lassen sich Situationen beschreiben bzw. eingrenzen, in denen Kompetenzen erlernt bzw. angewandt werden sollen? In welchem Maße werden motivationale und volitionale Aspekte in den Kompetenzbegriff integriert?

Die erste Frage nach dem Grad der Lernbarkeit spannt ein Kontinuum von Fähigkeiten auf, an deren einem Ende die bei Individuen verfügbaren Fähigkeiten, solchen Fähigkeiten am anderen Ende gegenüberstehen, die von ihm erlernt werden müssen. Versteht man Kompetenzen eher als allgemeine, bereits vorhandene intellektuelle Fähigkeiten, die ein Individuum bei der Bewältigung von Situationen benötigt, dann rückt der Begriff der Kompetenz näher an das, was in der Psychologie als ein allgemeines psychologisches Konstrukt, in diesem Falle das Intelligenzkonstrukt, beschrieben wird (vgl. Weinert, 1999;

Hartig & Klieme, 2006). Das andere Ende des Kontinuums dagegen beschreibt eher bereichsspezifische Kompetenzen, die von Individuen erst erlernt werden müssen, um bestimmte Situationen bewältigen zu können. Mit der Beschreibung von Situationen wird die zweite Frage nach der Domänenspezifität von Kompetenzen gestellt. Ähnlich wie der Kompetenzbegriff ist auch der Begriff der Domäne schwierig zu fassen. Auch hier lässt sich wieder ein Kontinuum von einem relativ breiten, fächerübergreifenden hin zu einem sehr konkreten Verständnis einer Domäne beschreiben. Ein fächerübergreifendes Verständnis liegt den PISA Untersuchungen im Bereich Problemlösen zugrunde:

“An individual’s capacity to use cognitive processes to confront and resolve real, cross-disciplinary situations where the solution path is not immediately obvious and *where the content areas or curricular areas that might be applicable are not within a single subject of mathematics, science or reading*” (OECD, 2004, S. 26; kursiv verändert durch S.E.).

Der Bereich Problemlösen fasst somit als Domäne nicht ein Fach, sondern in einen realweltlichen Kontext eingebundene (Problem)-situationen, die sich nicht durch Routinehandlungen lösen lassen und die eine Integration von Wissen aus mehreren Fächern erfordern (OECD, 2004). So wird an der Beschreibung von *cross-curricular competencies* deutlich, dass die Situationen, die von Schülerinnen und Schülern bewältigt werden müssen, zur Domänendefinition herangezogen werden sollten.

Die Klieme-Expertise legt ihrem Domänenbegriff Erkenntnisse der Expertiseforschung zugrunde, wo unter Domäne ein Lernbereich bzw. ein Fach verstanden wird. Andere Kompetenzmodelle, wie z.B. Kompetenzen beim Experimentieren (Hammann, 2004) oder Kompetenzen im Umgang mit Diagrammen (Lachmayer et al., 2007), grenzen den Bereich einer Domäne stärker ein. Je enger dabei die Kompetenzdomäne gefasst wird, desto bereichsspezifischer müssen auch die erforderlichen Kompetenzen definiert werden. Generell lässt sich für den Begriff der Domäne zusammenfassen, dass die genannten Ansätze sich nicht ausschließen sondern vielmehr ergänzen, da fachbezogene Kompetenzen eine Voraussetzung für fächerübergreifende Kompetenzen darstellen (Klieme et al., 2003).

Neben der Lernbarkeit und Domänenspezifität wird das dritte Problem, die Integration von motivationalen und volitionalen Aspekten in den Kompetenzbegriff nicht minder kontrovers diskutiert. Auch wenn Einigkeit darüber besteht, dass für die Bewältigung von Situationen neben kognitiven Leistungsdispositionen immer auch motivationale und volitionale Faktoren eine Rolle spielen, stellt diese Forderung eine große Herausforderung für die Operationalisierung und Messung von Kompetenzen dar. So konstatiert auch Weinert in

seiner Beschreibung von Kompetenzen im Rahmen des OECD Projekts „Defining and Selecting Key Competencies“, dass

„der Vorteil einer derartig breiten Definition des Kompetenzbegriffs zugleich auch seine größte Schwäche darstellt [...], wozu es weder eine theoretische noch praktische Lösung gibt“ (Weinert, 1999, S. 26).

Ähnlich beschreibt auch Klieme, dass die facettenreichen Anforderungen, die die Weinertsche Definition an die Bildungsstandards stellt, von diesen nicht vollständig erfüllt werden können (Klieme, 2003, S. 12). Letztendlich empfiehlt Weinert selbst, aus theoretischen aber wohl auch aus pragmatischen Gesichtspunkten, den Kompetenzbegriff auf domänenspezifische, *kognitive* Leistungsdispositionen einzugrenzen und damit in gewissem Maße Komponenten zu trennen, die eigentlich nicht getrennt werden sollen. Diesem Ansatz folgten dann auch die internationalen Schulleistungsstudien TIMSS und PISA sowie das DFG-Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“, welches Kompetenzen als kontextspezifische kognitive Leistungsdispositionen beschreibt (Klieme & Leutner, 2006). Diese Beschränkung auf *kognitive* Fähigkeiten stellt eine sinnvolle Vorgehensweise für die Entwicklung von Messinstrumenten für spezifische Kompetenzen dar. Ungelöst bleiben jedoch die Schwierigkeiten des Kompetenzbegriffs. So wird die vorliegende Arbeit ebenfalls diesem pragmatischen, auf kognitive Kompetenzen fokussierenden, Ansatz folgen.

2.2 Bewertungskompetenz - Ansätze aus fachdidaktischer und psychologischer Sicht

Bewertungskompetenz stellt seit der Verabschiedung der Bildungsstandards nicht nur im Fach Biologie, sondern in allen naturwissenschaftlichen Fächern einen von vier verbindlich vorgeschriebenen Kompetenzbereichen dar (KMK, 2004). Ähnlich wie der Kompetenzbegriff ist auch der Bewertungskompetenzbegriff schwierig zu fassen und führt oftmals zu Missverständnissen in fachdidaktischen, schulpraktischen oder kompetenzdiagnostischen Diskussionen. Dies wird exemplarisch an den Bildungsstandards der naturwissenschaftlichen Fächer deutlich. In allen einleitenden Kapiteln zu den Fächern Biologie, Chemie und Physik wird Bewertungskompetenz als eine Fähigkeit verstanden, verschiedenste Sachverhalte in unterschiedlichen Kontexten erkennen und bewerten zu können und bei einer derartigen Bewertung neben fachlichen auch vielfältige

gesellschaftliche, darunter beispielsweise ökonomische und soziale, Aspekte in eine Bewertung integrieren zu können. An den konkreten Aufgabenbeispielen, die der Erläuterung der Bildungsstandards dienen, wird jedoch deutlich, dass die Fächer im Hinblick auf Bewertungskompetenz unterschiedliche Schwerpunkte setzen. So wird im Fach Chemie, v.a. aber im Fach Physik in den praktischen Aufgabenbeispielen unter Bewertungskompetenz die Fähigkeit verstanden, aufgrund von Experimenten, Untersuchungen oder wissenschaftlichen Ergebnissen Schlussfolgerungen ziehen und die Aussagekraft von derartigen Ergebnissen eher *beurteilen* als bewerten zu können. Bewertungskompetenz im Fach Biologie wird im Vergleich dazu vielfältiger gefasst und versteht eine fachliche Beurteilung des biologischen Sachverhalts als Ausgangspunkt für eine anschließende Bewertung:

„Bevor Schülerinnen und Schüler die eigentliche Bewertung vornehmen, klären sie die biologischen Sachverhalte und erfassen die mögliche Problematik. Dabei nehmen sie [verschiedenste Perspektiven] ein und setzen beim systematischen Bewerten von Handlungsmöglichkeiten diese mit Werten in Beziehung“ (KMK, 2004, S. 15).¹

Bewertungskompetenz im Biologieunterricht erfordert somit die Fähigkeit, Probleme multiperspektivisch betrachten zu können und dabei eine integrierende naturwissenschaftliche und ethische Bewertung von Problemen durchführen zu können (KMK, 2004). Diese Beschreibung der Bildungsstandards stellt einen wertvollen Ausgangspunkt für die Bearbeitung und Untersuchung von Bewertungskompetenz im Biologieunterricht dar. Im Hinblick auf eine Identifikation zentraler Strukturen sind die Bildungsstandards jedoch relativ unsystematisch und die für Bewertung ausformulierten Teilstandards eher additiv formuliert, als dass sie eine Basis für eine Messung von Bewertungskompetenz darstellen könnten. Aus diesem Grund werden im Folgenden die für eine Präzisierung von Bewertungskompetenz relevanten fachdidaktischen und psychologischen Ansätze kurz vorgestellt.

Wie im ersten Kapitel beschrieben können Kompetenzen nach Weinert (2001) sowie Hartig & Klieme (2006) als kontextspezifische kognitive Leistungsdispositionen definiert werden. Die kontextspezifische Beschreibung wird in der biologiedidaktischen Diskussion auch für den Kompetenzbereich Bewertung angenommen. So lassen sich in der nationalen

¹ Eine ähnliche Unterscheidung wird auch im Kerncurriculum Biologie für die gymnasiale Oberstufe vorgenommen, wo evidenzbasiertes Bewerten von Hypothesen z.B. beim Experimentieren einem Bewerten im Sinne von moralischem Urteilen bzw. ethischem Begründen gegenübergestellt wird (vgl. Mayer et al., 2004).

Forschungstradition zu Bewertungskompetenz v.a. zwei Ansätze ausmachen, deren Unterscheidung v.a. auch auf unterschiedliche Kontexte zurückzuführen ist. Dies ist zum einen Forschung zur moralischen Urteilsfähigkeit im Kontext von bio- oder medizinethischen Themen, zum anderen Forschung zu Bewertungs- und Gestaltungskompetenz im Rahmen von Bildung für eine Nachhaltige Entwicklung². Im Gegensatz zur nationalen Forschungstradition wird diese Kontextualisierung im angloamerikanischen Raum im Bereich von *socio-scientific decision-making* jedoch nicht vorgenommen. Hier werden unter *socio-scientific issues* alle gesellschaftlich relevanten Themen verstanden, die multiperspektivische Lösungsansätze erfordern.

Moralische Urteilsfähigkeit

In bio- bzw. medizinethischen Kontexten werden im Rahmen der Förderung moralischer Urteilsfähigkeit v.a. Dilemmasituationen bearbeitet. Beispiele für solche Dilemmasituationen sind Fragen nach der Legitimation von Schwangerschaftsabbruch, Fragen der Präimplantationsdiagnostik, oder aber Fragen nach der Legitimation von Klonierungsversuchen (vgl. u.a. Höhle, 2006). Viele der fachdidaktischen Ansätze zu diesem Kontextbereich schlagen ähnliche Vorgehensweisen in der Bearbeitung derartiger Themen im Biologieunterricht vor (Bayrhuber, 1992; Dulitz & Kattmann, 1999; Höhle, 2001; Meisert & Kierdorf, 2001). In allen diesen fachdidaktischen Unterrichtsmodellen³ steht am Anfang der Bearbeitung von Dilemmasituationen die Wahrnehmung und Definition einer ethisch relevanten Situation. Daran schließt sich in den meisten Fällen die Generierung von Handlungsoptionen an, wobei in bioethischen Dilemmasituationen zumeist nur zwei Handlungsoptionen möglich sind, die in einer kategorialen „Ja / Nein-Entscheidung“ münden können (vgl. u.a. Barkmann & Bögeholz, 2003). Anschließend werden die mit den Handlungsoptionen in Verbindung stehenden Argumente sowie die dahinter stehenden Werte herausgearbeitet. Es folgt eine Abwägung der erarbeiteten Argumente sowie eine eingehende Betrachtung der berührten Werte, bevor ein individuelles Urteil gefällt wird. In vielen Modellen werden die angesprochenen Werte der deontologischen bzw. der konsequenzialistischen Argumentationsweise zugeordnet. Ein komplexeres Vorgehen in der Analyse der berührten Werte schlagen Kattmann & Dulitz vor, indem sie drei Ebenen in Bezug auf eine Wertereflexion vorschlagen, die einer zu vereinfachten Darstellung ethischer Problemstellungen vorbeugen soll (vgl. Kattmann & Dulitz, 1990). An die individuelle

2 Eine Übersicht zentraler Ansätze zu beiden Kontextbereichen liefern Bögeholz et al., 2004.

3 Der Begriff Modell wird in diesem Zusammenhang eher in seiner Alltagssprachlichen Bedeutung gebraucht und ist daher nicht mit dem Modellbegriff in der Kompetenzdiagnostik oder dem Modellbegriff in der klassischen und probabilistischen Testtheorie zu verwechseln.

Urteilsbildung schließt sich eine Reflexion anderer Urteile sowie eine Reflexion möglicher Konsequenzen an. Die Abfolge der einzelnen Schritte in der Urteilsbildung kann dabei in den verschiedenen Modellen abweichen (Bögeholz et al., 2004). Allen Modellen ist gemein, dass sie für bzw. aus der Unterrichtspraxis heraus entwickelt wurden und somit noch keine Kompetenzstrukturmodelle, wie sie im Hinblick auf eine Operationalisierung und Messung von Kompetenzen notwendig sind, darstellen. Eine Grundlage für eine Messung stellt das auf den Arbeiten von Hößle (2001) entwickelte Kompetenzstrukturmodell für bioethische Fragestellungen dar (Reitschert et al., 2007).

Zusammenfassend lässt sich für eine Bearbeitung bioethischer Dilemmasituationen konstatieren, dass sie sich oftmals auf eine Gegenüberstellung von zwei Handlungsoptionen zuspitzen lassen. Die für die jeweilige Handlungsoption angeführten Argumente basieren dabei auf unterschiedlichen Werthaltungen, die oftmals zwei ethischen Denktraditionen (vgl. u.a. Bayrhuber 1992; Hößle, 2001) bzw. drei Reflexionsebenen (Kattmann & Dulitz, 1990) zugeordnet werden.

Bewertungskompetenz im Rahmen von Bildung für eine Nachhaltige Entwicklung (BNE)

Im Gegensatz zu bioethischen Dilemmasituationen spitzen sich Bewertungssituationen im Kontext Nachhaltiger Entwicklung in der Regel nicht auf zwei einander unvereinbar gegenüberstehende Handlungsoptionen⁴ zu. Vielmehr können in diesen Situationen mehrere Handlungsoptionen entwickelt werden, wovon auch mehrere unter dem Leitbild der Nachhaltigen Entwicklung als Lösung in Frage kommen können. Aus diesem Grund spricht man im Kontext Nachhaltiger Entwicklung auch von Gestaltungssituationen bzw. -optionen im Sinne von Gestaltungskompetenz:

„Unter Gestaltungskompetenz wird das nach vorne weisende Vermögen bezeichnet, die Zukunft von Sozietäten, in denen man lebt, in aktiver Teilhabe im Sinne nachhaltiger Entwicklung modifizieren und modellieren zu können. [...] Dabei werden den Bürgern erhebliche Fähigkeiten⁵ (z.B. vorausschauendes Planen, eigenständige Informationsaneignung und -bewertung sowie neue Anforderungen in Bezug auf Kommunikation und Kooperation) bei der Beteiligung an Verständigungs- und Entscheidungsprozessen abverlangt“ (de Haan & Harenberg, 1999, S. 62).

4 Ausnahmen stellen Ökologisch-Soziale Dilemmata dar, in denen bei der Nutzung einer gemeinsamen Ressource die Forderung einer nachhaltigen Nutzung der Forderung nach Grundbedürfnisorientierung gegenüberstehen (vgl. Ernst, 1997).

5 In einer aktuelleren Definition von Gestaltungskompetenz definiert de Haan diese Fähigkeiten als spezielle Handlungs- und Problemlösefähigkeiten (vgl. de Haan 2006, S. 22).

Das de Haansche Konzept der Gestaltungskompetenz stellt einen Maximalanspruch für Bildung und Schule dar und sollte somit als eine Vision und Leitbild für BNE verstanden werden. Bewertungskompetenz ist dabei eine wichtige Voraussetzung für Gestaltungskompetenz (vgl. u.a. de Haan & Harenberg, 1999; Bögeholz & Barkmann, 2005). Sie beschreibt die Fähigkeit, in komplexen Problemsituationen begründet und systematisch bei unterschiedlichen Handlungsoptionen zu entscheiden, um kompetent am gesellschaftlichen Diskurs - beispielsweise um die Gestaltung von Nachhaltigen Entwicklungen - teilhaben zu können (Bögeholz et al., 2004; Bayrhuber et al., 2007).

Im Hinblick auf fachdidaktisch fokussierte Ansätze im deutschsprachigen Raum sind v.a. zwei Modelle für Bewertungskompetenz zu nennen. Dies sind die Modelle der ökologischen Bewertungs- und Urteilskompetenz (Bögeholz & Barkmann, 2003) bzw. der sozio-ökologischen Umweltbildung (Kyburz-Graber, 1997). Beide Modelle sind dahingehend vergleichbar, als dass sie das Lösen von Gestaltungssituationen durch die Entwicklung von Gestaltungsoptionen zum Ziel haben, wobei sowohl auf Sachwissen als auch auf relevante Werte und Normen Bezug genommen werden soll. Unterschiede bestehen in den Schwerpunkten der beiden Modelle. Das Modell der sozio-ökologischen Umweltbildung analysiert v.a. Handlungssysteme und stellt dabei Interessen und Werthaltungen beteiligter Personengruppen in den Vordergrund. Das Modell der ökologischen Bewertungs- und Urteilskompetenz ermöglicht einen Vergleich mehrerer Gestaltungsoptionen durch die Entwicklung eines geeigneten Bewertungsverfahrens, d.h. einer Bewertungsmethode, die den Bewertungs- und Entscheidungsprozess systematisieren und für Dritte transparent machen soll (Bögeholz et al., 2004; Bögeholz, 2006). Für den Einsatz im Unterricht wurde dazu eine Unterrichtsmethode entwickelt, die Schülerinnen und Schüler im Umgang mit Gestaltungssituationen unterstützt und wichtige Teilschritte in Bewertungsverfahren offen legt („Methode der expliziten Bewertung“; vgl. Ahlf-Christiani, 2003; Bögeholz, 2006).

Decision-making about Socio-Scientific Issues

Im angloamerikanischen Raum stellt das Konzept der *decision-making-competence* das Pendant zu Bewertungskompetenz im deutschsprachigen Raum dar. Unter *socio-scientific issues* (SSI) werden alle gesellschaftlich relevanten Kontexte verstanden, die zwar eine naturwissenschaftliche Basis haben, aber nicht ausschließlich mit Hilfe naturwissenschaftlicher Konzepte gelöst werden können (vgl. Ratcliffe & Grace, 2003; Sadler, 2007). Aufgrund ihrer komplexen, multiperspektivischen Natur stellen SSI realweltliche, oft an der Forschungsfront liegende („*frontier science*“) Probleme dar, die von vielen Wissenschaftler(inne)n als besser geeignet angesehen werden Schülerinnen und Schüler zu unterstützen die Probleme unserer Umwelt lösen zu können, als rein

naturwissenschaftliche Fragestellungen (vgl. u.a. Sadler, 2007; Seethaler & Linn, 2004, Science Education for Public Understanding Project (SEPUP) 1995; Zohar & Nemet, 2002). Als wichtigste Forschungsstränge im Hinblick auf SSI lassen sich die folgenden ausmachen: Der Einfluss von *socio-scientific issues*...

1. auf das Erarbeiten und Entwickeln naturwissenschaftlicher Konzepte,
2. auf das Verständnis von Schülerinnen und Schülern über die Generierung naturwissenschaftlicher Erkenntnisse bzw. den Erkenntnisprozess in den Naturwissenschaften⁶ sowie
3. auf die Bewertungskompetenz und die Argumentationsfähigkeit von Schülerinnen und Schülern.

Für die vorliegende Studie ist v.a. der dritte Forschungsstrang von Bedeutung, da er die Kernaspekte des Kompetenzbereichs Bewertung behandelt. Er beschäftigt sich mit den Bewertungs- und Entscheidungs- bzw. Argumentationsprozessen, die bei der Bearbeitung von SSI durchlaufen werden. *Decision-making-competence* wird dabei definiert als die Fähigkeit, informierte Entscheidungen auf der Basis (natur-) wissenschaftlicher Evidenz und dem anschließenden Abwägen der Vor- und Nachteile gegebener Optionen zu treffen (vgl. u.a. SEPUP, 1995; Roberts et al., 1997). Zentral für einen kompetenten Vergleich mehrerer Optionen ist die Fähigkeit *trade-offs* anwenden zu können. Damit ist ein Abwägen der Vor- und Nachteile verschiedener Optionen gemeint, indem negative Aspekte einer Option durch positive Aspekte kompensiert werden können und diese *trade-offs* anschließend dazu genutzt werden, Optionen zu hierarchisieren. Da Entscheidungssituationen und deren Handlungsoptionen im Bereich der SSI oftmals dadurch gekennzeichnet sind, dass es nicht die eine richtige sondern mehrere legitime Lösungen gibt, stellen *trade-offs* eine zentrale Strategie im Umgang mit Bewertungs- und Entscheidungssituationen dar (vgl. u.a. Hogan, 1999; Ratcliffe, 1997; Seethaler & Linn, 2004; SEPUP, 1995; Wilson & Sloane, 2000; Yang & Anderson, 2003).

Bewertungskompetenz aus Sicht der Entscheidungstheorie

Bewertungskompetenz beschreibt die Fähigkeit, in komplexen Problemsituationen begründet und systematisch bei unterschiedlichen Handlungsoptionen zu entscheiden. Bewertungskompetenz fasst somit sowohl das Bewerten mehrerer Handlungsoptionen als

⁶ Im deutschsprachigen Raum wird dieser Forschungsstrang mit dem Begriff „Natur der Naturwissenschaften“ verbunden, im englischsprachigen mit dem Begriff „nature of science (NOS)“.

auch das Treffen einer Entscheidung bzw. eines entscheidungsvorbereitenden Urteils (Bögeholz et al., 2004). In der psychologischen Forschungstradition werden derartige Prozesse in der deskriptiven Entscheidungstheorie (vgl. u.a. Abelson & Levi, 1985, Jungermann, Pfister & Fischer, 2004) mit Hilfe von Modellen der Entscheidungsfindung beschrieben. Entscheidungsfindungsprozesse sind dabei in der Regel durch mehrere aufeinander folgende Phasen gekennzeichnet⁷. Ein deskriptives Metamodell der Entscheidungsfindung stellt das Modell der Entscheidungsfindung von Betsch & Haberstroh (2005, vgl. Abb. 2.1) dar.

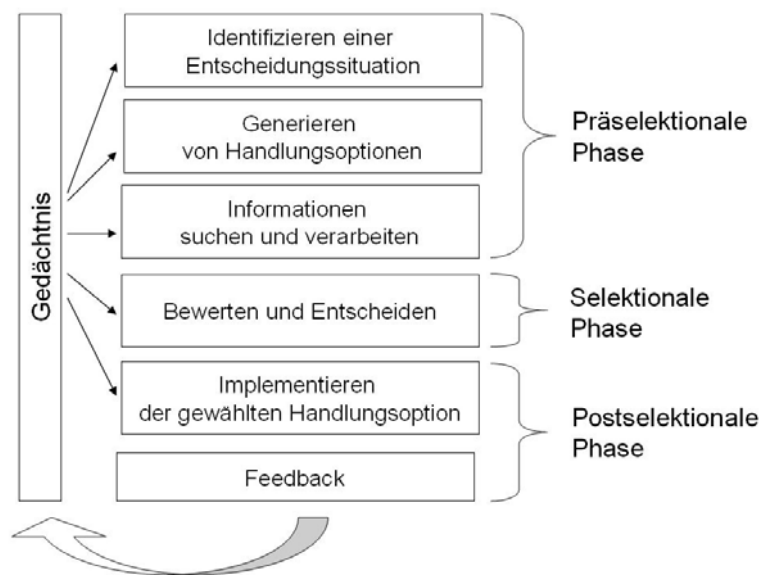


Abb. 2.1: Phasenmodell der Entscheidungsfindung (verändert nach Betsch & Haberstroh, 2005, S. 263).

In der präselektionalen Phase findet die Identifikation einer entscheidungsrelevanten Situation statt, woran sich die Entwicklung geeigneter Verhaltensoptionen anschließt. Die Entwicklung von Verhaltensoptionen erfordert dabei Informationssuch- und -verarbeitungsprozesse. In der selektionalen Phase werden die zuvor entwickelten Handlungsoptionen unter Verwendung einer Entscheidungsregel oder -strategie miteinander verglichen und bewertet. Dabei zum Tragen kommende Entscheidungsstrategien sind vielfältig und unterscheiden sich v.a. in der Intensität ihrer Informationsverarbeitung (vgl. u.a. Abelson & Levi, 1985; Jungermann et al. 2004). Neben intuitivem, spontanem Entscheidungsverhalten (vgl. Haidt, 2001; Jungermann et al., 2004) werden in der

⁷ Es kann aber auch auf Teilprozesse einer früheren Phase zurückgegriffen werden, bzw. eine Phase mehrfach durchlaufen werden.

Entscheidungstheorie non-kompensatorische von kompensatorischen Entscheidungsstrategien sowie Mischformen aus diesen beiden beschrieben (Jungermann et al., 2004; Payne, Bettmann & Luce, 1998). Non-kompensatorische Entscheidungsstrategien sind durch die Anwendung eines Ausschlussverfahrens charakterisiert. Optionen, die einen bestimmten Schwellenwert im Hinblick auf ein zentrales Entscheidungskriterium nicht erfüllen, werden ausgeschlossen. Dieser Prozess wird solange durchgeführt, bis schließlich nur noch eine Handlungsoption übrig bleibt, welche dann gewählt wird. Kompensatorische Entscheidungsregeln sind dadurch gekennzeichnet, dass Handlungsoptionen explizit miteinander verglichen werden und dabei negative Aspekte einer Handlungsoption durch positive Aspekte kompensiert werden können. Derartige Entscheidungsstrategien lassen sich mit der Anwendung von trade-offs im Rahmen von socio-scientific issues vergleichen (s.o.). Welche Entscheidungsstrategie von einem Individuum angewendet wird, hängt maßgeblich von der jeweiligen Entscheidungssituation sowie ihrer Charakteristika ab, als auch von den Erfahrungen oder Routinen, die ein Individuum in einer bestimmten Entscheidungssituation mitbringt (vgl. dazu auch die Feedback Komponente in Abb. 2.1). Das Ende der selektionalen Phase stellt dann letztendlich die Auswahl einer Handlungsoption sowie die Formulierung einer Handlungsintention dar. In der postselektionalen Phase wird die gewählte Handlungsoption in der Realität umgesetzt, wobei v.a. volitionale Aspekte eine Rolle spielen (Betsch, Haberstroh & Höhle, 2002).

Zusammenfassend lässt sich festhalten, dass das vorgestellte Metamodell wichtige Hinweise liefern kann, welche Prozesse für eine Identifikation zentraler Strukturen von Bewertungskompetenz wichtig sein könnten. Für eine Definition von empirisch überprüfbaren Kompetenzen werden dies v.a. die ersten beiden Phasen des Metamodells sein, da eine Überprüfung einer kompetenten Umsetzung von Handlungsoptionen in der Realität zwar anzustreben ist, im schulischen Kontext oftmals aber nicht realisiert werden kann.

3 Modellierung von Kompetenzen

Um von allgemeinen Beschreibungen, was eine bestimmte Kompetenz ausmacht bzw. welche Fähigkeiten darunter zu verstehen sind, zu einer Modellierung und damit auch empirischen Messbarmachung von Kompetenzen zu kommen, bedarf es mehrerer Schritte, darunter die Identifikation von zentralen Strukturen in Form von Kompetenzstrukturmodellen. Um in der empirischen Umsetzung Schwierigkeiten zu reduzieren, müssen dabei bereits in dieser theoretischen Phase empirische, d.h. auch messtheoretische, Voraussetzungen mitberücksichtigt werden (vgl. u.a. Hartig & Klieme, 2006). Diese Voraussetzungen werden im Folgenden vorgestellt. Darauf aufbauend werden im Anschluss die wichtigsten, und in dieser Studie zum Einsatz gekommenen, Messmodelle der Item-Response-Theorie⁸ (IRT) vorgestellt.

3.1 Operationalisierung und Messung von Kompetenzen

Für eine Messung von Kompetenzen ist neben den beschriebenen Forderungen der Lernbarkeit und der Spezifikation der Domäne, d.h. der Kontextualisierung von Kompetenzen (siehe Kapitel 1), v.a. die Definition von (Binnen)-strukturen hervorzuheben (Hartig & Klieme, 2006). Die Ausschärfung derartiger Binnenstrukturen ergibt sich dabei auch durch die Anforderungen der spezifischen Situation, bzw. des Kontextes, in dem sie bewältigt werden müssen (Rychen, 2001). Die theoretische Beschreibung von Binnenstrukturen mündet dabei typischerweise in der Formulierung von Kompetenzstrukturmodellen. Kompetenzstrukturmodelle beschreiben Teilkompetenzen, die aus theoretischer Sicht getrennt werden sollten und von denen man ausgeht, dass sie sich möglicherweise auch empirisch voneinander trennen lassen. Während sich Kompetenzstrukturmodelle mit Teilkompetenzen einer Kompetenz befassen (Hartig &

⁸ Im Folgenden wird ausschließlich der Begriff Item-Response-Theorie verwendet. Ursprünglich wurden der Item-Response-Theorie nur das Rasch Modell, das 2-parametrische Birnbaummodell sowie das 3-parametrische Modell zugeordnet. Heute werden jedoch alle probabilistischen Testmodelle unter dem Sammelbegriff Item-Response-Theorie subsumiert (vgl. Rost, 2006).

Klieme, 2006), beschreiben Kompetenzentwicklungsmodelle darüber hinaus die verschiedenen Ausprägungen oder Qualitäten, die auf einer Teilkompetenz erfasst werden können (ebd.; Hammann, 2004). Diese verschiedenen Ausprägungen werden als Kompetenzniveaus oder -stufen bezeichnet. Ähnlich dem Kompetenz- und Domänenbegriff liefern auch die Begriffe Niveaus, Stufen oder auch Levels Ansatz für kontroverse Diskussionen. Insbesondere der Begriff der Stufen wird im Rahmen von Kompetenzdiagnostik und Assessments heutzutage eher abgelehnt (vgl. u.a. Helmke & Hosenfeld, 2004), da er eine Vorstellung von Kompetenzentwicklung suggeriert, die mit der kognitiven oder moralischen Entwicklung eines Individuums wie in den Arbeiten von Piaget und Kohlberg vergleichbar ist (vgl. Hartig & Klieme, 2006). In der kognitiven Entwicklungspsychologie sind Entwicklungsstufen dadurch charakterisiert, dass von einer Stufe zur nächsten qualitativ neue Leistungen oder Handlungen erbracht werden können, die vorher nicht möglich waren (vgl. u.a. Helmke & Hosenfeld, 2004). Eine derart strikte Unterteilung in Stufen, die zudem mit eindeutig qualitativen Entwicklungssprüngen einhergeht, ist für Kompetenzmessungen eher unrealistisch.

Eine Formulierung von Kompetenzniveaus ist derzeit stark an empirischen Ergebnissen orientiert. Auch wenn es ideal ist, eine theoretische Vorstellung von der Entwicklung einer Kompetenz zu haben, welche als Basis für eine Beschreibung von Kompetenzniveaus dienen kann, so bestimmen in der Praxis häufig zu einem großen Maße die statistischen Auswertungsverfahren, welche Kompetenzniveaus definiert werden und an welchem Punkten der Kompetenzskala die Grenzen zwischen zwei Kompetenzniveaus gezogen werden. So konstatiert auch Rost, dass es wohl noch keine empirische Studie gegeben habe, wo die a priori angenommene Struktur von Daten auf Anhieb bestätigt wurde (Rost, 2006, S. 8).

In den PISA Untersuchungen geschieht die Setzung von Kompetenzniveaus typischerweise durch die Einteilung der Kompetenzskala in mehrere Teilabschnitte, deren Festlegung auf Basis von statistischen Auswertungsverfahren, dem sogenannten *scale anchoring*, geschieht (Beaton & Allen, 1992) und somit zu einem gewissen Grad nicht unbedingt willkürlich aber pragmatisch ist:

„Dividing [...] these continua into levels, though useful for communication about students' development, is essentially arbitrary” (Adams & Wu, 2002, S. 197).

Im Anschluss werden die einzelnen Kompetenzniveaus mittels post-hoc Analysen der Inhalte von Aufgaben, die sich in diesem Niveaubereich befinden, beschrieben und anschließend mit den z.T. zuvor theoretisch beschriebenen Niveaus in Deckung gebracht. Diese Vorgehensweise, Kompetenzniveaus anhand von zentralen Aufgabencharakteristiken

festzulegen, wird auch als kriteriumsorientierte Vorgehensweise beschrieben (vgl. u.a. Watermann & Klieme, 2002).

Im Gegensatz zu dieser v.a. empirisch geleiteten Vorgehensweise der Formulierung von Kompetenzniveaus, gibt es andere kompetenzdiagnostische Programme, in denen eine eher theoriegeleitete Vorgehensweise favorisiert wird, die jedoch mit anderen Problemen umgehen müssen. Ein Beispiel stellen die VERA („Vergleicharbeiten in der Grundschule“; vgl. Helmke & Hosenfeld, 2004) sowie die DESI (Deutsch-Englisch-Schülerleistungen International; vgl. Beck & Klieme, 2007) Untersuchungen dar. In DESI wurden die Testaufgaben vor ihrem Einsatz anhand von schwierigkeitsbestimmenden Charakteristika beschrieben (Hartig, 2007). Mit Hilfe dieser Merkmale sollen Aufgabenschwierigkeiten sowie Schwierigkeitsunterschiede zwischen Aufgaben vorhergesagt werden können und darüber hinaus als Grundlage auch für eine Neukonstruktion von Aufgaben dienen (Hartig, 2007). Doch auch bei einem derartigen eher theoriegeleiteten Vorgehen treten in der Formulierung von Kompetenzniveaus Probleme auf, darunter z.B. die Spannweite der einzelnen Kompetenzniveaus. Während in den PISA Studien durch eine post-hoc Festlegung der Schwellen zwischen den Niveaus gleichgroße Kompetenzniveaus festgelegt wurden, sind die Kompetenzniveaus in den DESI Untersuchungen, z.B. für das Lese- und Hörverstehen im Fach Englisch, unterschiedlich groß. Zudem unterscheiden sich die Niveauschwellen zwischen den beiden Kompetenzbereichen Lese- und Hörverstehen im Hinblick auf ihre Schwierigkeit. Beide Aspekte stellen Hindernisse für eine Kommunizierbarkeit gegenüber verschiedensten Interessengruppen dar (Willenberg, 2007).

Zusammenfassend für die Modellierung von Kompetenzen lässt sich feststellen, dass nicht so sehr die Identifikation von Teilkompetenzen, und somit von relevanten Strukturen einer Kompetenz, Schwierigkeiten bereitet, sondern vielmehr die Definition von Entwicklungsniveaus einer Teilkompetenz bzw. allen Teilkompetenzen einer Kompetenz. Dabei muss konstatiert werden, dass eine theoriegeleitete Modellentwicklung wünschenswert ist und gefordert werden sollte, dass diese aber auch zukünftig an ihre empirischen Grenzen stoßen wird und somit

„die Grenzen zwischen einem Post-Hoc-Vorgehen und einem streng hypothesengeleiteten und modellbasierten Vorgehen *wohl fließend bleiben werden*“ (Hartig, 2007, S. 97; *kursiv verändert durch S.E.*).

3.2 Modellierung mit quantifizierenden eindimensionalen IRT-Modellen

Um theoretisch formulierte Kompetenzstrukturmodelle empirisch zu überprüfen und ggf. zu modifizieren, bedarf es geeigneter Testinstrumente. Eine Eignung kann jedoch erst mit Hilfe von psychometrischen Testmodellen geprüft werden. Typischerweise werden dazu in *large scale assessments* wie den TIMSS und PISA Studien, aber auch in *small scale assessments* Testmodelle der Item-Response Theorie eingesetzt. Diese psychometrischen Modelle haben im Vergleich zu klassischen Testmodellen besonders für Kompetenzmessungen folgende Vorteile:

- a. die Möglichkeit, Fähigkeiten von Personen und Schwierigkeiten von Aufgaben in einem Modell zu integrieren sowie
- b. die Möglichkeit, Kompetenzentwicklungen kriteriumsorientiert beschreiben zu können.

Auf Basis dieser Beschreibungen können dann Schülerinnen und Schüler gezielt gefördert werden. Das am häufigsten verwendete psychometrische Modell aus der Familie der Item-Response Modelle ist das Raschmodell für dichotome Daten (Rasch, 1960). Eine Erweiterung für polytome Daten stellt das Partial Credit Modell dar (Masters, 1982)⁹. Im Folgenden wird näher auf diese beiden Modelle eingegangen, da diese für die Modellierung von Bewertungskompetenz in der vorliegenden Arbeit zentral sind.

Das Raschmodell

Das Raschmodell nimmt an, dass das beobachtete Antwortverhalten von Individuen in einem bestimmten Test (bzw. in einem bestimmten Testitem) auf eine latente Personenfähigkeit, ein latentes trait, zurückzuführen ist. Im Raschmodell wird diese Fähigkeit bzw. diese Kompetenz der getesteten Personen in Beziehung gesetzt zur Lösungswahrscheinlichkeit der eingesetzten Items (Hartig & Klieme, 2006). Zentrale Größen im Raschmodell sind somit Personenfähigkeit und Itemschwierigkeit:

$$P(X_i = 1 \mid \theta, \sigma_i) = \frac{e^{(\theta - \sigma_i)}}{1 + e^{(\theta - \sigma_i)}} \quad (1.1)$$

9 Beide Modelle gehören zu den sogenannten 1-Parameter-Modellen, da sie lediglich einen Itemparameter, nämlich die Itemschwierigkeit, berücksichtigen. Darüber hinaus gibt es auch 2- bzw. 3-Parameter-Modelle, die neben dem Itemparameter zusätzlich einen Trennschärfeparameter bzw. einen Rateparameter berücksichtigen (vgl. Bühner, 2006).

wobei P die Wahrscheinlichkeit beschreibt das Item i korrekt zu lösen, θ die Personenfähigkeit und σ die Itemschwierigkeit des Items i . Entscheidend für die Lösungswahrscheinlichkeit ist dabei die Differenz zwischen Personenfähigkeit und Itemschwierigkeit. Die Beziehung zwischen Personenfähigkeit und Itemschwierigkeit lässt sich durch sogenannte Itemcharakteristikkurven, auch Item-Characteristic-Curves (ICCs) genannt, visualisieren (vgl. Abb. 3.1).

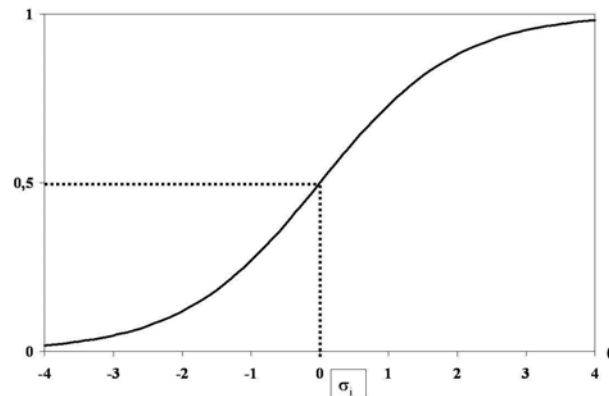


Abb. 3.1: Item-Characteristic-Curve: Zusammenhang zwischen der Personenfähigkeit (θ) und der Wahrscheinlichkeit das Item i zu lösen (verändert nach Rost, 2004, S. 120)

Die Abbildung zeigt eine ICC für ein Item, wobei auf der horizontalen Achse die Personenfähigkeit bzw. die Itemschwierigkeit und auf der vertikalen Achse, die Wahrscheinlichkeit das Item i korrekt zu lösen dargestellt sind. Sowohl Personenfähigkeit als auch Itemschwierigkeit sind auf einer logarithmierten Metrik¹⁰ abgebildet. Die Abbildung zeigt, dass bei einer Itemschwierigkeit von 0 *logits* die Personenfähigkeit ebenfalls bei 0 *logits* liegt. Die Wahrscheinlichkeit, das Item i korrekt zu lösen, liegt bei 50%. Diese Lösungswahrscheinlichkeit lässt sich exemplarisch anhand der Gleichung des Raschmodells nachvollziehen:

$$\begin{aligned}
 P(X_i = 1 \mid 0.0, 0.0) &= \frac{e^{(0.0-0.0)}}{1 + e^{(0.0-0.0)}} & (1.2.) \\
 &= \frac{e^{(0.0)}}{1 + e^{(0.0)}}
 \end{aligned}$$

¹⁰ Die Umwandlung von Rohwerten in logit measures, auch logits genannt, hat den Vorteil, dass die Fähigkeiten von Personen am unteren sowie am oberen Ende einer Kompetenzskala adäquater abgebildet werden als es bei Rohwerten der Fall ist. Um von Rohwerten zu logits zu kommen, muss in einem ersten Schritt die Lösungswahrscheinlichkeit einer Person in Bezug auf ein Item durch seine Gegenwahrscheinlichkeit geteilt werden (der sogenannte Wettquotient). In einem zweiten Schritt wird der Wettquotient logarithmiert, was man dann als den Logit der Lösungswahrscheinlichkeit bezeichnet (vgl. Rost, 2004, S. 117 ff.).

$$= \frac{1}{1+1}$$

$$= 0.50.$$

Steigt die Personenfähigkeit an, so erhöht sich die Lösungswahrscheinlichkeit, sinkt die Personenfähigkeit unter 1 logit, so verringert sich auch die Wahrscheinlichkeit, das Item korrekt zu lösen. In die Abb. 3.1 könnten theoretisch alle Items eines eindimensionalen Tests integriert und somit die Items ihrer Schwierigkeit nach geordnet werden¹¹.

Die beschriebene gemeinsame Metrik von Personenfähigkeiten und Itemschwierigkeiten im Raschmodell ist eine Voraussetzung für eine spätere Unterteilung einer Kompetenzskala in unterschiedliche Kompetenzniveaus und einer kriteriumsorientierten Beschreibung dieser Niveaus (Hartig & Klieme, 2006).

Das Partial Credit Modell

Eine Erweiterung des dichotomen Raschmodells stellt das Raschmodell für polytome Daten, das sogenannte Partial Credit Modell, dar (Masters, 1982). Ebenso wie das dichotome Raschmodell beschreibt das Partial Credit Modell eine beobachtete Personenfähigkeit durch ein latentes trait. Der Unterschied zum dichotomen Raschmodell besteht darin, dass ordinale Daten mit mehreren Antwortkategorien modelliert werden können.¹² In Abb. 3.2 sind analog zu einer ICC für ein dichotomes Item die sogenannten *Characteristic-Category-Curves* (CCCs) für ein polytomes Item dargestellt.

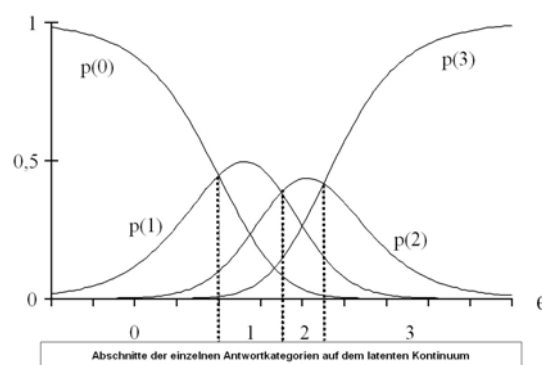


Abb. 3.2: Category-Characteristic-Curves für ein polytomes Item im Partial Credit Modell (verändert nach Rost, 2004, S. 204)

¹¹ Eine vereinfachte Darstellung aller in einem Test untersuchten Individuen sowie aller eingesetzten Items stellen sogenannte Person-Item-Maps dar, die für eine erste Visualisierung der Schwierigkeit eines Tests häufig hinzugezogen werden.

¹² Im Partial Credit Modell können auch Kombinationen von dichotomen und polytomen Items modelliert werden.

Die CCCs im Partial Credit Modell sind im Vergleich zur ICC im dichotomen Raschmodell nicht alle monoton steigend bzw. monoton abfallend. Die Abb. 3.2. zeigt die Kategorienfunktionen für ein polytomes Item mit vier Antwortkategorien (0-3). Die CCCs beschreiben die Antwortwahrscheinlichkeit der einzelnen Kategorien von der latenten Variablen. Wenn man auf dem latenten Kontinuum, d.h. einer bestimmten Personenfähigkeit, von links nach rechts wandert, dann hat zunächst die erste Antwortkategorie die höchste Wahrscheinlichkeit. Diese Wahrscheinlichkeit verringert sich dann zur nächsten Antwortkategorie hin, bis diese dann wiederum zur nächsten hin abnimmt usw. Die Kurvenschnittpunkte unterteilen das latente Kontinuum in mehrere Abschnitte, in denen jeweils eine Antwortkategorie die höchste Wahrscheinlichkeit hat. Die Schnittpunkte zwischen zwei Antwortkategorien werden auch als Schwellen (englisch: *thresholds*) bezeichnet. Für die Entwicklung eines Testinstruments mit polytomen Items ist es notwendig, dass diese Schwellen, und damit die verschiedenen Antwortkategorien auf dem latenten Kontinuum geordnet sind. Ist dies nicht der Fall, dann kann einer bzw. mehreren Antwortkategorien kein Abschnitt auf dem latenten Kontinuum zugeordnet werden. Daraus folgt, dass ein, durch die verschiedenen Antwortkategorien angenommener Schwierigkeitsanstieg in den Daten nicht abbildbar ist. Daraus ergeben sich vielfältige Konsequenzen für die psychometrische Modellierung, darunter u.a. die Notwendigkeit Antwortkategorien, die nicht in ihrer Schwierigkeit ansteigen, zusammenzuschmelzen.

3.3 Modellierung mit IRT-Modellen jenseits der Eindimensionalität

Kann ein Test nicht mit Hilfe eines eindimensionalen Modells modelliert werden, bedeutet dies, dass keine homogene Personengruppe bzw. kein homogenes Antwortverhalten vorliegt. Eine Ursache kann darin begründet sein, dass der Test schlecht konstruiert ist, eine andere mögliche Ursache ist, dass ihm eine andere psychometrische Struktur zugrunde liegt (Rost, 2006, S. 267). Rost (ebd.) nennt drei alternative psychometrische Strukturen, die das Antwortverhalten auf einen bestimmten Test erklären können, die im Folgenden erläutert werden sollen.

Latente Klassenanalyse, Multidimensionale Raschmodelle und Mixed Rasch Modelle

Zum einen kann ein Test nicht ein latentes trait, sondern mehrere latente Typen und damit unterschiedliche Arten von Antwortverhalten oder Lösungsstrategien abbilden. Dieses

Modell wäre demnach kein quantifizierendes sondern ein qualifizierendes, welches Personen unterschiedlichen Klassen von Antworttypen zuordnet. Ein derartiges Modell stellt die Latent-Class-Analyse (LCA) dar. Die zweite Alternative umfasst Modelle, die eine Mehrdimensionalität von Testitems in einem Test modellieren können. Dabei lässt sich zwischen *between item multidimensionality* und *within item multidimensionality* unterscheiden (Abb. 3.3). Im ersten Fall lassen sich die Items eines Tests genau je einer Dimension zuordnen. In der Abbildung sind dies drei latente Dimensionen, die jeweils durch drei Items gemessen werden. Dadurch wird der Test in mehrere voneinander abzugrenzende Subtests unterteilt. Die Beziehungen zwischen den Dimensionen werden anschließend anhand der gewonnenen Messwerte untersucht. Im zweiten Fall der *within item multidimensionality* laden die Items auf mehrere Dimensionen gleichzeitig. In einem solchen Modell werden die Korrelationen zwischen den einzelnen Dimensionen mitmodelliert. Der Ansatz der *between item multidimensionality* ist ein häufig verwendetes Verfahren in der Modellierung von Kompetenzen. Man nimmt jedoch dabei niedrigere Korrelationen zwischen den einzelnen Dimensionen in Kauf als beim Ansatz der *within item multidimensionality*¹³.

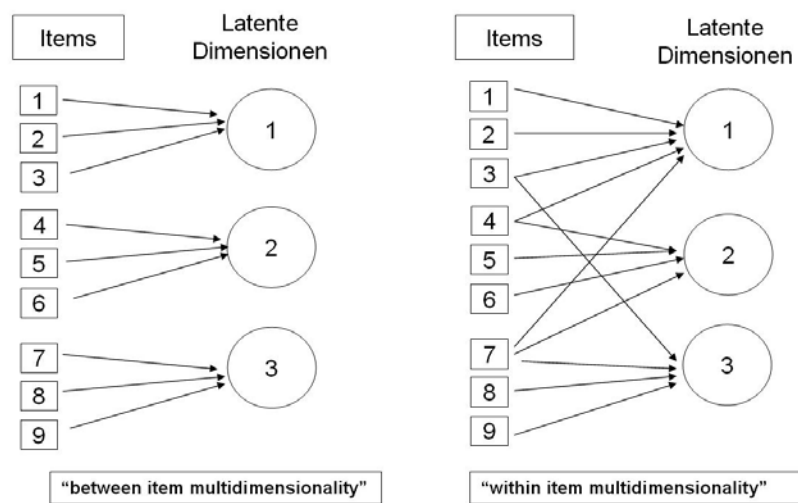




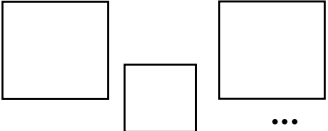
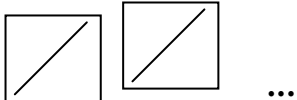
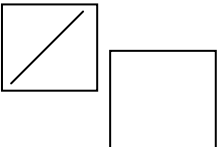
Abb. 3.3: Unterschiede in der Modellierung von Mehrdimensionalität (verändert und übersetzt nach Wu et al., 1998, S. 115)

Die dritte von Rost beschriebene Alternative ist das Mixed Rasch Modell (MRM). Das MRM nimmt mehrere Teilpopulationen innerhalb einer Population an, in denen jeweils das

¹³ Auf diese multidimensionalen Modelle wird im Folgenden nicht näher eingegangen, da sie im Rahmen dieser Arbeit als Konsequenz der Forschungsstrategie (siehe Kapitel 4.2) nicht verwendet wurden.

Raschmodell gilt. Somit vereint das Mixed Rasch Modell das quantifizierende Raschmodell mit der qualifizierenden Latent-Class-Analyse. Die Zusammenhänge dieser Modelle werden deutlich, wenn man lediglich eine latente Klasse annimmt, in der das Raschmodell gilt. In diesem Fall ist das Mixed Rasch Modell gleichbedeutend mit dem Raschmodell (Rost, 2004). Generell wird die optimale Anzahl der latenten Klassen in einem Mixed Rasch Modell durch modellvergleichende Tests ermittelt, die idealerweise mit der Anzahl theoretisch angenommener, qualitativ unterschiedlicher, Klassen übereinstimmen (Rost, 2006, S. 269). Eine Modifizierung des Mixed Raschmodells stellen sogenannte Hybridmodelle dar (Yamamoto, 1989; Davier & Yamamoto, 2007). Hybridmodelle sind eine Mischung aus Teilpopulationen in denen das Raschmodell gilt (skalierbare Klassen) sowie Teilpopulationen die unskalierbar sind und somit eine latente Klasse darstellen. In der folgenden Tabelle sind alle beschriebenen psychometrischen Modelle, die in der Modellierung von Bewertungskompetenz im Rahmen dieser Arbeit zum Einsatz gekommen sind, dargestellt.

Tab. 3.1: Überblick der für die Studie relevanten IRT-Modelle

Modell	Beschreibung	Skalierung bzw. Unskalierbarkeit der Klassen*
Raschmodell (Rasch, 1960)	quantifizierendes, eindimensionales Modell für dichotome Daten	
Partial Credit Modell (Masters, 1982)	quantifizierendes, eindimensionales Modell für polytome Daten	
Latent-Class-Analyse (Lazarsfeld & Henry, 1968)	qualifizierendes Modell für dichotome als auch polytome Daten	
Mixed Rasch Modell (Rost, 1990)	quantifizierendes und qualifizierendes Modell für dichotome und polytome Daten, wobei in allen latenten Klassen das Raschmodell gilt	
Hybrid Modell (Yamamoto, 1989)	quantifizierendes und qualifizierendes Modell für dichotome und polytome Daten, wobei eine Kombination aus skalierbaren und unskalierbaren latenten Klassen angenommen werden kann	

*Anmerkung: Die grafische Darstellung der einzelnen Modelle wurde in Anlehnung an Rost, 2004, S. 171 vorgenommen. Eine Gerade innerhalb eines Kästchens repräsentiert Skalierbarkeit im Sinne einer quantitativen PersonenvARIABLE.

3.4 Modellgeltungstests für IRT-Modelle

„Jede Testauswertung beruht auf einem Modell über das Antwortverhalten der Personen in einem Test. Ob die Testwerte etwas über die getesteten Personen aussagen und was sie bestenfalls aussagen können, hängt davon ab, ob das bei der Auswertung angewendete Modell überhaupt auf die Daten passt“ (Rost, 2004, S. 330).

Der Frage, welches eingesetzte Modell am besten auf die Daten passt, wird durch Modellgeltungstests nachgegangen (ebd.). Zu den statistischen Modellgeltungsmaßen gehören a) inferenzstatistische Maße, b) informationstheoretische Maße und c) weitere speziellere Modellgeltungstests. Zu den ersten gehören der Likelihoodquotiententest, die Chi²-Statistik sowie das sogenannte Bootstrapverfahren. In der vorliegenden Arbeit kam aufgrund der für IRT-Verfahren relativ geringen Stichprobe sowie der, im Vergleich zu den theoretisch möglichen, beobachteten Antwortpattern das Bootstrapverfahren (Davies, 1997; Rost, 2004) zum Einsatz. Das Bootstrapverfahren ist ein Resimulationsverfahren, bei dem auf Basis des vorhandenen Datensatzes künstliche Datensätze simuliert werden. Für jeden simulierten Datensatz werden, wie für den echten Datensatz, zwei Prüfgrößen (Cressie-Read-Statistik und Pearson-Statistik) berechnet. Die Prüfgrößen der simulierten Datensätze werden dann mit der Prüfgröße für den echten Datensatz verglichen. Üblicherweise wird ein Modell dann bestätigt, wenn die Fit-Statistik des echten Datensatzes kleiner ist als mindestens 5% Werte der simulierten Datensätze. Somit stellt die 5% Signifikanzgrenze in diesem Fall die Grenze dar, bei der die Null-Hypothese bestätigt wird und damit das eingesetzte Modell die Daten hinreichend gut beschreiben kann. Können mehrere Modelle in Bezug auf diese Signifikanzgrenze die Daten erklären, sollte das einfachste Modell, d.h. dasjenige mit den wenigsten Klassen, gewählt werden (Davies, 1997, S. 302).

Zu den informationstheoretischen Maßen zählen die sogenannten Informationskriterien. Mit Hilfe von Informationskriterien können verschiedene Modelle miteinander verglichen und Aussagen darüber getroffen werden, welches Modell die Daten relativ am besten erklärt. Informationskriterien liefern aber keine Aussage darüber, welches Modell das eindeutig Beste ist (vgl. Rost, 2004). Generell lassen sich drei informationstheoretische Maße unterscheiden: das Akaiikes information criterion (AIC), das Bayes information criterion (BIC) sowie das consistent Akaiikes information criterion (CAIC), wobei bei den letzten beiden Maßen die Stichprobengröße berücksichtigt wird. Für das Raschmodell gibt es darüber hinaus weitere spezielle Modellgeltungstests, die v.a. die Forderung auf Personen- bzw. Itemhomogenität untersuchen. Die bekanntesten Tests sind der Andersen-Test auf

Personenhomogenität und der Martin-Löf Test auf Itemhomogenität. Die Personenhomogenität kann alternativ mit Hilfe von grafischen Modelltests untersucht werden. Dabei werden die Itemparameter für spezielle Subgruppen, z.B. Klassenstufen, in einem Kreuzdiagramm miteinander verglichen. Wenn die Itemparameter der beiden Subgruppen übereinstimmen, liegen alle Punkte auf einer 45° Linie (vgl. Rost, 2004).

Von den beschriebenen Modellgeltungstests kamen in den drei empirischen Studien des zweiten Forschungsschwerpunkts das Bootstrapverfahren, der Modellvergleich anhand von Informationskriterien sowie grafische Modelltests zum Einsatz.

4 Ableitung der Forschungsstrategie

4.1 Forschungsschwerpunkt I: Theoretische Entwicklung eines Kompetenzstrukturmodells für Bewertung

In den einleitenden Kapiteln wurde deutlich, dass die Begriffe Kompetenz und Bewertungskompetenz facettenreich sind. Heterogenitäten in der Begriffsdefinition sowie in der Abgrenzung zu anderen Konstrukten wurden dargelegt. Im Bezug auf Bewertungskompetenz wurden fachdidaktische Ansätze vorgestellt, die in Abhängigkeit von ihrem spezifischen Kontext, in unterschiedlichen Forschungstraditionen stehen. Diese Ansätze stellen dabei für bzw. aus der Unterrichtspraxis heraus entwickelte Unterrichtsmodelle dar.

Mit der Outputorientierung der Bildungsstandards rückte in den sprachlichen, mathematischen und den naturwissenschaftlichen Fächern die Notwendigkeit einer Überprüfung der in den Standards formulierten Kompetenzen ins Zentrum der fachdidaktischen und schulpolitischen Diskussionen. Diese Forderung trifft in besonderem Maße für den Bereich Bewertungskompetenz zu. Er stellt im Gegensatz zu den für den Unterricht in den naturwissenschaftlichen Fächern etablierten Kompetenzbereichen Fachwissen und Erkenntnisgewinnung insbesondere für Lehrkräfte einen relativ neuartigen Kompetenzbereich dar. Mit dieser hervorgehobenen Bedeutung ist gleichzeitig eine Zurückhaltung auf Seiten der Lehrkräfte aber auch auf Seiten von Lehrerbildungsinstitutionen zu verzeichnen, Bewertungskompetenz stärker in den Unterricht zu integrieren. Um eine Integration stärker voranzutreiben und um die Erreichung der Standards bei Schülerinnen und Schülern abprüfen zu können, bedarf es in einem ersten Schritt einer Auspräzisierung des Kompetenzbereichs Bewertung in Form eines Kompetenzstrukturmodells. Ein derartiges Modell, welches bereits auf eine empirische Überprüfung ausgerichtet ist, lag für Bewertungskompetenz zum Zeitpunkt dieser Arbeit nicht vor. Aus diesem Grund stellt die theoretische Herleitung zentraler Strukturen von Bewertungskompetenz in Form eines Kompetenzstrukturmodells den ersten Forschungsschwerpunkt dieser Arbeit dar.

Dazu wird ein Modell für Bewertungskompetenz, kontextualisiert für den Bereich Nachhaltige Entwicklung, auf Basis bestehender fachdidaktischer Ansätze sowie Erkenntnissen aus der psychologischen Entscheidungstheorie abgeleitet. Für eine anschließende Operationalisierung im zweiten Forschungsschwerpunkt werden kompetenzdiagnostische Voraussetzungen und Erkenntnisse aus Schulleistungsstudien im Rahmen von naturwissenschaftlicher Grundbildung im Bezug auf die Identifikation von Strukturen und im Hinblick auf eine mögliche Entwicklung von Kompetenzen berücksichtigt.

Aus den genannten Zielen generieren sich spezifische Forschungsfragen für den Forschungsschwerpunkt I:

1. Welche zentralen Strukturen von Bewertungskompetenz lassen sich auf Basis der fachdidaktischen Vorarbeiten sowie der psychologischen Entscheidungstheorie ableiten?
2. Welche möglichen Entwicklungsverläufe für Bewertungskompetenz - insbesondere für die im zweiten Forschungsschwerpunkt operationalisierte Teilkompetenz - lassen sich beschreiben?

4.2 Forschungsschwerpunkt II: Operationalisierung und Messung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“

Studie 1

Ergebnis des Forschungsschwerpunkts I ist das Kompetenzstrukturmodell für Bewertung. Das Modell besteht aus vier, theoretisch sinnvoll voneinander zu trennenden, Teilkompetenzen (vgl. Kap. 5). Für die empirische Überprüfung im Forschungsschwerpunkt II wurde zunächst die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ ausgewählt.¹⁴ Diese Teilkompetenz stellt, im Vergleich zu den drei anderen Teilkompetenzen, eine für den Biologieunterricht neuartige Kompetenz sowohl aus der Perspektive der Forschung als auch aus der Perspektive der Unterrichtsentwicklung dar. Für diese Teilkompetenz besteht somit der größte Handlungsbedarf. Unter „Bewerten,

¹⁴ Die Studien des zweiten Forschungsschwerpunktes fokussieren zunächst auf eine Teilkompetenz und damit auf die Entwicklung von Testaufgaben für diese Teilkompetenz. Daraus ergibt sich für die zukünftige psychometrische Modellierung zunächst ein between item multidimensionality Ansatz.

Entscheiden und Reflektieren“ wird die Fähigkeit verstanden, mögliche Optionen in einer Bewertungs- und Entscheidungssituation unter Anwendung einer Entscheidungsstrategie miteinander vergleichen zu können. Fachdidaktische Forschung zur Anwendung von Entscheidungsstrategien fokussiert dabei besonders auf die Fähigkeit von Schülerinnen und Schülern, *trade-offs* in Entscheidungssituationen anwenden zu können.

Unter *trade-off* versteht man die Fähigkeit, die Vor- und Nachteile möglicher Optionen gegeneinander abzuwägen. Dabei können negative Aspekte einer Option durch positive kompensiert werden. Forschungsergebnisse zu diesem Aspekt von Bewertungskompetenz sind divers. Einige Studien konnten zeigen, dass Schülerinnen und Schüler Schwierigkeiten in der Anwendung von *trade-offs* haben. Ratcliffe (1997) analysierte Gruppendiskussionen von 15-Jährigen und konnte zeigen, dass Vor- und Nachteile möglicher Optionen zwar genannt werden, diese aber nicht systematisch miteinander verglichen wurden. Andere Studien zeigten, dass beispielsweise 13-Jährige ganz auf die Anwendung von *trade-offs* verzichteten, und mögliche Optionen anhand eines oder weniger Aspekte analysierten und dabei ebenfalls unsystematisch vorgehen (Hong & Chang, 2004; Hogan, 1999). Andere Studien zeigten bessere Ergebnisse im Hinblick auf die Anwendung von *trade-offs*. Ergebnisse im Rahmen des SEPUP-Projekts weisen darauf hin, dass sich die Anwendung von *trade-offs* bei Siebt- und Neuntklässlern v.a. nach längeren Unterrichtsinterventionen verbesserte (Wilson & Sloane, 2000). Diese Ergebnisse konnten eingeschränkt von Siegel (2006) im Hinblick auf 16-Jährige sowie ohne Einschränkungen von Seethaler & Linn (2004) bei Acht-Klässlern bestätigt werden. Darüber hinaus konnten zahlreiche Studien zeigen, dass Schülerinnen und Schüler, aber auch Studierende generell Schwierigkeiten haben, Entscheidungskriterien bzw. vorgenommene *trade-offs* zu gewichten und diese Gewichtung durch individuelle Werthaltungen zu begründen (vgl. u.a. Barkmann & Bögeholz, 1999; Große & Bögeholz, 2005; Jimenez-Aleixandre, 2002; Seethaler & Linn, 2004). Ursachen hierfür können darin begründet sein, dass durch eine Offenlegung von individuellen Werthaltungen eine Wertehierarchisierung vorgenommen werden muss und dabei eindeutig einige Werte anderen untergeordnet werden.

Integriert man die beschriebenen Ergebnisse fachdidaktischer Forschung mit der psychologischen Entscheidungstheorie, lässt sich die Fähigkeit *trade-offs* anzuwenden als eine kompensatorische Entscheidungsstrategie beschreiben (vgl. Kap. 2). Darüber hinaus lassen sich jedoch bei Individuen weitere Entscheidungsstrategien beobachten. Hong & Chang (s.o.) beobachteten in einem Kaufentscheidungsszenario, dass Schülerinnen und Schüler aufgrund von einem Kriterium entschieden und Optionen ausschlossen. Ein Ausschließen von Optionen aufgrund eines Kriteriums, bzw. der qualitativen Ausprägung einer Option im Hinblick auf ein Kriterium, wird als non-kompensatorische Strategie

bezeichnet, Positive Aspekte einer Option fallen nicht ins Gewicht, sobald eine Option durch eine schlechte Ausprägung ausgeschlossen wird. Diese *elimination-by-aspects* Strategie wurde zuerst von Tversky (1972) beschrieben. In Abhängigkeit der unterschiedlichen Komplexität von Entscheidungssituationen, d.h. v.a. der Anzahl von Optionen und/oder Entscheidungskriterien, werden die beschriebenen Strategien auch miteinander kombiniert bzw. sequenziert (Jungermann et al., 2004). Ergebnisse experimenteller Studien konnten zeigen, dass mit einem Anstieg gegebener Optionen sich die Wahrscheinlichkeit erhöht, dass Individuen eine non-kompensatorische Entscheidungsstrategie anwenden (Payne et al., 1998).

Neben diesen eher bewusst angewandten Entscheidungsstrategien, lässt sich bei Individuen auch immer spontanes bzw. intuitives Entscheidungsverhalten beobachten. Ein Beispiel hierfür sind Routineentscheidungen (Betsch & Haberstroh, 2005). Haidt postuliert sogar, dass, zumindest für den Bereich der moralischen Urteilsbildung, alle Entscheidungen, die wir treffen, zunächst intuitiv getroffen werden (Haidt, 2001). Erst im Nachhinein werden Gründe gesucht und angeführt, die eine getroffene Entscheidung stützen sollen. Typischerweise werden dabei Gründe, die eine Entscheidung in Frage stellen könnten, abgeschwächt bzw. ignoriert. Haidt spricht in diesem Zusammenhang von rechtfertigenden Entscheidungen. Integriert man die Erkenntnisse der Entscheidungstheorie und der fachdidaktischen Forschung, so lassen sich mehrere Entscheidungsstrategien identifizieren, die für eine Operationalisierung von „Bewerten, Entscheiden und Reflektieren“ berücksichtigt werden sollten. Dies sind auf der einen Seite kompensatorische und non-kompensatorische, sowie Mischformen aus diesen beiden Strategien und intuitive bzw. rechtfertigende Entscheidungsstrategien auf der anderen Seite.

Ziel der ersten Studie war es somit nicht nur auf eine Anwendung von *trade-offs* zu fokussieren, sondern zu untersuchen, welche der in der Entscheidungstheorie beschriebenen Entscheidungsstrategien sich bei Schülerinnen und Schülern beobachten lassen. Da die Auswahl und Anwendung einer bestimmten Strategie von der Komplexität (s.o.), dem Kontext sowie der Darstellungsweise der Entscheidungssituation abhängt, wurden diese Aspekte bei der Entwicklung der Testaufgaben berücksichtigt. Im Hinblick auf die Komplexität wurden alle Testaufgaben mit vier Entscheidungsoptionen konstruiert. Dies repräsentiert eine realistische Anzahl von Optionen in Gestaltungssituationen Nachhaltiger Entwicklung, die Schülerinnen und Schüler bearbeiten können sollten. In gleicher Weise wurde die Anzahl der Kriterien, anhand derer die Entscheidungsoptionen näher beschrieben wurden, beschränkt. Zudem wurde ein offenes Aufgabenformat gewählt, da der Fokus des Forschungsinteresses auf den Entscheidungsstrategien, und damit auf der Argumentationsstruktur, und nicht so sehr auf der getroffenen Entscheidung lag. Im Bezug

auf den Kontext der Aufgaben wurden in dieser Studie sowohl nicht biologische, teilweise biologische sowie Themen Nachhaltiger Entwicklung ausgewählt, um mögliche Unterschiede in der Auswahl einer Entscheidungsstrategie in Abhängigkeit des Kontextes identifizieren zu können. Die Variation des Kontextes geht dabei auch der Frage nach, inwiefern es sich bei der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ eher um eine fächerübergreifende Kompetenz, in Anlehnung an die Messung von *cross-curricular-competencies* in den PISA Studien handelt (vgl. Kap. 2), oder ob der Kontext einen Einfluss auf die Auswahl von Entscheidungsstrategien hat.

Ein zweiter Schwerpunkt der ersten, sowie der im Folgenden beschriebenen zweiten und dritten Studie, stellt die psychometrische Modellierung dar. Im Rahmen von Kompetenzmessungen in Schulleistungsstudien kommen in der Regel dichotome, teilweise auch polytome, Raschmodelle zum Einsatz, da man davon ausgeht, dass zumindest die Teilkompetenzen einer Kompetenz auf einer eindimensionalen Kompetenzskala abgebildet werden können. Eindimensionale Kompetenzskalen haben auch den nicht zu unterschätzenden Vorteil, dass verschiedene Kompetenzniveaus von Schülerleistungen auch für methodische Laien verständlich kommuniziert werden können. Alle Teilkompetenzen einer Kompetenz werden dann entweder nach dem *between item dimensionality* Ansatz (vgl. Kap. 3) einzeln oder nach dem *within item dimensionality* Ansatz gleichzeitig in einem Modell modelliert.

Derzeit existieren, nach Kenntnis der Autorin, bis auf die Kompetenzmodellierung im Rahmen des SEPUP Projekts, noch keine Ansätze, Bewertungskompetenz mit Hilfe von IRT-Modellen zu modellieren. Im SEPUP Projekt wird im Rahmen der Operationalisierung der Teilkompetenz *evidence and trade-offs*, wiederum auf die Anwendung von *trade-offs* fokussiert. Diese Teilkompetenz wird dann mit den weiteren Teilkompetenzen von *decision-making-competence* (vgl. Kap. 5) in einem multidimensionalen Raschmodell modelliert.

Für eine Modellierung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ sollen im Gegensatz zum SEPUP Vorgehen alle in der Entscheidungsstrategie beschriebenen Entscheidungsstrategien in der Modellierung berücksichtigt werden. Aufgrund der theoretischen Annahme von verschiedenen Entscheidungsstrategien, sind neben dem Raschmodell weitere psychometrische Testmodelle denkbar (vgl. Kap. 3). Die Annahme des Raschmodells würde die Anwendung von Entscheidungsstrategien auf eine latente Personenfähigkeit zurückführen und Unterschiede im beobachteten Antwortverhalten quantitativ interpretieren. In diesem Falle würden sich die Entscheidungsstrategien als einfachere und schwierigere auf dem latenten Kontinuum darstellen. Eine Latent-Class-Analyse würde die beschriebenen Entscheidungsstrategien als unterschiedliche Qualitäten von Antwortverhalten und somit unterschiedliche Typen von Entscheidungsstrategien

interpretieren. Ein Mixed-Rasch-Modell würde innerhalb von Klassen von Entscheidungsstrategien unterschiedliche Kompetenzniveaus abbilden. Dies würde bedeuten, dass sich z.B. innerhalb einer Klasse von kompensatorischen Entscheidungsstrategien mehr und weniger kompetente Ausprägungen finden lassen. Schließlich könnte es auch der Fall sein, dass sich einige der identifizierten Klassen nicht skalieren lassen, sondern ein homogenes Antwortverhalten zeigen. In diesem Fall würde ein Hybridmodell das Antwortverhalten am besten erklären. Zusammenfassend lässt sich sagen, dass die erste Studie v.a. einen explorativen Charakter hat, indem untersucht werden soll, inwiefern sich die in der Entscheidungstheorie beschriebenen Entscheidungsstrategien bei Schülerinnen und Schülern beobachten lassen und inwiefern zentrale Faktoren wie Kontext und Darstellungsweise der Entscheidungssituationen einen Einfluss auf die Auswahl der Entscheidungsstrategie haben. Sollten sich mehrere Entscheidungsstrategien in den Schülerantworten bestätigen lassen, ist zu klären, welches psychometrische Modell dieses Antwortverhalten am besten erklären kann.

Auf Basis der fachdidaktischen und entscheidungspsychologischen Forschung generieren sich spezifische Forschungsfragen für die Studie 1 des Forschungsschwerpunkts II:

- Welche unterschiedlichen Entscheidungsstrategien lassen sich bei Schülerinnen und Schülern in der Bearbeitung von Entscheidungssituationen beobachten?
- Inwiefern beeinflussen Kontext und Darstellungsweise von Entscheidungssituationen die Auswahl einer Entscheidungsstrategie?

Daran schließt sich die Frage nach der psychometrischen Modellierung an:

- Welches IRT-Modell beschreibt das Antwortverhalten in Bezug auf die Anwendung von Entscheidungsstrategien am besten?

Studien 2 und 3

Die erste empirische Studie lieferte wichtige Ergebnisse im Hinblick auf eine Optimierung und Weiterentwicklung der Testaufgaben. Insbesondere die Auswahl des Kontextes bzw. der Darstellungsweise der Entscheidungssituationen wurde in der zweiten Studie optimiert. Analysen in Bezug auf den Einfluss des Kontextes zeigten, dass der Kontext in zwei identifizierten Subgruppen der Stichprobe einen Einfluss auf die Auswahl der Entscheidungsstrategie hatte. Als Konsequenz wurden in der zweiten Studie ausschließlich

Kontexte Nachhaltiger Entwicklung gewählt. Darüber hinaus wurde der Aufgabenpool um Aufgaben erweitert, in denen Schülerinnen und Schüler nicht mehr nur selbst Entscheidungen treffen, sondern darüber hinaus über die Strukturen von Entscheidungsprozessen reflektieren. Dazu wurden ihnen Entscheidungsprozesse von fiktiven Schülerinnen und Schülern vorgelegt, die auf Basis von prototypischen Schülerantworten aus der ersten Studie entwickelt wurden. Durch diese Testaufgaben konnte auch der dritte Aspekt der Teilkompetenz „Bewerten, Entscheiden und *Reflektieren*“ stärker in die Operationalisierung integriert werden. Aufgaben zur Reflexion über den eigenen Entscheidungsprozess, die bereits in der ersten Studie eingesetzt wurden, hatten sich als nicht hinreichend geeignet erwiesen, um Reflexionsprozesse anstoßen zu können.

Zudem wurde die Vorgehensweise in der Auswertung der offenen Antworten verändert. Dabei wurden die in der ersten Studie identifizierten Typen von Entscheidungsstrategien im Hinblick auf ihren Komplexitäts- und Elaborationsgrad hierarchisiert und in einem mehrkategorialen Auswertungsschema verarbeitet. Analog dazu wurden auch die Antwortkategorien im Hinblick auf die Reflexionsfähigkeit konstruiert. Die beschriebene Hierarchisierung der Entscheidungsstrategien in einem mehrkategorialen Auswertungsschema war eine Voraussetzung, die in der ersten Studie herausgefundenen Typen von Entscheidern auf einem latenten Kontinuum anordnen und somit mit Hilfe eines eindimensionalen polytomen Raschmodells modellieren zu können. Im Zentrum der zweiten Studie stand somit die Fragestellung, ob das favorisierte eindimensionale polytome Raschmodell das Antwortverhalten der Schülerinnen und Schüler adäquat abbilden kann. Zusammenfassend lässt sich sagen, dass die zweite Studie zwei große Schwerpunkte aufweist: zum einen der Versuch die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ eindimensional modellieren zu können und zum anderen das entwickelte Testinstrument im Hinblick auf seine Reliabilität und Validität zu optimieren. Da in der zweiten Studie v.a. Fragen nach der Reliabilität geklärt werden konnten, wurde eine dritte Studie durchgeführt, die Fragen nach der Validität des Testinstruments eingehender untersuchte. Aufgrund eines besseren Überblicks werden die zweite und die dritte Studie jedoch im Folgenden gemeinsam beschrieben.

Für die Studien 2 und 3 des Forschungsschwerpunkts II lassen sich die folgenden spezifischen Forschungsfragen formulieren:

- Kann die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ durch ein eindimensionales polytomes Raschmodell modelliert, und damit die Anwendung von Entscheidungsstrategien und deren Reflexion auf einem latenten Kontinuum dargestellt werden?
- Inwiefern genügt das entwickelte Testinstrument den Gütekriterien der Reliabilität und Validität?

5 Göttinger Modell der Bewertungskompetenz - Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung¹⁵

5.1 Zusammenfassung

Ein Schwerpunkt biologiedidaktischer Forschung ist derzeit die Entwicklung und empirische Überprüfung von Kompetenzmodellen. Dieser Trend ist ein Resultat internationaler Schulleistungsstudien sowie der derzeitigen Umsetzungsbestrebungen der nationalen Bildungsstandards für den mittleren Schulabschluss. Derzeit liegen den vier ausgewiesenen Kompetenzbereichen Sachwissen, Erkenntnisgewinnung, Bewertung und Kommunikation nur teilweise theoretisch fundierte und empirisch überprüfte Kompetenzmodelle zu Grunde. Dieser Beitrag leitet ein Strukturmodell für den Kompetenzbereich Bewertung mit besonderem Fokus auf den Kontext Nachhaltiger Entwicklung her, welches auf Erkenntnissen der Entscheidungstheorie, der Forschung zu naturwissenschaftlicher Grundbildung sowie Bewertungskompetenz basiert. Dabei wird detailliert erläutert, welche Konstruktionsprinzipien dem Modell zu Grunde liegen. Darüber hinaus werden eine mögliche Vorgehensweise zur empirischen Überprüfung des Kompetenzmodells sowie

15 Quelle: Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz – Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung. Zeitschrift für Didaktik der Naturwissenschaften, 12, 177-197.

Unser Dank gilt der Göttinger Arbeitsgruppe für die kritisch konstruktiven Diskussionen bei der Entwicklung des Kompetenzmodells. Diese Arbeit wird im Rahmen von Biologie im Kontext vom Bundesministerium für Bildung und Forschung (BMBF) gefördert.

postulierter Niveaustufen beispielhaft für die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ vorgestellt.

Abstract

One central aim of research in the area of didactics in biology is the development and empirical validation of competence models that can predict and describe pupils' competencies in central areas of biological education. This trend is due to the results of international assessment studies – such as TIMSS and PISA – as well as national German curriculum development in the area of science education. At the moment, only few theoretically based and empirically validated models concerning the formulated competence areas within the national curriculum exist. This article presents a theoretically based competence model for decision making in the area of sustainable development. The structure of the model is described in detail. A procedure for empirical validation is exemplified for one core aspect of decision making competence: “evaluation, choice and reflection”.

5.2 Einleitung

Viele Situationen, denen wir im Alltag begegnen, erfordern Bewertungen und Entscheidungen von uns. Biologieunterricht soll Schüler(innen) befähigen, begründete und reflektierte Entscheidungen zu treffen. Möchte man z.B. ein Haustier halten, so sollte man das für einen persönlich am besten geeignete Tier auswählen. Will man aus gesundheitlichen Gründen mehr Sport treiben, ist es ratsam in Frage kommende Sportarten einander gegenüber zu stellen und durch Abwägen eine gesundheitsfördernde und passende Sportart auszuwählen. Wohl überlegte Entscheidungen in gesellschaftlich relevanten Fragen angewandter Biologie, wie z.B. des Umweltschutzes oder der Nachhaltigen Entwicklung unseres Planeten, erfordern Entscheidungsfindungsprozesse (im Folgenden: Entscheidungsprozesse) und damit Bewertungskompetenz. Biologieunterricht soll Schüler(innen) in einem systematischen Umgang mit Entscheidungssituationen unterstützen. Mit der Formulierung der Bildungsstandards für den mittleren Schulabschluss in den naturwissenschaftlichen Fächern (KMK, 2004) und den Empfehlungen für die gymnasiale Oberstufe (Harms et al., 2004) wurde dies im Kompetenzbereich Bewertung für den Biologieunterricht curricular verankert bzw. angeraten.

In dem vorliegenden Artikel wird zunächst ein allgemein gültiger Prozess von Bewertungen und Entscheidungen auf der Basis bestehender Prozessmodelle und Entscheidungstheorien

rekonstruiert (Betsch & Haberstroh, 2005). Das Modell wird anschließend für den Kontext der Nachhaltigen Entwicklung konkretisiert und an einem Beispiel für den Biologieunterricht erläutert. Darauf aufbauend wird das Göttinger Strukturmodell für Bewertungskompetenz formuliert. Die Konstruktion des Modells greift dabei auf Prinzipien bestehender Kompetenzmodelle im Rahmen von *Scientific Literacy* bzw. *Science Education* (Bybee, 1997; Deutsches PISA-Konsortium, 2001/2004; Hammann, 2004; Wilson & Sloane, 2000) zurück.

5.3 Theoretischer Hintergrund: Prozessmodelle zu Bewerten und Entscheiden

Entscheidungsprozesse werden in der deskriptiven Entscheidungstheorie in der Regel durch Prozessmodelle, die sich aus verschiedenen Phasen zusammensetzen, beschrieben (für eine umfassende Zusammenstellung siehe Abelson & Levi, 1985, sowie Betsch et al., 2002). Die Abfolge dieser Phasen ist dabei nicht streng chronologisch zu verstehen; es können auch iterative Teilprozesse im Entscheidungsprozess auftreten. Viele Prozessmodelle für Entscheidungsfindungen lassen sich in einem dreiphasigen Metamodell (siehe Tab. 5.1) zusammenfassen (Betsch & Haberstroh, 2005).

Die *Präselektionale Phase* beginnt mit der Identifikation einer entscheidungsrelevanten Situation (Entscheidungssituation) (Betsch et al., 2002). Entscheidungsrelevant ist eine Situation, in der eine hinreichende Diskrepanz zwischen Ist- und Sollzustand wahrgenommen wird. Der Identifikation einer Entscheidungssituation schließt sich die Auswahl bzw. Generierung möglicher Entscheidungsoptionen an. Jede identifizierte Option wird dabei durch zentrale Kriterien (Jungermann et al., 1998) näher beschrieben¹⁶. Erforderlich wird damit eine gezielte Informationssuche. Die Intensität der Informationssuche, d.h. der in das Entscheidungsproblem investierte kognitive Aufwand, hat entscheidenden Einfluss (a) auf die Präselektionale Phase und (b) auf den gesamten Bewertungs- und Entscheidungsprozess (vgl. u.a. Svenson, 1990). Die Informationssuche mündet letztendlich in der Aufbereitung einer Entscheidungssituation.

In der *Selektionalen Phase* findet der Bewertungsprozess im engeren Sinne, d.h. das Vergleichen der identifizierten Optionen, statt. Beim Vergleich der Optionen in Bezug auf

¹⁶ In der Literatur wird für den Begriff „Option“ vielfach auch der Begriff „Alternative“ verwendet. Für den Begriff „Kriterium“ findet man oftmals den Begriff „Attribut“ (vgl. Jungermann et al., 1998).

zentrale Kriterien werden eine Vielzahl von Entscheidungsstrategien oder Heuristiken¹⁷ (Jungermann et al., 1998) verwendet. Die Anwendung von Entscheidungsstrategien kann helfen, einen komplexen Entscheidungsprozess zu systematisieren und zu vereinfachen, d.h. sie kann erfolgreiches Entscheiden unterstützen. Entscheidungsstrategien sind zahlreich (vgl. u.a. Jungermann et al., 1998, Payne et al., 1998, Hogarth, 1987, Abelson; Levi 1985); so können beispielsweise kompensatorische und nicht-kompensatorische Entscheidungsstrategien unterschieden werden (siehe Kasten 5.1). Bei der Anwendung von kompensatorischen Strategien ist ein Abwägen der verschiedenen Kriterien über die gegebenen Optionen hinweg möglich. Es findet ein *trade-off*¹⁸ der verschiedenen Kriterien statt. Bei der Anwendung von non-kompensatorischen Strategien kann ein „schlechtes“ Kriterium nicht durch ein „gutes“ Kriterium aufgewogen werden. Es findet kein *trade-off* statt. Auf Simon (1976) geht eine Unterteilung von Entscheidern in *Satisficer* und *Optimizer* zurück. *Satisficer* wenden v.a. non-kompensatorische Entscheidungsstrategien an und wählen eine zufrieden stellende Option aus. *Optimizer* hingegen versuchen, die bestmögliche Option auszuwählen. Diese Unterteilung macht deutlich, dass eine non-kompensatorische Vorgehensweise zu einer Entscheidung führen kann, in der wichtige Aspekte unbeachtet bleiben. Welche Regel letztendlich ausgewählt wird, hängt von zahlreichen Faktoren ab. Zu diesen Faktoren zählt die Komplexität des Entscheidungsproblems oder des Entscheidungskontextes (vgl. u.a. Beach & Mitchell, 1978). So werden bei einer geringen Anzahl von Optionen eher kompensatorische, bei einer hohen Anzahl von Optionen eher non-kompensatorische Regeln angewandt. Auch können Kombinationen von Entscheidungsstrategien auftreten (Jungermann et al., 1998). Den Abschluss der *Selektionalen Phase* bildet die Formulierung einer Handlungsintention zur Umsetzung der Entscheidung (Betsch et al. 2002, 457). In der *Postselektionalen Phase* dominieren v.a. volitionale Prozesse, die zur Umsetzung der formulierten Handlungsintention führen. Hierzu gehören z.B. Abschätzungen über die Realisierbarkeit der Handlungsintention, den optimalen Zeitpunkt bzw. die Zeitdauer einer Handlung (Betsch et al., 2002, 457), aber auch Überlegungen, die einen erfolgreichen Abschluss der Implementation unterstützen. Neben der eher bewussten Anwendung von kompensatorischen oder non-kompensatorischen Entscheidungsstrategien in der Selektionalen Phase werden Entscheidungen im täglichen

17 Der Begriff Heuristik stammt ursprünglich aus dem Griechischen und bedeutet „Finden“ bzw. „Auffinden“ (Wissenschaftlicher Rat der Dudenredaktion). Heuristiken beschreiben Regeln oder Strategien, die helfen können, ein Problem zu lösen oder zu einer Entscheidung zu gelangen (vgl. u.a. Polya, 1954).

18 Der Begriff *trade-off* bezeichnet den Prozess des Abwägens bei der Anwendung von kompensatorischen Entscheidungsregeln.

Leben auch oftmals intuitiv getroffen, wo sie auch zu einem guten Ergebnis führen. Haidt (2001) geht in seinem *social intuitionist model* davon aus, dass Menschen sich in einer Vielzahl von Entscheidungssituationen eher intuitiv entscheiden bzw. ihre intuitiv getroffene Entscheidung post-hoc rechtfertigen. Aus bildungswissenschaftlicher und bildungspolitischer Sicht ist der Anspruch an Bewertungskompetenz aber, Schüler(innen) zu systematischen und begründeten Entscheidungen in komplexen Situationen Nachhaltiger Entwicklung zu befähigen (Eggert & Höhle, 2005; Bögeholz, 2006). Elaborierte Entscheidungsstrategien (non-kompensatorisch und kompensatorisch s.o.) sind hierbei ein geeigneter Ansatz.

Tab. 5.1: Prozessmodell der Entscheidungsfindung und Konkretisierung für Gestaltungsaufgaben Nachhaltiger Entwicklung

<p>Phase</p>	<p>Metamodell Entscheidungsfindung (verändert nach Betsch & Haberstroh 2005)</p>	<p>Entscheidungsfindung – konkretisiert für Gestaltungsaufgaben Nachhaltiger Entwicklung (aufbauend auf Bögeholz & Barkmann 2003)</p>	<p>Aufgabenbeispiel Entscheidungsfindung - Gestaltung eines Fließgewässers</p>
<p>Prä-Selektional</p>	<p>Identifikation einer Entscheidungssituation und Generierung von Verhaltensalternativen durch:</p> <ul style="list-style-type: none"> • Informationssuche 	<p>Identifikation einer Gestaltungssituation und Generierung von Gestaltungsoptionen durch:</p> <ul style="list-style-type: none"> • Identifikation relevanter Entscheidungskriterien • Aufarbeitung der ökologischen, ökonomischen und sozialen Sachinformationen 	<p>Gestaltungssituation: Wie soll der Abschnitt eines Fließgewässers gestaltet werden? Generierung von Gestaltungsoptionen durch:</p> <ul style="list-style-type: none"> • Planung, Durchführung und Analyse von Erhebungen zum Istzustand / zu Istzuständen • Aufarbeitung aller relevanten ökologischen, ökonomischen und sozialen Sachinformationen für alle relevanten Optionen unter Einbezug vermuteter Auswirkungen durch deren Implementation • Ergebnis sind drei Gestaltungsoptionen: <ul style="list-style-type: none"> ○ Option A: Mäandrieren ○ Option B: Wasserspielplatz ○ Option C: Hobbyforelleenteiche
<p>Selektional</p>	<p>Bewertung und Entscheidung</p> <ul style="list-style-type: none"> • Anwendung von Entscheidungsstrategien • Evaluationsprozesse • Formulierung einer Handlungsintention 	<p>Bewertung und Entscheidung</p> <ul style="list-style-type: none"> • Vergleichen der Sachinformation unter Berücksichtigung ... <ul style="list-style-type: none"> • eigener Wünsche, der Wünsche anderer sowie • gemeinschaftlich und gesellschaftlich verhandelbarer Werte und Normen • unter Anwendung von Bewertungsstrukturwissen und Entscheidungsstrategien • Reflexion bzw. Evaluation des Bewertungsprozesses sowie der getroffenen Entscheidung • Formulierung einer Handlungsintention zur Umsetzung einer ausgewählten Option 	<p>Bewertung und Entscheidung</p> <ul style="list-style-type: none"> • Vergleichen der Sachinformation unter Berücksichtigung von ... <ul style="list-style-type: none"> • Wünschen: z.B. Erhöhung der Artenvielfalt und des Erholungscharakters • Werten: z.B. ökologischer, gesundheitlicher Wert • Normen: z.B. Leitbild Nachhaltiger Entwicklung • unter Anwendung von Bewertungsstrukturwissen und Entscheidungsstrategien • Mögliche Reflexionsfragen: <ul style="list-style-type: none"> • Sind die Wünsche bei der Bewertung angemessen berücksichtigt worden? • Sind die Wichtigkeiten der Kriterien angemessen einbezogen worden? • Ist die gewählte Option für alle zufrieden stellend? • Option A (bzw. B oder C) soll umgesetzt werden!
<p>Post-selektional</p>	<p>Implementation des gewählten Verhaltens</p>	<p>Implementation der ausgewählten Option</p>	<p>Implementation der Option A, B oder C</p>

Kasten 5.1: Beispiele von ausgewählten Entscheidungsstrategien (in Anlehnung an Jungermann et al., 1998, S. 123)

Kompensatorische Entscheidungsstrategien (Anwendung bei geringer Anzahl von Optionen)	Non-Kompensatorische Entscheidungsstrategien (Anwendung bei hoher Anzahl von Optionen)																				
Kontext Wohnungssuche: Ein Paar sucht eine geeignete Wohnung. Drei Wohnungen A, B und C sind in die engere Auswahl gekommen.																					
<p><u>MAU (Multi-Attribute Utility)-Regel:</u></p> <p>Drei zur Auswahl stehende Wohnungen A, B, C werden in Bezug auf drei zentrale Kriterien analysiert. Für das Paar ist die Ausstattung der Wohnung am wichtigsten, gefolgt von der Busanbindung (Infrastruktur) und der Wohnlage (Nähe zu Erholungsgebieten wie z.B. Park, Wald). Jede Option wird bezüglich jedes Kriteriums auf einer Skala von 1-10 beurteilt (Berechnung des Partialnutzenwerts). Abschließend wird der Gesamtnutzen für jede Option berechnet, indem pro Kriterium die Wichtigkeit mit den Partialnutzenwerten multipliziert wird. Die Wohnung mit dem höchsten Gesamtnutzenwert wird ausgewählt. In diesem Fall wäre dies Wohnung A.</p>	<p><u>Satisficing-Regel (SAT)</u></p> <p>Nimmt man an, dass die Wohnungen A, B und C nicht gleichzeitig zur Verfügung stehen, sondern erst nacheinander angeboten werden, so bietet sich die Anwendung der <i>Satisficing-Regel</i> an. Dabei wird diejenige Option ausgewählt, die zuerst als hinreichend zufrieden stellend empfunden wird bzw. einen bestimmten Schwellenwert überschritten hat.</p>																				
<table border="1" data-bbox="272 981 813 1370"> <thead> <tr> <th data-bbox="272 981 459 1115">Option \ Kriterium (Wichtigkeit)</th> <th data-bbox="459 981 579 1115">Wohnung A</th> <th data-bbox="579 981 699 1115">Wohnung B</th> <th data-bbox="699 981 813 1115">Wohnung C</th> </tr> </thead> <tbody> <tr> <td data-bbox="272 1115 459 1189">Ausstattung (0,5)</td> <td data-bbox="459 1115 579 1189">4 (2,0)</td> <td data-bbox="579 1115 699 1189">4 (2,0)</td> <td data-bbox="699 1115 813 1189">2 (1,0)</td> </tr> <tr> <td data-bbox="272 1189 459 1263">Infrastruktur (0,3)</td> <td data-bbox="459 1189 579 1263">3 (0,9)</td> <td data-bbox="579 1189 699 1263">3 (0,9)</td> <td data-bbox="699 1189 813 1263">4 (1,2)</td> </tr> <tr> <td data-bbox="272 1263 459 1337">Wohnlage (0,2)</td> <td data-bbox="459 1263 579 1337">6 (1,2)</td> <td data-bbox="579 1263 699 1337">3 (0,6)</td> <td data-bbox="699 1263 813 1337">1 (0,2)</td> </tr> <tr> <td data-bbox="272 1337 459 1370">Gesamtnutzen</td> <td data-bbox="459 1337 579 1370" style="text-align: center;">4,1</td> <td data-bbox="579 1337 699 1370" style="text-align: center;">3,5</td> <td data-bbox="699 1337 813 1370" style="text-align: center;">2,4</td> </tr> </tbody> </table>	Option \ Kriterium (Wichtigkeit)	Wohnung A	Wohnung B	Wohnung C	Ausstattung (0,5)	4 (2,0)	4 (2,0)	2 (1,0)	Infrastruktur (0,3)	3 (0,9)	3 (0,9)	4 (1,2)	Wohnlage (0,2)	6 (1,2)	3 (0,6)	1 (0,2)	Gesamtnutzen	4,1	3,5	2,4	<p><u>Lexikographische Regel (LEX)</u></p> <p>Zuerst werden die Kriterien in Bezug auf ihre Wichtigkeit hierarchisiert. Anschließend wird jede Option hinsichtlich des wichtigsten Kriteriums beurteilt. Die Option, die das wichtigste Kriterium am besten erfüllt, wird ausgewählt. Erfüllt keine Option dieses Kriterium hinreichend, so wird das zweitwichtigste Kriterium betrachtet usw. Wenn der Schwellenwert für jedes Kriterium beispielsweise bei > 4 liegt, würde nach der Betrachtung der Kriterien Ausstattung und Infrastruktur keine Option ausgewählt werden. Erst nach der Betrachtung des drittwichtigsten Kriteriums Wohnlage würde Wohnung A ausgewählt werden.</p>
Option \ Kriterium (Wichtigkeit)	Wohnung A	Wohnung B	Wohnung C																		
Ausstattung (0,5)	4 (2,0)	4 (2,0)	2 (1,0)																		
Infrastruktur (0,3)	3 (0,9)	3 (0,9)	4 (1,2)																		
Wohnlage (0,2)	6 (1,2)	3 (0,6)	1 (0,2)																		
Gesamtnutzen	4,1	3,5	2,4																		
<p><u>Equal Weights Regel (Spezialfall der MAU-Regel):</u></p> <p>Bei der Anwendung der <i>Equal Weights Regel</i> werden alle Kriterien gleich gewichtet. Auch in diesem Fall würde Wohnung A gewählt werden.</p> <p><u>Additive Differenzenregel:</u></p> <p>Diese Regel ist bei der Auswahl zwischen zwei Optionen sinnvoll. Würde man z.B. Wohnung A und B miteinander vergleichen, so kann man die Differenzen zwischen jedem Kriterium bezüglich der Optionen bestimmen. Abschließend wird die Gesamtdifferenz berechnet. Nach der Anwendung dieser Regel würde Wohnung A ausgewählt.</p>	<p><u>Elimination by Aspects Regel (EBA)</u></p> <p>Die EBA-Regel ähnelt in ihrer Struktur der LEX-Regel. Es wird ebenfalls eine Hierarchisierung der Kriterien vorgenommen aufgrund derer die Optionen betrachtet werden. Diejenige Option, die zuerst auf dem in der Hierarchie wichtigsten Kriterium am besten abschneidet, wird ausgewählt.</p> <p>Im Gegensatz zur LEX-Regel handelt es sich bei der EBA-Regel um eine stochastische Entscheidungsregel. Stochastische Entscheidungsregeln unterscheiden sich von deterministischen in der Annahme, dass in wiederholten Entscheidungen nicht immer wieder dieselbe Option (hier Wohnung A) ausgewählt wird. Das bedeutet, dass z.B. das Kriterium Ausstattung nicht zwangsläufig immer das wichtigste Kriterien sein muss, sondern nur mit einer bestimmten Wahrscheinlichkeit das wichtigste Kriterium ist.</p>																				

5.4 Bewerten und Entscheiden im Kontext Nachhaltiger

Entwicklung

Entscheidungssituationen im Kontext Nachhaltiger Entwicklung sind von ihrer Struktur her allgemeinen Entscheidungssituationen sehr ähnlich (siehe Tab. 5.1). Der Kontext stellt jedoch spezifische Anforderungen an ein *erfolgreiches* Bewerten und Entscheiden (siehe Kasten 5.2). In Fragen Nachhaltiger Entwicklung geht es um Maßnahmen zur Gestaltung bzw. Entwicklung z.B. eines Ökosystems, einer Region oder einer Gemeinde, d.h. derartige Entscheidungssituationen sind durch eine gestalterische Komponente charakterisiert. Aus diesem Grund wird statt von Entscheidungs- auch von Gestaltungssituationen gesprochen.

In der *Präselektionalen Phase* erfolgt die Identifikation einer Gestaltungsaufgabe. Eine mögliche Gestaltungsaufgabe ist, inwiefern ein bestehendes Fließgewässer - unter Berücksichtigung von Gesetzen wie der Europäischen Wasserrahmenrichtlinie - entwickelt bzw. verändert werden kann. Dazu muss zuerst ein tatsächlicher Handlungsbedarf, d.h. eine Gestaltungsrelevanz, erkannt werden. Anschließend werden durch Informationssuche und -verarbeitung Gestaltungsoptionen entwickelt und identifiziert. Sinnvolle, nachhaltige Optionen sind diejenigen, die nicht nur ökologische, sondern auch ökonomische und soziale Aspekte berücksichtigen und somit der geforderten Gesamtvernetzung (Retinität, siehe Tab. 1) Rechnung tragen. Für ein Fließgewässer, welches einen eher schlechten ökologischen Zustand aufweist und damit gestaltungsrelevant ist, sind beispielsweise folgende Gestaltungsoptionen denkbar:

- a) *Mäandrieren des Flussverlaufs* (siehe Tab. 5.1). Die Mäandrierung ist eine für Renaturierungsmaßnahmen typische Option. Die Umsetzung dieser Option wird v.a. ökologische Konsequenzen für das Fließgewässer haben: Verbessert wird sich die Gewässerstruktur und damit verbunden die Artenvielfalt sowie die biologische und chemische Gewässergüte. Zudem kommen bei dieser Option auch ökonomische und soziale Aspekte zum tragen. So wird eine Mäandrierung Folgen für das Fließgewässer als Naherholungsziel haben. Nicht zuletzt spielen die finanziellen Möglichkeiten zur Umsetzung dieser Option eine entscheidende Rolle.
- b) *Gestalten eines Wasserspielplatzes an einem dafür angelegten Altarm* (siehe Tab. 5.1). Durch das Anlegen eines Altarmes werden ebenfalls die Gewässerstruktur sowie die biologische und chemische Gewässergüte verbessert. Daneben hätte diese Option im Vergleich zur ersten einen höheren Freizeit- und Erholungswert. Die Kosten beziehen sich auf eine Realisierung sowie auf deren Pflege und Instandhaltung.

- c) *Anlegen von zwei Hobbyforellenteichen* (siehe Tab. 5.1). Das Anlegen von Hobbyforellenteichen ist im Kontext Nachhaltiger Entwicklung eine legitime Option. Hier steht der ökonomische Aspekt im Vordergrund, jedoch dürfen auch ökologische und soziale Aspekte nicht außer Acht gelassen werden.

Das Ergebnis der Präselektionalen Phase ist eine aufbereitete Gestaltungssituation, d.h. die möglichen Gestaltungsoptionen werden beschrieben. Diese Gestaltungsoptionen sind dabei nicht als beliebig zu verstehen. Unterschieden werden muss zwischen legitimen und im Hinblick auf die Situation irrelevanten Optionen. Ein Korrektiv kann dabei das Leitbild der Nachhaltigen Entwicklung sein.

In der *Selektionalen Phase* werden die vorhandenen Gestaltungsoptionen unter Anwendung von Entscheidungsstrategien miteinander verglichen.¹⁹ Für das Fließgewässerbeispiel würden in dieser Phase die Maßnahmen „Mäander“, „Wasserspielplatz“ und „Hobbyforellenteich“ anhand von konkreten Daten miteinander verglichen werden. Dabei hängen Bewertungsprozess und Auswahl der Gestaltungsoption von der bewertenden Person bzw. ihren Zielen und Präferenzen ab. Dies ist vergleichbar mit allgemeinen Entscheidungssituationen. Für den Kontext spezifisch ist eine Reflexion der eigenen Werte und Normen sowie der Werte und Normen anderer im Hinblick auf das Leitbild der Nachhaltigen Entwicklung. Gefordert ist somit ein hohes Maß an Perspektivübernahme.

Die *Postselektionale Phase* umfasst eine Umsetzung samt deren Vorbereitung. Im Hinblick auf das Fließgewässerbeispiel wären dies zum einen Fragen bezüglich des zeitlichen Rahmens oder möglicher Hindernisse einer konkreten Umsetzung. Zum anderen gehören dazu aber auch Überlegungen, die einen erfolgreichen Abschluss der geplanten Umsetzung unterstützen.

Das Beispiel der Gestaltung eines Fließgewässers macht deutlich, dass in allen Phasen des Entscheidungsprozesses auf gesellschaftliche, gemeinschaftliche sowie eigene Ziele und Werthaltungen rekurriert werden muss, um zu einer begründeten und reflektierten Lösung zu gelangen. Eine Nichtbeachtung normativ relevanter Schritte in Entscheidungsprozessen in der Selektionalen Phase oder eine Vermischung von naturwissenschaftlichen Fakten und Werten bzw. Normen kann zu Hindernissen oder Fehlern im Entscheidungsprozess führen. Eine Bewusstmachung der faktischen und der normativen Ebene ist somit ein zentrales Desiderat für einen kompetenten Umgang mit Entscheidungssituationen. Der Kontext

¹⁹ Somit ist auch in dieser Phase noch der Ausschluss einer nicht legitimen Version im Sinne der Nachhaltigkeit, wie bereits für die Präselektionale Phase beschrieben, möglich.

Nachhaltiger Entwicklung ist insofern geeignet, als das er zur Auseinandersetzung mit dieser doppelten – faktischen und ethischen – Komplexität (Bögeholz & Barkmann, 2003) auffordert.

Kasten 5.2: Das Leitbild der Nachhaltigen Entwicklung – Herausforderung für den Biologieunterricht

Nachhaltige Entwicklung ist für viele Staaten das normativ legitimierte Leitbild für das 21. Jahrhundert (Rio 1992). Auch die Bundesrepublik Deutschland hat sich diesem Leitbild verpflichtet. Eine Aufbereitung für den Bildungsbereich wurde in der BRD seit 1999 verfolgt (vgl. BLK 21 „Bildung für eine Nachhaltige Entwicklung“ siehe de Haan & Harenberg 1999, Ahlf-Christiani et al. 2003). Im Angloamerikanischen und Angelsächsischen Raum haben diese Konzepte im Rahmen von STSE Education (Science-Technology-Society and Environment; vgl. u.a. SEPUP 1995, Zeidler 2003) oder citizenship education (vgl. u.a. Ratcliffe 1997) bereits eine längere Tradition. Mit dem Jahr 2005 wurde die UN-Dekade Bildung für eine Nachhaltige Entwicklung eingeläutet. Nicht zuletzt ist Bildung für eine Nachhaltige Entwicklung in den Bildungsstandards für den mittleren Schulabschluss ausgewiesen (KMK 2004, S.17 -20).

Das Leitbild Nachhaltiger Entwicklung fordert eine permanente, lebenslange Auseinandersetzung um eine nachhaltige Gestaltung des gemeinsamen Lebensraums Erde – insbesondere auch mit Blick über das eigene Leben hinaus. Obschon vielfältige Vorstellungen und Definitionen über Nachhaltige Entwicklung bestehen (z.B. Eigner-Thiel & Bögeholz 2004), kann man sich auf drei Essentials Nachhaltiger Entwicklung einigen (Bögeholz 2000, aufbauend auf WCED 1987, SRU 1994):

- a) Gesamtvernetzung von Ökologie, Ökonomie und Sozialem („Retinität“)
- b) Inter- und intragenerationelle Gerechtigkeit
- c) (Grund-)Bedürfnisorientierung

Retinität bezeichnet die Gesamtvernetzung der bei einem Problem oder Bewertungsgegenstand relevanten ökologischen, ökonomischen und sozialen Systeme. Die Forderung nach **Gerechtigkeit** bezieht sich sowohl auf die jetzige Generation als auch auf die Ansprüche zukünftiger Generationen. Dabei wird immer von den betroffenen Menschen bzw. der betroffenen Generation aus argumentiert, damit die **Grundbedürfnisse** aller Beteiligten gewährleistet sind.

5.5 Bewertungskompetenz im Biologieunterricht

Der beschriebene Prozess der Entscheidungsfindung (siehe Tab. 5.1) wird in der Terminologie der Bildungsstandards für das Fach Biologie unter Bewertung subsummiert, obwohl Bewertung nach dem Metamodell (Betsch & Haberstroh, 2005) lediglich Teil der Selektionalen Phase ist. Im Folgenden wird, dem Ansatz der Bildungsstandards folgend, von Bewertungskompetenz gesprochen, obschon sich im englischen Sprachraum *decision making competence* etabliert hat (vgl. u.a. Ratcliffe, 1997; Ratcliffe & Grace, 2003).

Um Bewertungskompetenz im Biologieunterricht zu fördern, bedarf es im ersten Schritt eines theoretischen Modells, welches zentrale Teilkompetenzen von Bewertung beschreibt. Ein derartiges Kompetenzstrukturmodell liegt den Bildungsstandards zur Zeit noch nicht zu Grunde. Als Ausgangspunkt eignet sich das beschriebene Prozessmodell der Entscheidungsfindung nach Betsch und Haberstroh (2005).

Darüber hinaus sollen Kompetenzmodelle auch Aussagen über Entwicklungsverläufe von Kompetenzerwerb machen können. Mit der Forderung nach Entwicklungsmodellen ist die Frage nach einer Messbarkeit von Kompetenzen und einer Graduierung auf verschiedenen Kompetenzniveaus verbunden. Diese Forderungen sind auch das Resultat bisheriger Schulleistungstudien (TIMSS und PISA).

Im Folgenden sollen zentrale Aspekte von Kompetenzmodellen sowie Vorgehensweisen bei der Messung von Kompetenzen am Beispiel von

- a) *Scientific Literacy* (Bybee, 1997; Deutsches PISA-Konsortium, 2001, 2004),
- b) Erkenntnisgewinnung durch Experimentieren (Hamman, 2004) sowie
- c) einem Modell im Rahmen von *science education for public understanding* (vgl. u.a. SEPUP: Wilson & Sloane, 2000) verdeutlicht werden.

Darauf aufbauend wird ein Kompetenzstrukturmodell für Bewertung formuliert sowie abschließend eine mögliche Vorgehensweise zur Messung dieser Kompetenz sowie mögliche Niveaustufen beschrieben.

5.6 Elemente von Kompetenzmodellen

Beispiel Scientific Literacy

Scientific Literacy ist das Modell, welches in den letzten zehn Jahren die Entwicklung des naturwissenschaftlichen Unterrichts sowie die Messung von Schülerkompetenzen in diesem Bereich maßgeblich beeinflusst hat. Aufbauend auf dem Scientific Literacy Konstrukt von Bybee (Bybee, 1997) werden in den OECD/PISA Untersuchungen drei zentrale Aspekte von naturwissenschaftlicher Kompetenz unterschieden: naturwissenschaftliche Konzepte, naturwissenschaftliche Prozesse sowie relevante Situationen bzw. Kontexte, in denen konzeptuelle und prozedurale Fähigkeiten angewendet werden müssen (Deutsches PISA-Konsortium, 2001, 197).

In den PISA Untersuchungen im Jahr 2000 wurden zunächst vier zentrale (Teil-)Prozesse formuliert:

- a) Verständnis naturwissenschaftlichen Konzepte,
- b) Verständnis der Besonderheiten naturwissenschaftlicher Untersuchungen,
- c) ein Umgehen mit Evidenz und
- d) Kommunizieren naturwissenschaftlicher Beschreibungen oder Argumente
(Deutsches PISA-Konsortium, 2001, 199).

Diese Prozesse wurden in den PISA Untersuchungen im Jahr 2003 weitestgehend wieder aufgegriffen (Deutsches PISA-Konsortium, 2004). Um Aussagen über mögliche Entwicklungsverläufe machen zu können, wurden die genannten Prozesse auf Basis der theoretischen Annahmen von Bybee in Niveaustufen unterteilt (siehe beispielhaft Tab. 5.2). Für die empirischen Untersuchungen wurden Aufgaben entwickelt, die v.a. Fähigkeiten in den vier relevanten Prozessen sowie jede der postulierten Niveaustufen möglichst gut abbilden. Messtheoretisch ist dies eine Voraussetzung, Schüler(innen), die ein bestimmtes Set an Aufgaben lösen, einem entsprechenden Kompetenzniveau zuordnen zu können. Die empirischen Ergebnisse der PISA Untersuchungen zeigen eine relativ gute Passung mit den theoretisch formulierten Niveaus. Die Schwellen, die die einzelnen Kompetenzniveaus voneinander trennen, wurden jedoch rein empirisch festgelegt. Sie stellen somit eine artifizielle Trennung der einzelnen Niveaustufen dar. Tatsächlich lässt sich die Entwicklung naturwissenschaftlicher Kompetenz eher auf einem Kontinuum beschreiben. Die empirisch generierten Niveaus wurden anschließend mit den theoretischen Vorannahmen abgeglichen, wobei die theoretischen Annahmen erweitert wurden. Allgemein lässt sich die Entwicklung

von einem basalen Niveau, welches durch Alltagswissen geprägt ist, hin zu einem differenzierten, auf Konzepte und naturwissenschaftliche Modelle fokussierenden Verständnis von Naturwissenschaft, beschreiben (siehe Tab. 5.2: Verständnis naturwissenschaftlicher Konzepte). Ähnlich lässt sich ein Argumentieren aufgrund von Alltagswissen beim Umgang mit Evidenz identifizieren (Niveau 1). Umgang mit Evidenz zielt jedoch auf die Fähigkeit, naturwissenschaftliche Daten und Erkenntnisse in die eigene Argumentation einzubringen und als Beleg für Aussagen und mögliche Schlussfolgerungen heranzuziehen (siehe Tab. 5.2).

Tab. 5.2: Scientific Literacy am Beispiel von zwei Prozessen und deren Niveaustufen (vereinfacht nach: Deutsches PISA-Konsortium, 2001, S. 204)

Niveaustufen	Verständnis naturwissenschaftlicher Konzepte	Umgehen mit Evidenz
1 Nominell	Einfaches Faktenwissen wiedergeben	Schlussfolgerungen auf der Basis von naturwissenschaftlichem Alltagswissen ziehen oder bewerten
2 Funktional	Naturwissenschaftliches Alltagswissen anwenden	Schlussfolgerungen unter Verweis auf Daten oder naturwissenschaftliche Informationen ziehen oder bewerten
3 Funktional	Naturwissenschaftliche Konzepte anwenden	Beim Ziehen und Bewerten von Schlussfolgerungen zwischen relevanten und irrelevanten Daten unterscheiden oder Argumentationsketten auswählen
4 Konzeptuell und Prozedural	Elaborierte naturwissenschaftliche Konzepte anwenden	Daten systematisch auf Aussagen über mögliche Schlussfolgerungen beziehen und eine Argumentationskette entwickeln
5 Konzeptuell und Prozedural (Modelle)	Einfache konzeptuelle Modelle entwickeln oder anwenden	Daten als Evidenz benutzen, um alternative Gesichtspunkte oder unterschiedliche Perspektiven zu beurteilen

Zusammenfassend lässt sich die Struktur des *Scientific Literacy* Modells durch zentrale, voneinander abgrenzbare, Teilprozesse beschreiben, die sowohl konzeptuelle als auch prozedurale Fähigkeiten aufgreifen, die in naturwissenschaftlich relevanten Situationen bzw. Kontexten angewandt werden müssen. Die Niveaustufen der einzelnen Prozesse zeigen einen in Systematik und Elaboration steigenden sowie qualitativ unterschiedlichen Umgang in der Performanz der Teilkompetenzen. Die Übergänge zwischen den Niveaustufen sind als Schwellen zu verstehen, die v.a. empirisch festgesetzt werden. Zur Bestimmung der Fähigkeit von Schüler(inne)n werden für jedes Kompetenzniveau Aufgaben entwickelt. Dadurch können Personen, die ein bestimmtes Set von Aufgaben mit einer bestimmten Wahrscheinlichkeit lösen, einem Kompetenzniveau zugeordnet werden. Die Situationen und Kontexte, in denen die Aufgaben verankert sind, sollen den gesamten Bereich der Naturwissenschaften abdecken und sind damit sehr weit gefasst.

Beispiel Erkenntnisgewinnung durch Experimentieren

Das Kompetenzmodell zur Erkenntnisgewinnung durch Experimentieren (Hammann, 2004) baut auf dem Scientific Literacy Modell auf und erweitert es um domänenspezifisches Wissen sowie Prozesse, die für den Erwerb von Kompetenzen beim Experimentieren notwendig sind (vgl. u.a. Hammann, 2004). Hammann identifiziert nach dem Scientific Discovery as Dual Search Model (SDDS-Modell; Klahr, 2000) drei Teilkompetenzen, welche von ihrer Struktur mit den Prozessen im Scientific Literacy Modell zu vergleichen sind. Dies sind

- a) Suche im Hypothesenraum,
- b) Suche im Experimentierraum sowie
- c) Analyse von Daten (Hammann, 2004, 198).

Die Teilkompetenz Suche im Hypothesenraum beschreibt Fähigkeiten in der Generierung und Überprüfung von Hypothesen. Die Teilkompetenz Suche im Experimentierraum beschreibt die Fähigkeit, Experimente systematisch zu planen und durchzuführen. Das Analysieren der Daten verbindet die Fähigkeiten der beiden genannten Teilkompetenzen, indem Daten im Hinblick auf die Hypothesen und dahinter liegenden Theorien ausgewertet und u.a. mit den Versuchsbedingungen in Beziehung gesetzt werden.

Die drei Teilkompetenzen werden ebenfalls in vier Niveaustufen unterteilt. Exemplarisch wird die Kompetenzentwicklung für die Teilkompetenz Suche im Experimentierraum vorgestellt, die sich durch steigende Fähigkeit zur Systematisierung beschreiben lässt. Schüler(innen) auf Kompetenzniveau 1 können Variablen in einem Experiment nur unsystematisch variieren, d.h. sie verändern mehrere oder alle Variablen (Hammann, 2004, 201). Schülerantworten auf den höheren Niveaus zeichnen sich dadurch aus, dass Variablen kontrolliert variiert werden können. Auf dem höchsten Niveau kann diese systematische Vorgehensweise auf andere Domänen übertragen werden. Dem Modell liegt somit – wie auch dem Scientific Literacy Modell – eine steigende Systematik zu Grunde. Für die empirischen Untersuchungen wurden Multiple Choice Aufgaben für jede der drei Teilkompetenzen entwickelt (vgl. Hammann et al., 2006c; Phan & Hammann, 2006). Die einzelnen Kompetenzniveaus für jede Teilkompetenz werden, im Unterschied zu PISA, durch die verschiedenen Antwortmöglichkeiten abgebildet. Dadurch werden Schüler(innen), die eine bestimmte Antwortmöglichkeit auswählen, dem entsprechenden Kompetenzniveau zugeordnet.

Zusammenfassend lassen sich für beide Modelle, Scientific Literacy und Erkenntnisgewinnung durch Experimentieren, zentrale Entwicklungslinien beschreiben. Zum einen ist dies ein steigender Elaborationsgrad, mit dem naturwissenschaftliche

Fragestellungen bearbeitet werden und wie relevantes Wissen bzw. relevante Prozesse zur Lösung herangezogen werden. Zum anderen ist ein steigender Systematisierungsgrad bei naturwissenschaftlichen Untersuchungen erkennbar. Die genannten Entwicklungslinien werden bei der Entwicklung des Strukturmodells für Bewertungskompetenz berücksichtigt.

Beispiel Science Education for Public Understanding Program (SEPUP)

Im Rahmen von *Science Education for Public Understanding*²⁰ wird *decision making in socio-scientific-issues*²¹ als ein Prozess verstanden, in dem auf relevantes biologisches Fachwissen sowie Erkenntnisse naturwissenschaftlicher Untersuchungen rekurriert werden muss, um mögliche Lösungsalternativen im Hinblick auf eine Entscheidungssituation zu prüfen und deren Vor- und Nachteile abwägen zu können. Darüber hinaus müssen Schüler(innen) in der Lage sein, naturwissenschaftliche Untersuchungen selbst durchzuführen, zu analysieren und Ergebnisse adäquat zu kommunizieren (nach Roberts, Wilson & Draney, 1997, 4). Diese identifizierten Strukturen dienten als Basis für eine Definition von 5 SEPUP Variablen, die zentrale Teilkompetenzen von *decision making* beschreiben:

- a) “Designing and Conducting Investigations“
- b) “Evidence and Trade-Offs” bzw. “Evidence to Make Trade-Offs”
- c) “Understanding Concepts”
- d) “Communicating Scientific Information”
- e) “Group Interaction” (Roberts, Wilson & Draney, 1997, 4).

Jede Variable²² wurde dabei a priori auf unterschiedlichen Niveaustufen konkretisiert und anschließend empirisch überprüft sowie optimiert. Das Prinzip wird im Folgenden an der für Bewertungskompetenz zentralen Variable „*Using Evidence to Make Trade-offs*“ beschrieben (siehe Tab. 5.3). Diese Variable beschreibt die Fähigkeit, in einer Entscheidungs- bzw. Problemsituation Vor- und Nachteile der gegebenen Optionen vergleichen zu können. Beispielhaft wird dies an einer Aufgabe zum Thema Verpackungsmaterial illustriert (siehe Kasten 5.4). Die Schülerantworten auf die Verpackungsmaterialaufgabe werden anhand eines vierstufigen *Scoring Guides* analysiert (siehe Tab. 5.3). Auf Niveau 1 nennen und erläutern Schüler(innen) lediglich eine für das Problem relevante Perspektive bzw. Option. Zur Erläuterung ziehen sie v.a. persönliche

20 im Folgenden: SEPUP, siehe Kasten 3

21 siehe Kasten 5.3

22 Die Begriffe Prozesse, Teilkompetenzen sowie Variablen, die in den verschiedenen Modellen benutzt werden, sind inhaltlich auf derselben Ebene verortet.

Meinungen heran, begründen jedoch nicht auf der Basis relevanter Fakten. Auf Niveau 2 können Schüler(innen) bereits mehrere Perspektiven bzw. Optionen nennen und erläutern. Darüber hinaus argumentieren sie nicht mehr auf der Basis von Meinungen, sondern können einige relevante Fakten für ihre Begründung anführen. Es fehlen aber zentrale Aspekte, die für eine Lösung der Aufgabe von Bedeutung sind. Auf Niveau 3 werden alle relevanten Fakten berücksichtigt. Schüler(innen) können auf diesem Niveau die gegebenen Optionen bzw. Perspektiven vollständig miteinander vergleichen. Der Aspekt der ansteigenden Perspektivübernahmefähigkeit geht auf das Konzept der SOLO Taxonomie²³ zurück, die dem Modell zu Grunde liegt (Biggs & Collis, 1982). Niveau 4 schließlich geht über die Bearbeitung der Aufgabe insofern hinaus, als das nicht in der Aufgabenstellung genannte Aspekte zur Lösung herangezogen werden oder aber die Angemessenheit der vorhandenen Informationen hinterfragt wird. Schülerantworten auf diesem Niveau unterscheiden sich vom Niveau 3 somit v.a. durch die Fähigkeit zur kritischen Reflexion über die zu bearbeitende Aufgabe.

23 Die „Structure of Observed Learning Outcome“ (SOLO) Taxonomie beschreibt Kompetenzentwicklungsverläufe durch eine steigende Perspektivübernahmefähigkeit.

Kasten 5.3: Science Education for Public Understanding Program (SEPUP)

Das *Science Education for Public Understanding Project* (SEPUP) wurde 1997 in den USA in Zusammenarbeit zwischen der Universität Berkeley sowie Curriculumentwicklern und Lehrkräften ins Leben gerufen. Das Programm setzt sich aus mehreren Bestandteilen zusammen. Zum einen wurden verschiedene Curricula jeweils für ca. ein Schuljahr konzipiert, auf deren Basis Unterrichtskonzepte entwickelt wurden. Diese Konzepte wurden dann von Lehrkräften im Unterricht eingesetzt.

Für den Kompetenzbereich Bewertung ist dabei v.a. das Unterrichtskonzept „*Issues, Evidence and You*“ (SEPUP 1995) relevant. „*Issues, Evidence and You*“ möchte Schüler(innen) in ihrem Umgang mit Themen angewandter Biologie fördern, welche im Englischsprachigen Raum unter dem Begriff *socio-scientific-issues* gefasst werden. Zum anderen beinhaltet das Programm ein *Assessment System*, mit dem die Kompetenz bzw. die Kompetenzentwicklung von Schüler(inne)n erhoben werden kann. Dazu werden Diagnoseaufgaben eingesetzt, die auch während einer laufenden Unterrichtseinheit eingesetzt werden können.

Tab. 5.3: Scoring Guide für die SEPUP Variable “Using Evidence to Make Trade-offs”
(übersetzt, leicht modifiziert nach Wilson & Sloane, 2000, S. 193)

Niveau	Beschreibung
	Schüler(innen):
0	geben keine Antwort bzw. geben eine Antwort, die sich nicht auf die Fragestellung bezieht.
1	nennen eine relevante Perspektive bzw. Option, aber geben nur persönliche Meinungen wieder und/oder führen falsche oder irrelevante Belege an.
2	nennen mindestens zwei Perspektiven bzw. Optionen, begründen unter Bezug auf einige relevante Fakten, aber zentrale Aspekte fehlen.
3	nennen mindestens zwei Perspektiven bzw. Optionen, begründen unter Bezug auf alle relevanten Fakten und können zwischen den genannten Perspektiven bzw. Optionen abwägen.
4	erreichen Niveau 3 und gehen darüber hinaus, indem zusätzliche, über die vorhandenen Informationen hinaus gehende Belege angeführt werden und/oder indem die gegebenen Sachinformationen und/oder naturwissenschaftlichen Erkenntnisse kritisch auf ihre Aussagekraft hin hinterfragt werden.

Kasten 5.4: SEPUP Beispielaufgabe „Verpackungsmaterial“

Stell Dir vor, ...

Du arbeitest in einer Firma, die Geschirr produziert. Du bist verantwortlich für den Export der Waren. Die Waren werden zum größten Teil mit dem Schiff transportiert. Damit die Waren nicht beschädigt werden, musst Du ein geeignetes Verpackungsmaterial auswählen. Es stehen Dir zerkleinertes Zeitungspapier, Styroporkugeln oder Verpackungsmaterial aus Maisstärke zur Verfügung. Styroporkugeln behalten beständig ihre Form bei, Zeitungspapier und Maisstärkekugeln nicht. Styroporkugeln und Maisstärkekugeln schwimmen auf dem Wasser. Styroporkugeln können wieder verwendet werden, jedoch nicht biologisch abgebaut werden. Maisstärke löst sich in Wasser auf. Zeitungspapier kann einfach recycelt werden.

Welches Material würdest Du auswählen?
Vergleiche die Vor- und Nachteile und wäge dabei ab!

(übersetzt, leicht modifiziert nach Wilson & Sloane 2000, S. 190)

Die unterschiedliche Graduierung in der vorgestellten SEPUP Variable lässt sich analog zum Scientific Literacy Modell und dem Modell zur Erkenntnisgewinnung durch Experimentieren durch steigende Systematik und Elaboration, sowie durch steigende Perspektivübernahmefähigkeit charakterisieren. Hinzu kommt die Unterscheidung der Niveaus 3 und 4 durch kritische Reflexion des zu bearbeitenden Aufgabenmaterials.

5.7 Das Göttinger Modell der Bewertungskompetenz – konkretisiert für Gestaltungsaufgaben Nachhaltiger Entwicklung

Auf der Basis der relevanten Phasen in Entscheidungs- bzw. Gestaltungssituationen (siehe Tab. 1) lassen sich vier zentrale Teilkompetenzen für Bewertungskompetenz beschreiben. Dies sind:

- a) Generieren und Reflektieren von Sachinformationen
- b) Bewerten, Entscheiden und Reflektieren
- c) Kennen und Verstehen von Werten und Normen
- d) Kennen und Verstehen von Nachhaltiger Entwicklung

Die Teilkompetenz „Generieren und Reflektieren von Sachinformationen“ (siehe Abb. 5.1) greift zentrale Aspekte der Präselektionalen Phase, darunter Informationssuch- und

Verarbeitungsprozesse, auf. Die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ beschreibt zentrale Aspekte der Selektionalen Phase, darunter den Bewertungsprozess im engeren Sinne, d.h. das Vergleichen möglicher Optionen unter Anwendung von Entscheidungsstrategien. Beide Teilkompetenzen beschreiben v.a. relevante Prozesse von Bewertung. Zusätzlich werden zwei v.a. konzeptuelle Teilkompetenzen postuliert, die sich auf das bei Themen angewandter Biologie relevante domänenspezifische Wissen beziehen: „Kennen und Verstehen von Werten und Normen“ sowie „Kennen und Verstehen von Nachhaltiger Entwicklung“. Die Teilkompetenz „Kennen und Verstehen von Nachhaltiger Entwicklung“ ist für andere Kontexte, wie z.B. der Medizin- oder Bioethik, durch das in dieser Domäne relevante Wissen auszutauschen.

Eine weitere Teilkompetenz, die die Aspekte der Postselektionalen Phase (siehe Tab. 5.1) aufgreift, wurde für das Kompetenzmodell diskutiert, jedoch zunächst verworfen. Eine erfolgreiche Umsetzung einer formulierten Handlungsoption sowie ein Lernen aus bereits umgesetzten Handlungsintentionen ist ein zentraler Aspekt in Entscheidungssituationen. Eine derartige Umsetzung kann im schulischen Kontext jedoch oftmals nicht realisiert werden. So lassen sich Performanzen von Schüler(inne)n in Entscheidungssituationen im Rahmen des Biologieunterrichts oftmals nur bis zur Formulierung einer Handlungsintention nachvollziehen (vgl. Beispiel der Fließgewässergestaltung). Das Kompetenzmodell stellt einen Orientierungsrahmen für Diagnose und kognitive Förderung von Bewertungskompetenz dar.

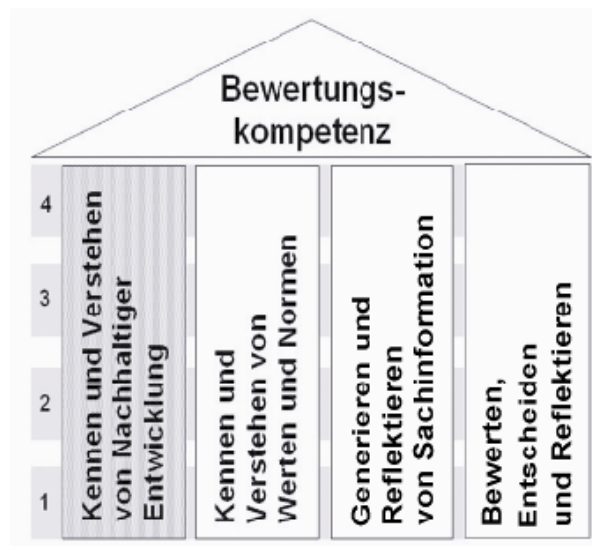


Abb. 5.1: Göttinger Modell der Bewertungskompetenz

Im Folgenden werden die vier Teilkompetenzen vorgestellt. Anschließend werden die Graduierungsprinzipien der Niveaustufen beispielhaft an der Teilkompetenz „Bewerten,

Entscheiden und Reflektieren“ veranschaulicht sowie eine geeignete Vorgehensweise zur Messung dieser Teilkompetenz erläutert.

Teilkompetenzen von Bewertungskompetenz

Die Teilkompetenz „**Generieren und Reflektieren von Sachinformationen**“ umfasst die Fähigkeit zur Identifikation möglicher Optionen zur Lösung einer Gestaltungsaufgabe. Schüler(innen) müssen u.a. in der Lage sein, Datenerhebungen bzw. Untersuchungen zu planen, durchzuführen und aufzubereiten. Das Ergebnis der Datenbearbeitungen ist die Beschreibung möglicher Optionen, d.h. ihrer ökologischen, ökonomischen und sozialen Aspekte. Um beispielsweise über die Gestaltung eines Fließgewässers begründet entscheiden zu können, müssen Schüler(innen) eine Gewässerkartierung planen, durchführen und die gewonnenen bzw. gegebenen Daten sinnvoll analysieren. Auf Basis des Istzustandes eines Gewässers müssen die Lernenden mögliche Gestaltungsoptionen, wie z.B. das Mäandrieren eines Flusslaufs, das Anlegen eines Wasserspielplatzes oder zweier Forellenteiche, beschreiben können (siehe Tab. 5.1). Wichtig ist hierbei auch das Abschätzen der Konsequenzen, die Umsetzungen derartiger Maßnahmen auf das Ökosystem Fließgewässer haben können. In der Umweltplanung bzw. dem Umweltschutz wird bei einer derartigen Vorgehensweise von Modellbildung gesprochen, welche den Anspruch hat, Gestaltungsoptionen möglichst genau und umfassend darzustellen (vgl. u.a. Poschmann et al., 1998). Unter Modell wird hierbei die Abbildung der Realität mit ihren Kernaspekten verstanden. Zu diesem Konzept von Modellbildung gehören auch das Erkennen der Defizite einer durchgeführten Datenerhebung sowie deren anschließende Analyse. Damit ist nicht nur die Reflexion über den eigenen Arbeitsprozess und das Erkennen möglicher Fehler gemeint. Schüler(innen) sollen vielmehr erkennen, dass jede Art von Datenerhebung und damit Beschreibung z.B. der ökologischen Zusammenhänge die Realität nicht vollständig abbilden kann, sondern immer nur eine Annäherung darstellt, welche auch kritisiert werden kann und muss (vgl. Bögeholz, 2006). Schüler(innen) werden hier in ihrer Fähigkeit zum Umgang mit unsicherem Wissen gefordert und sollen selbst bei defizitärer Datenlage fähig sein, Sachinformationen beurteilen zu können. Die Teilkompetenz „Generieren und Reflektieren von Sachinformationen“ beschreibt somit Fähigkeiten, die nicht nur im Rahmen von Bewertungskompetenz zentral sind, sondern Bedeutung für viele Bereiche des Biologieunterrichts haben.

Die Teilkompetenz „**Bewerten, Entscheiden und Reflektieren**“ greift zentrale Aspekte der Selektionalen Phase auf (siehe Tab. 5.1). In einer Gestaltungssituation müssen

Schüler(innen) Optionen miteinander vergleichen bzw. gegeneinander abwägen. Im Hinblick auf komplexe Gestaltungssituationen Nachhaltiger Entwicklung reicht ein intuitives oder rechtfertigendes (Haidt, 2001) Entscheidungsverhalten nicht aus. Um Optionen systematisch miteinander vergleichen zu können, ist das Kennen von Entscheidungsstrategien (siehe auch Kasten 5.2) zentral. Dabei sind Strategien wie das Anwenden von KO Kriterien durch Festlegung von Schwellenwerten (non-kompensatorische Vorgehensweise) oder das Vergleichen von Optionen durch Abwägen (kompensatorische Vorgehensweise) notwendig. Um diese Strategien in verschiedenen Kontexten bewusst auszuwählen und einsetzen zu können, ist Metastrategiewissen (Kuhn, 1999) bzw. ein Wissen über die Strukturen von Bewertungen (Bewertungsstrukturwissen, vgl. Bögeholz & Barkmann, 2003; Bögeholz, 2006) eine Voraussetzung. Wie schon für die Teilkompetenz „Generieren und Reflektieren von Sachinformationen“ formuliert, stellt das Ergebnis - die ausgewählte Option - nicht das Ende des Entscheidungsprozesses dar. Sowohl die Entscheidung als auch der Prozess mit allen normativen Teilentscheidungen muss von Schüler(inne)n reflektiert werden. Zu dieser kritischen Reflexion gehört neben eine Fehleranalyse der angewandten Strategien auch eine kritische Überprüfung der gefällten Wertentscheidungen.

Die Teilkompetenz „**Kennen und Verstehen von Werten und Normen**“ beschreibt zentrale Grundelemente ethischen Basiswissens, die für einen Umgang mit Gestaltungsaufgaben bei Themen angewandter Biologie zentral sind. Schüler(innen) müssen in der Lage sein, Werte und Normen von Meinungen, Aussagen oder Emotionen unterscheiden zu können. Erst, wenn sie faktische von ethischen Aussagen trennen können, ist eine Voraussetzung geschaffen, um normative Entscheidungen zu erkennen und zu reflektieren. Diese Unterscheidung scheint v.a. in bioethischen Kontexten auf den ersten Blick offensichtlich zu sein. In anderen Kontexten, wie z.B. der Nachhaltigen Entwicklung, können faktische und ethische Aussagen mangels noch weitgehend fehlenden Wissens um Nachhaltige Entwicklung in der Wahrnehmung leicht „verschwimmen“. Gerade in derartigen Kontexten sollten Schüler(innen) in der Lage sein, ethische Aspekte aufzudecken und bewusst Werte und Normen in Entscheidungen zu berücksichtigen (vgl. u.a. Eggert & Höhle, 2005). Über diesen bewussten Umgang mit Werten und Normen hinaus, sollen Schüler(innen) ein Verständnis über den gesellschaftlichen Aushandlungsprozess, in dem Normen entstehen, erlangen. Sie sollen erkennen, dass Normen und daraus resultierende Gesetze und Regelungen nur solange Bestand haben, wie sie durch die Gesellschaft legitimiert sind. Dazu gehört letztendlich, dass Schüler(innen) im Sinne eines postkonventionellen Verständnisses (vgl. u.a. Kohlberg, 1976) jenseits von Normen und Gesetzen denken und argumentieren und dadurch absolute Werte und Prinzipien in den

Vordergrund stellen können (vgl. transpersonal-autonomer Argumentationstyp nach Eckensberger et al., 1999, 183ff.).

Die Teilkompetenz „**Kennen und Verstehen von Nachhaltiger Entwicklung**“ beschreibt zentrale Aspekte des Konzepts der Nachhaltigen Entwicklung. Schüler(innen) sollen über eine Definition des Konzepts hinaus verstehen, dass eine Nachhaltige Entwicklung unserer Umwelt nur durch eine zusammenhängende Betrachtung der drei Sphären Ökologie, Ökonomie und Soziales, sowie der Berücksichtigung der Bedürfnisse der heutigen und zukünftiger Generationen möglich ist. Dazu gehört auch die Erkenntnis, dass diese Zusammenhänge aufgrund unterschiedlicher Zielvorstellungen zu Konflikten führen können. Das Erkennen derartiger Zielkonflikte wiederum ist die Grundlage zur Entwicklung von tragfähigen, nachhaltigen Lösungsoptionen.

Auf den ersten Blick könnte diese Trennung zwischen der allgemein formulierten Teilkompetenz „Kennen und Verstehen von Werten und Normen“ und der durch das Leitbild der Nachhaltigen Entwicklung konkretisierten Teilkompetenz als unnötig erscheinen, da auch bei der letzteren ein Verständnis von Werten und Normen erforderlich ist. Dennoch kann diese Trennung begründet sein, da der Kontext der Nachhaltigen Entwicklung andere Anforderungen an Schüler(innen) stellt als andere Kontexte angewandter Biologie. Es ist eine empirisch zu klärende Frage, inwiefern die Teilkompetenz „Kennen und Verstehen von Nachhaltiger Entwicklung“ domänenspezifisches Wissen neben allgemeinem ethischem Basiswissen erfordert.

Kompetenzniveaus am Beispiel von „Bewerten, Entscheiden und Reflektieren“

Graduierungen einer Kompetenz sind immer mit der Frage nach ihrer empirischen Überprüfbarkeit sowie Messbarkeit verbunden. Aus diesem Grund werden im Folgenden analog zu den Modellen von Scientific Literacy, zur Erkenntnisgewinnung durch Experimentieren sowie von SEPUP Niveaustufen von Bewertungskompetenz formuliert, die Entwicklungsverläufe beschreiben könnten. Hierbei wird auf Untersuchungen zur Bewertungskompetenz (Bögeholz & Barkmann, 2003; Große & Bögeholz, 2005) zurückgegriffen. Beispielhaft wird die Vorgehensweise für die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ vorgestellt.

Tab. 5.4: Kompetenzniveaus der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“

Niveau	Beschreibung
	Schüler(innen)...
1	<ul style="list-style-type: none"> • bewerten und entscheiden intuitiv bzw. rechtfertigend ohne Anwendung einer Entscheidungsstrategie • wählen eine Option auf der Basis von Alltagsvorstellungen aus und / oder berücksichtigen dabei maximal 1 Kriterium
2	<ul style="list-style-type: none"> • bewerten und entscheiden unter Berücksichtigung von mindestens zwei relevanten Kriterien • vergleichen gegebene Optionen teilweise im Hinblick auf die Kriterien und dokumentieren ihren Entscheidungsprozess unvollständig • entscheiden v.a. non-kompensatorisch
3	<ul style="list-style-type: none"> • bewerten und entscheiden unter Berücksichtigung von mindestens drei relevanten Kriterien • vergleichen gegebene Optionen vollständig im Hinblick auf die Kriterien und dokumentieren vollständig • entscheiden non-kompensatorisch und / oder kompensatorisch • reflektieren zentrale normative Entscheidungen im Bewertungsprozess
4	<ul style="list-style-type: none"> • bewerten und entscheiden unter Berücksichtigung von mindestens drei relevanten Kriterien • vergleichen gegebene Optionen vollständig im Hinblick auf die Kriterien und dokumentieren vollständig • entscheiden v.a. kompensatorisch • reflektieren zentrale normative Entscheidungen im Bewertungsprozess und können die Grenzen in der Anwendung von Entscheidungsstrategien erkennen

Bisherige Forschung zu Bewertungskompetenz konnte zeigen, dass Schüler(innen) Schwierigkeiten im Umgang mit einer systematischen Vorgehensweise in Entscheidungsprozessen haben und über kaum Metastrategie- bzw. Bewertungsstrukturwissen verfügen (Bögeholz & Barkmann, 2003; Große & Bögeholz, 2005; Müller, 2006). Ohne derartiges Wissen können Schüler(innen) in einer Entscheidungssituation nur wenig systematisch vorgehen. Sie entscheiden vielmehr intuitiv oder rechtfertigend. Diese Vorgehensweise entspricht dem postulierten Niveau 1 (siehe Tab. 5.4). Schüler(innen) wählen eine Option intuitiv und/oder unter Berücksichtigung eines Kriteriums aus und greifen dabei auf Alltagserfahrungen zurück (Müller, 2006). Die Charakterisierung des Niveaus 1 durch Argumentationen auf der Basis von Alltagsvorstellungen folgt den Prinzipien von Scientific Literacy, sowie dem SEPUP Modell. Diese eher unsystematische Vorgehensweise wird auch für das Niveau 1 für die Erkenntnisgewinnung durch Experimentieren beschrieben. Schüler(innen), die sich rechtfertigend entscheiden, verbalisieren ihre Entscheidung und rechtfertigen diese post-hoc durch die Anführung guter und die Ignorierung oder Abschwächung schlechter Eigenschaften. Schülerantworten auf den Kompetenzniveaus 2-4 sind durch ein systematischeres Vergleichen von Optionen, und damit eine Berücksichtigung mehrerer relevanter Kriterien, charakterisiert. Der steigende Systematisierungsgrad ist vergleichbar den Modellen von Scientific Literacy, zur Erkenntnisgewinnung durch Experimentieren und von SEPUP. Auf Niveau 2 berücksichtigen Schüler(innen) mindestens zwei Kriterien im Entscheidungsprozess und begründen die Auswahl bzw. Gewichtung dieser Kriterien durch

ihre Wünsche bzw. Präferenzen. Im Entscheidungsprozess werden v.a. non-kompensatorische Strategien angewandt. Die Dokumentation des Prozesses kann noch unvollständig sein, d.h. es werden einige Optionen und/oder Kriterien nicht berücksichtigt. Schüler(innen) auf Niveau 3 berücksichtigen mindestens drei Kriterien im Entscheidungsprozess und begründen diese mit ihren Wünschen bzw. Präferenzen. Eine Berücksichtigung von drei Kriterien bei mehreren Optionen erfordert im Vergleich zu Niveau 2 einen hohen Anstieg in der Komplexität und fordert von Schüler(inne)n ein systematischeres Vorgehen. Auf diesem Niveau wenden Schüler(innen) sowohl non-kompensatorische als kompensatorische Entscheidungsstrategien an. Sie sind in der Lage, die in Entscheidungsprozessen vorgenommenen normativen Teilentscheidungen zu reflektieren und mögliche Fehler zu korrigieren. Auf diesem Niveau werden Aufgaben von Schüler(inne)n analog zum Niveau 3 der entsprechenden SEPUP Variable vollständig bearbeitet. Für beide Modelle ist hierbei eine kompensatorische Vorgehensweise bzw. das Anwenden von *trade-offs* kennzeichnend. Auf Niveau 4 gehen Schüler(innen) über eine vollständige Bearbeitung der Aufgabe noch hinaus. Sie erkennen die Grenzen in der Anwendung von Entscheidungsstrategien. Niveau 3 und 4 unterscheiden sich daher auch durch eine stärkere Elaboration im Hinblick auf die Reflexionsfähigkeit.

Messung von Bewertungskompetenz

Bei der Messung von Bewertungskompetenz sowie der Erfassung von Niveaustufen sind verschiedene Vorgehensweisen möglich. Generell werden bei der Konstruktion von Messinstrumenten in den vorgestellten Modellen relevantes Wissen, relevante Prozesse sowie naturwissenschaftlich relevante Situationen bzw. Kontexte beachtet.

In den PISA Untersuchungen wurden Aufgaben zu Scientific Literacy konstruiert, die die verschiedenen Kompetenzniveaus jedes Prozesses abbilden. Zum einen können somit die Fähigkeiten für jeden relevanten Prozess getrennt analysiert werden. Zum anderen sind die Fähigkeiten von Personen mit dem Schwierigkeitsgrad der Aufgaben direkt vergleichbar. Das für eine Aufgabe relevante Faktenwissen wurde vorgegeben, um eine Performanz innerhalb eines Prozesses unabhängig vom Vorwissen zu gestalten. Einschränkend ist hier jedoch anzumerken, dass der Einfluss anderer Teilkompetenzen bei einer Bearbeitung nie vollkommen auszuschließen ist. Alle Aufgaben sind in einem der für PISA definierten Kontexte: Leben und Gesundheit, Erde und Umwelt sowie Technologie und Wissensgenese (vgl. Deutsches PISA Konsortium, 2004, 113) verortet. Für das Modell der Erkenntnisgewinnung durch Experimentieren wurden ebenfalls Aufgaben für jede

Teilkompetenz entwickelt. Die Niveaus wurden jedoch über die verschiedenen Antwortmöglichkeiten in den Multiple Choice Aufgaben abgebildet.

Für SEPUP Aufgaben werden als relevante Kontexte Themen angewandter Biologie verwendet. Die Aufgaben sind so konstruiert, dass der Schwerpunkt auf einer der fünf Variablen, d.h. Teilkompetenzen, liegt. Bei der Abbildung der Schwierigkeitsniveaus innerhalb einer Teilkompetenz wurde im Vergleich zu den beiden erstgenannten Modellen anders verfahren. Für jede Variable liegen offene Aufgaben für die gesamte Teilkompetenz und – im Gegensatz zu PISA – nicht für einzelne Kompetenzniveaus vor. Die Schülerantworten werden pro Variable mit Hilfe eines *Scoring Guides* kodiert und anschließend anhand des erreichten Scores einem Kompetenzniveau zugeordnet. D.h., dass in diesem Falle das unterschiedliche Antwortverhalten die Kompetenzniveaus definiert und nicht die durch die Personen gelösten Aufgaben. Das Ergebnis, Schüler(innen) anhand ihrer Fähigkeit – entweder anhand der gelösten Aufgaben oder anhand des gezeigten Elaborationsniveaus in den Antworten – einem Kompetenzniveau zuzuordnen, ist bei beiden Vorgehensweisen jedoch gegeben.





Für die Messung von Bewertungskompetenz ist es wichtig, die Anwendung bestimmter Strategien sowie den Entscheidungsprozess in Verbindung mit der Entscheidung abzubilden und nicht ausschließlich das letztendliche Ergebnis zu beurteilen. Offene Aufgaben geben Schüler(inne)n diese Möglichkeit mehr als geschlossene oder stark vorstrukturierte offene Aufgaben (vgl. auch Hammann, 2006a). Aus diesem Grund wurde eine dem SEPUP Programm ähnliche Vorgehensweise ausgewählt. Es wurden Aufgaben entwickelt, die die Fähigkeiten von Schüler(inne)n in Bezug auf die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ messen. Das für die Bearbeitung der Aufgaben relevante Faktenwissen ist gegeben, um den Einfluss des Vorwissens gering zu halten. In Zukunft werden Aufgabensets für die weiteren drei Teilkompetenzen entwickelt.

Im Folgenden wird die Konstruktion beispielhaft an der Aufgabe „Kaufentscheidung Apfelsorte“ konkretisiert (siehe Abb. 5.2) Der Kontext Nachhaltiger Entwicklung lässt sich durch eine Matrix aus Bedeutungsdimension und Inhaltsdimension beschreiben (siehe Tab. 5.5). Die Bedeutungsdimension beschreibt die Relevanz einer Aufgabe für den Lernenden. Es werden Aufgaben unterschieden, die einen *Fokus* auf eine unmittelbare Bedeutung für Schüler(innen) und ihr direktes Umfeld legen (interpersonale Bedeutung). Dazu gehören Aufgaben, die sich z.B. mit der Haltung von Haustieren oder dem Kauf von Lebensmitteln beschäftigen. Andere Aufgaben haben v.a. eine Bedeutung für die lokale bzw. regionale Gemeinschaft. Hier ist das Beispiel der Fließgewässergestaltung verortet. Aufgaben mit einer globalen Bedeutung sind z.B. Aufgaben zur Produktion und zum Handel von Konsumgütern globaler Herkunft oder zur Lösung der weltweiten

Überfischungsproblematik. Ausgehend von dem Fokus der Aufgabe können dann Bezüge zu anderen Ebenen der Bedeutungsdimension hergestellt werden. Die Inhaltsdimension greift die Essentials Nachhaltiger Entwicklung auf (siehe Kasten 2, vgl. auch Bögeholz, 2000). Diese Essentials berücksichtigen die ersten beiden Bereiche der relevanten PISA-Kontexte tägliches Leben und Gesundheit sowie Erde und Umwelt (siehe Tab. 5.5). Eine Verortung einer Aufgabe in allen Feldern der Matrix in gleichem Maße ist dabei nicht notwendig. Die Matrix erlaubt vielmehr eine gezielte Konzeption und Auswahl exemplarischer Aufgaben.

Um die Schülerantworten zu kodieren wurde ein spezifischer *Scoring Guide* entwickelt. Dieser *Scoring Guide* besteht aus für Entscheidungsprozesse zentralen Konstrukten. Anhand der Scores auf den zentralen Konstrukten können Schülerantworten den postulierten Kompetenzniveaus zugeordnet werden. Die Auswertung erfolgt wie in den PISA Untersuchungen und dem SEPUP Programm probabilistisch. Die Analysen sollen zum einen die Frage klären, ob sich wie in den SEPUP Auswertungen die Fähigkeiten von Schüler(inne)n auf einer latenten Dimension „Bewerten, Entscheiden und Reflektieren“ abbilden lassen oder ob unterschiedliche Typen von Antwortverhalten angenommen werden müssen. Zum anderen soll im Hinblick auf die Entwicklung der Teilkompetenz untersucht werden, ob sich das Antwortverhalten über Klassenstufen hinweg unterscheidet und sich eine Entwicklung auch tatsächlich empirisch beschreiben lässt.

Kaufentscheidung Apfelsorte
Du möchtest im Supermarkt ein Kilo Äpfel kaufen. Die folgenden Informationen aus einem Werbeprospekt stehen Dir zur Verfügung:

Meran	Borsdorfer	Mörker	Alkmene
			
- leicht säuerlich	- säuerlich	- leicht säuerlich und leicht süß	- süß
- Süd-Italien	- Göttinger Umland	- Bayern	- Neuseeland
- ökologischer Anbau	- konventioneller Anbau	- ökologischer Anbau	- konventioneller Anbau
- Preis: 2,49€ / kg	- Preis: 1,88€ / kg	- Preis: 3,09€ / kg	- Preis: 2,09€ / kg

Aufgabe:
Wäge die Kriterien aus der Tabelle gegeneinander ab und erkläre dabei genau, wie Du zu Deiner Entscheidung gekommen bist.

Abb. 5.2: Beispielaufgabe zur Überprüfung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“

Tab. 5.5: Kontextmatrix für eine systematische Aufgabenentwicklung zur Umsetzung von Bildung für eine Nachhaltige Entwicklung (verändert nach Hammann, 2006a, 86)

Inhalts- dimension →	Retinität			Gerechtigkeit		Grund- bedürfnisse
	Sphäre Ökologie	Sphäre Ökonomie	Sphäre Soziales	Intra- generationell	Inter- generationell	Bedürfnis- orientierung
Bedeutungs- dimension ↓						
<i>Interpersonal:</i> Freunde und Familie	Auswahl Haustier, Kaufentscheidung Apfel					
<i>Gemeinschaft:</i> lokal und regional	Fließgewässerbewertung					
<i>Gesellschaft:</i> national und global	Produktion und Handel mit Konsumgütern z.B. Produktion von Tiefkühlshrimps in Shrimp-Aquakulturen					

5.8 Zusammenfassung und Ausblick

In diesem Artikel wird ein Prozessmodell für Bewertung auf Basis relevanter Aspekte der Entscheidungstheorie formuliert. Darauf aufbauend wurde ein Kompetenzstrukturmodell von Bewertung hergeleitet, welches bestehende Kompetenzmodelle aus dem Bereich der Naturwissenschaftlichen Grundbildung, dem Bereich von *decision making competence* sowie bestehende Forschung zu Bewertungskompetenz berücksichtigt. Vorgestellt wird eine mögliche Graduierung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ in Kompetenzniveaus sowie die Vorgehensweise bei der empirischen Überprüfung.

Ziel der Forschungsarbeit ist es, die entwickelten Aufgaben für die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ zunächst in Querschnittserhebungen zu testen. Zentrale Fragestellungen der Auswertung sind, ob die entwickelten Aufgaben zur Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ tatsächlich eine latente Dimension abbilden oder ob verschiedene Typen von Schüler(inne)n und damit unterschiedliche Qualitäten der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ identifiziert werden können. Damit hängt unmittelbar auch die Frage nach der Validierung bzw. Optimierung der postulierten Niveaustufen zusammen. Diese Forschungsfragen sind gleichermaßen für die weiteren Teilkompetenzen relevant.

Für die Umsetzung des Kompetenzbereichs Bewertung im Biologieunterricht bietet das Modell einen Orientierungsrahmen, was Bewertungskompetenz im Kern und in der Struktur

ausmacht. Eine Klärung des Kompetenzbereichs Bewertung ist dabei - gerade auch wegen der vielfältigen nebeneinander existierenden alltagsweltlichen Verwendungen - notwendig. Das Kompetenzstrukturmodell stellt einen Orientierungsrahmen für eine systematische Aufgabenentwicklung für eine kumulative Förderung von Bewertungskompetenz dar.

Eine Auswahl der Aufgabenkontexte mit Hilfe der vorgestellten Kontextmatrix bietet einen Referenzrahmen nicht nur für Diagnose- sondern gerade auch für Lernaufgaben im Unterricht. So können Bewertungs- bzw. Gestaltungsaufgaben entwickelt werden, die systematisch einzelne Aspekte auf dem Weg zur Lösung von Gestaltungsaufgaben Nachhaltiger Entwicklung bearbeiten und Teilkompetenzen von Bewertungskompetenz fördern. So können Schüler(innen) an einen Umgang mit faktischer und ethischer Komplexität von Themen angewandter Biologie herangeführt werden.

6 Identifying students' decision making strategies in everyday life contexts and contexts of sustainable development²⁴

6.1 Abstract

This study analyzed students' decision making strategies in two everyday life contexts and in one context of sustainable development. Using an open item questionnaire, individuals were requested to compare possible options on the basis of given information and to reach an informed decision. Results based on probabilistic test models (item response theory) showed that the sample (n=291) could be divided into three groups that each exhibited a different decision making strategy. A first group mainly used "trade-offs" to compare given options. This group used "trade-offs" constantly across all contexts and task formats. A second group used "cut-offs" to eliminate given options that did not meet a certain level. The use of "cut-offs" was partly influenced by the decision making context. A third group focused on the decision criteria that were provided to describe the presented options. Individuals "weighed" criteria according to personal importance, but did not explicitly use any "trade-offs" or "cut-offs". Especially this group was influenced by the task format and the decision making context. The identification of these different decision making strategies can provide a valuable starting point both for the development of measuring instruments and for the integration of such strategies into the science classroom as a tool for working on decision making problems of applied science.

²⁴ Eggert, S.*, Schatz, J.*, & Bögeholz, S. (2008). Identifying students' decision making strategies in everyday life contexts and context of sustainable development. Manuscript submitted for publication. Diese Arbeit wird gefördert vom Bundesministerium für Bildung und Forschung.
*joint first authorship

6.2 Introduction

The need to foster students' competence in decision making has been highlighted by curriculum development as well as science education research worldwide (c.f. AAAS, 1991; National Research Council, 1995; Ratcliffe & Grace, 2003; Roberts, Wilson & Draney 1997; SEPUP, 1995; Siegel, 2006; Seethaler & Linn, 2004; Zeidler, 2003). In Germany, decision-making competence has only lately been recognized by the National Standards as one of four key competence areas in science education (KMK, 2004). Here, decision making competence is defined as the ability to make informed decisions in order to be able to participate in societal discussions. The implementation of National Educational Standards was also a consequence of Germany's mediocre results in the PISA studies. The formulation of National Standards initiated a major shift in the German educational system from an input orientated to an output orientated curriculum. This change can be compared to changes that science education curricula already underwent in other countries such as the USA (c.f. Scantlebury, Boone, Butler-Kahle & Fraser, 2001).

To test whether the implementation of National Standards will be successful, the institute for educational progress (IQB²⁵) will conduct national written assessments in 2012. For the natural sciences, the focus of assessment will be on conceptual scientific knowledge and scientific inquiry. These competencies represent those areas that have traditionally been in the focus of science education for a long time. In contrast, decision making competence and scientific communication will not or only be considered marginally. This fact is extremely dissatisfactory, because students' decision making competence - especially about problems of applied science - is acknowledged as one major aspect for preparing students to become informed citizens of our society (c.f. Ratcliffe & Grace, 2003). One reason for postponing the assessment of decision making competence is that it is difficult to assess with respect to quantitative assessment. In contrast to scientific knowledge, assessment of decision making competence needs to capture processes rather than conceptual knowledge. One important cognitive process with respect to decision making competence is students' ability to reason and to compare different possible solutions within a decision making situation to be able to reach an informed decision (Ratcliffe & Grace, 2003). As a starting point, this quantitative study presents one approach for the assessment of students' reasoning abilities. More specifically, it will focus on the strategies that students apply in order to make a choice between different possible solutions.

25 IQB = Institut zur Qualitätsentwicklung im Bildungswesen

6.3 Decision making competence and decision theory

Central to reasoning processes in decision making situations is the ability to use trade-offs. This aspect has been analyzed by several researchers in the field of science education (Kortland, 1996; Ratcliffe, 1997; Ratcliffe & Grace, 2003; Roberts et al., 1997; SEPUP, 1995; Wilson & Sloane, 2000). Besides, individuals also frequently use cut-offs in decision making situations to eliminate options that do not meet an individual's expectations on a specific decision criterion (c.f. Abelson & Levi, 1985+; Jungermann et al., 2004). These decision making strategies can also be accompanied by explicitly weighing decision criteria or trade-offs according to personal importance (ibid.). These reasoning processes are not only limited to science subjects but are crucial for all complex decision making situations that cannot be solved spontaneously or intuitively.

Concerning students' abilities to use trade-offs in a decision making situation, Ratcliffe (1997) analyzed students' group discussions about a controversial issue. She found that although students were able to state the advantages and disadvantages in a decision making situation, they did so in an unsystematic way. Students used only a few most dominant decision criteria. In particular, environmental or ethical considerations were used less frequently. Recent findings have contributed to the notion that students have difficulties in reasoning and systematically comparing different options (Hogan, 1999; Hong & Chang, 2004; Yang, 2004). Results from a year-long middle high-school course about controversial issues within the Science Education for Public Understanding Project (SEPUP) (Roberts et al., 1997; Wilson & Sloane, 2000) presented different results. Here, students were asked to respond to open-ended questions to make decisions about contemporary issues. Analysis focused on students' ability to make trade-offs on the basis of scientific evidence. Students who attended a SEPUP science course performed significantly better on the decision making tasks than students who attended a normal science course (Wilson & Sloane, 2000). Following students over a whole school term, performances developed with regard to the use of trade-offs and the ability to incorporate multiple perspectives into their decisions. However, a recent study on students' decision making about issues of sustainable development, which used teaching material as well as scoring procedure from the SEPUP project, found only limited improvement of students' performance (Siegel, 2006). Seethaler and Linn (2004) analyzed students' decision making competence on the issue of genetically modified food. Students were asked to respond to open-ended questions and to write "position papers" about which method of crop production they favoured. Results showed that students were able to trade-off advantages and disadvantages of either method and were able to present evidence not only for their chosen but also for options that were rejected.

However, students were not able to include weighing of trade-offs in their decision making processes.

Descriptive decision theory describes the ability to use trade-offs in decision making situations as a compensatory strategy (e.g. Abelson & Levi, 1985; Jungermann et al., 2004; Westenberg & Koele, 1994). Apart from using trade-offs, individuals also often use non-compensatory decision-making strategies that are characterized by the use of cut-offs (ibid.). Options that do not meet a certain level on a specific criterion are eliminated and not considered any further. As a result, a remaining option that has fulfilled all established cut-offs will be chosen. It is also frequently observed that individuals use a combination of non-compensatory and compensatory strategies (Svenson, 2003). Considering the aspect of weighing, research indicates that individuals often do not explicitly prioritize given criteria according to importance but weigh all criteria with equal importance (Gigerenzer et al., 1999). Moreover, the decision-making strategy that an individual applies in a specific situation depends on the complexity of the task (e.g. Jungermann et al., 2004), the decision context and the way the task is structured (Rettinger & Hastie, 2003). In addition, emotions (Pfister & Böhm, 1992) as well as past experience in decision making situations (Betsch & Haberstroh, 2005) also play a role.

In view of the above, three elements emerge as being central to the analysis of cognitive decision making strategies. Firstly, it is the use of trade-offs and secondly, the use of cut-offs as characteristics of compensatory and non-compensatory decision making strategies. Thirdly, it is the aspect of weighing criteria or options according to the individual's own personal importance and thus to be explicit about individual value considerations. More specifically, the first aim was to analyze in an exploratory approach which patterns of decision making strategies are used by high school students as well as university undergraduates. Secondly, we were interested in whether the use of decision making strategies can be regarded as a more general argumentative skill or whether the application of a specific decision making strategy is strongly influenced by the decision making context as well as the format of the decision making task.

6.4 Methods

Sample

We developed a questionnaire to analyze the use of decision making strategies in three different decision making contexts covering everyday life contexts as well as one context of sustainable development. The questionnaire was used in a cross sectional study (n=256²⁶) with lower and upper secondary students (5 to 12 years of education) as well as undergraduate university students (n=35) in biology education. All secondary students attended grammar or comprehensive schools in Germany. At least two classes were tested at each grade. The questionnaire was completed by all individuals within 45 minutes²⁷. In order to cover a high range of performance we included students as well as undergraduates (n=99) who had prior experience with decision making tasks and the use of decision-making strategies.

Development of test tasks²⁸

Since the aim was to analyze the decision making strategies that students used and the process that they exhibited in their written reports, an open item format was used. We varied the decision making context as well as the format of the decision making tasks to analyze possible differences in the application of decision making strategies. Three decision making tasks were developed: a decision for a class strip, a decision for a sporting activity (c.f. Appendix A) and a decision for the sustainable development of a river ecosystem (c.f. Appendix B). All tasks comprised four different options that were characterized by four or five provided criteria to keep the amount of factual information relatively comparable. All presented options were equally legitimate. In the sporting activity and the river development situations, an additional question was included into the task (Appendix A and B; test tasks a) where individuals were explicitly asked to weigh given criteria according to their own personal importance. On the basis of this ranking, they were then asked to compare the presented options in as much detail as possible and finally to reach an informed decision (c.f.

26 In total, 269 individuals completed the questionnaire. However, 13 individuals exhibited extreme z-scores with respect to person fit. Thus, they were excluded from final data analysis (see results).

27 Die Testzeit von 45 Minuten gibt die Zeit an, die die Probandinnen und Probanden für die Bearbeitung der Aufgaben benötigten, die letztendlich in der Auswertung im Rahmen dieses Artikels berücksichtigt wurden. Die Bearbeitung aller in der ersten Studie eingesetzten Testaufgaben dauerte 90 Minuten.

28 Der komplette Fragebogen kann bei der Erstautorin angefragt werden. Eine gekürzte Version des Fragebogens ist in Appendix A / B (S. 83-84) gegeben.

Appendix A and B; test tasks b). The class trip situation did not have this additional question. This task also presented four options, which were characterized by five criteria. Here, students had to choose from two city trips and two trips to the countryside. Concerning the river development task, the presented information was reduced in a way that also junior high school students could deal with this task. Since all individuals received the same test tasks and performance was analyzed through the quality of individuals' responses, we aimed to find contexts that were comparably relevant for junior high school students as well as first year undergraduates.

Development and Application of Scoring Guide

To analyze individuals' responses, a scoring guide was developed based on decision making strategies described in literature. The first version of the scoring guide consisted of ten items. All responses²⁹ were analyzed with respect to these ten items for each decision making task. Thus, each individual received 30 scores in total. Scoring was completed by two researchers who coded all questionnaires independently. In the case of disagreement, discussion took place until an inter-coder agreement could be reached. As a next step, all scores were entered into SPSS (SPSS Inc., 2003) in order to analyze whether the ten identified items were logically independent from each other, which was a prerequisite for the subsequent data analysis. More specifically, this meant that every score combination had to be possible. Results showed that some of the items were logically dependent and as a consequence, had to be merged. Finally, three core items were identified, where any combination of these items is possible: weighing criteria, using trade-offs and using cut-offs. Tab. 6.1 presents the final scoring guide. The scoring guide was used for scoring all three decision making tasks. For example, a student who took all given criteria into account equally, scored zero on the item weighing criteria. A student who eliminated a given option because it was too expensive and who consequently did not consider it any further, scored 1 with respect to the item using cut-offs. A student who first eliminated two given options by using a cut-off and secondly traded-off the remaining two options by examining positive and negative aspects scored 1 on both items. Note that the dichotomous scoring procedure was also a result of preliminary data analysis. We started with polytomous scoring for each item at first. However, results showed that some scoring categories were either assigned seldom or scoring categories were not in order with regard to difficulty. A common procedure to deal with this problem is to collapse categories together (Bond & Fox, 2001). Individuals' final scores for every decision making task and every item (3 items per decision making task = 9

²⁹ In the current study, individuals' performances were not measured by the number of items that were solved, but by the quality of their responses with respect to the items in the scoring guide.

scores) were then analyzed with the WINMIRA software program (von Davier, 2001) using probabilistic test theory (Rost, 2004a).

Tab. 6.1: Final Scoring Guide used to analyze individuals' responses to the three decision making tasks

Item	Students and university undergraduates...	Assigned Score
Weighing criteria in decision process	do not weigh given criteria.	0
	weigh given criteria (in test task a) but do not refer to this weighing in their answer (test task b).	0
	weigh given criteria (test task a) and refer to this weighing in their answer (test task b).	1
Using cut-offs in decision process	do not use cut-offs during the decision process.	0
	eliminate given options by using cut-offs during the decision process.	1
Using trade-offs in decision process	do not use trade-offs to compare given options.	0
	only use trade-offs when comparing positive and negative features of the option that is chosen.	0
	use trade-offs to compare two or more given options in the decision process.	1

Data Analysis

The data was analysed with four different item-response-theory models. Most of these models are frequently used in the area of competence assessment. The most prominent one is the Rasch model (Rasch, 1960). Besides three other IRT-models were used, namely latent class analysis, mixed Rasch models, and hybrid models. These four models as well as their accompanying statistic are described briefly. The Rasch model (Rasch, 1960) is a psychometric test model that serves to explain an individual's performance on a particular test or test item through one specific trait. This specific trait is called a latent trait (e.g. Rost, 2004a; Boone & Scantlebury, 2005). In the current study, this was identified as students' ability to apply decision making strategies in a decision making situation. The Rasch model assumes that for an observed sample, any individual's performance can be explained by this latent trait. Individuals do not respond in qualitatively different ways, but only quantitatively different with respect to given test items. In terms of competence assessment this often means that test persons can be differentiated according to "how well" items are solved.

In contrast to the Rasch model, latent class analysis (LCA; Lazarsfeld & Henry, 1968) does not assume different qualities of response patterns in an observed sample. These are called latent classes. Individuals in the same class are characterized by using the same response pattern and are thus defined as a homogeneous subgroup. Individuals in different classes show maximal heterogeneous response behaviour. It is also possible to explain empirical data with statistical models that combine aspects of the Rasch model and latent class analysis. These are called mixed Rasch (Rost, 1990) and hybrid models (Yamamoto, 1989). Mixed Rasch models assume that an observed sample can be divided into different latent classes that show qualitatively different response patterns. Besides, response patterns in these classes show a quantitative variation. Hybrid models represent a mixture of the mixed Rasch model and latent class analysis. For example, an observed sample can be divided into classes, where in class one the Rasch model holds and in class two a latent class structure is assumed. Such hybrid models allow for free mixtures of Rasch models and latent class models (von Davier, 1997). It is also possible that more than one model can explain the data satisfyingly. In this regard, probabilistic test theory refrains from defining one model as “correct” and another as “false”. Instead, it identifies the model which serves to explain the data relatively best (Rost, 2004b).

Used statistics

The data was first analyzed with respect to global model fit. Statistical values for identifying the best model are the so called information criteria, namely the Best-Information-Criterion (BIC) and the Consistent Akaike's Information Criterion (CAIC). If the same data is analyzed with different competing models, the model that exhibits the smallest BIC and CAIC values is considered to fit best (Rost, 2004a).

Secondly, the bootstrapping method (von Davier, 1997) was used. This simulation technique is a test for significance. It has to be applied when the data is defined as being relatively sparse. This means that the number of observed response patterns was much smaller than the number of possible response patterns. “The bootstrap method is a resimulation technique where a number of data sets are simulated on the basis of the real data set. A common decision rule is to accept the model, where the fit statistic of the real data set is smaller - and thus better - than at least 5% of the simulated values of that fit statistic” (von Davier & Rost, 1997, p. 302). In this sense, the p-values represent the level of significance under which the null-hypothesis of model fit can be retained. Empirical p-values for both Cressie Read and Pearson from the bootstrap procedure were compared across the different models. If it was found that more than one model fitted the data with respect to the 5% significance level, “then the most parsimonious model was preferred” (von Davier & Rost, 1997, p. 303).

Thirdly, individuals were analyzed with regard to class membership. Every individual was assigned to one of the identified classes according to their maximum probability value. After that, the mean probability of class membership within the three classes was calculated. In comparison with classical test theory, mean probability values for class membership can be interpreted as a test for reliability (Rost, 2004b). Item profiles within each class were analyzed with regard to the model that fitted the data best to be able to explain response behaviors shown within the different classes in detail.

Finally, item fit as well as person fit was analyzed. Item fit was calculated using the Q-index. Q-index values for each item allow conclusions about how well an item fits the theoretical construct that is aimed to be measured. Thus, the Q-index is a procedure for analyzing reliability and can be compared to the KR-20 or Cronbach's alpha in classical test theory. Acceptable values for the Q-index range between 0.00-0.30 (Rost, 2004a). A Q-index of 0.50 usually indicates random response behaviour. Additionally, person fit was analyzed to identify individuals who showed unexpected response behaviour. These individuals can have a great impact on the identification of a specific test model and thus, should be considered carefully.

6.5 Results and Discussion

256 individuals were analyzed with respect to the models described above. The information provided in Tab. 6.2 shows that the Rasch model had higher BIC and CAIC values when compared to all other models. This suggests that the analyzed sample is not homogenous with respect to the use of decision making strategies. Secondly, two, three, four and five latent class models were applied to the data. BIC and CAIC values indicated that either the three or the four class solution should be considered further. As a next step, we analyzed three and four class mixed Rasch models to see whether a model that also assumes quantitative variations within the identified classes can explain our data better. Values for BIC and CAIC indicated that a three class mixed Rasch model fitted the data much better. However, the best model according to the information criteria is a three class hybrid model made up of two Rasch classes and one latent class (2RM1LCA; Tab. 6.2). Results from the bootstrap procedure strengthened the analysis from the information criteria. The four latent class solution, the three class mixed Rasch and the three class hybrid model all fulfilled the 5% significance level for p-values for Pearson. With regard to p-values for Cressie Read, only the three class hybrid model fulfilled the 5% significance level. Thus, this model can be considered as the one that has relatively best model fit. However, the four latent class

solution and the three class mixed Rasch model only just failed to meet that significance level. Further analysis of these models showed that they can be interpreted in a similar way. The four latent class solution can be seen as an extension of a three class model since one class is split into two classes that merely differ in their degree of performance. Comparing the three class mixed Rasch and the three class hybrid model, a cross tabulation showed that all individuals are assigned to the same classes in both models. Thus, both models explain individuals' decision making behavior in the same way only that the hybrid model does assume one latent class whereas the mixed Rasch model assumes only Rasch classes. Since the three class hybrid model showed best statistical test results, further analysis will concentrate on this model. The model was analyzed with regard to class membership (Tab. 6.3), item (Tab. 6.4) and person fit, as well as class profiles (fig. 6.1).

Tab. 6.2: Fit statistics for all applied models

model	log-likelihood	BIC	CAIC	no. of parameters	p-values*	
					Cressie Read	Pearson
Rasch model (RM)	-1552	3161	3171	10	0.000	0.000
LCA 2	-1498	3104	3123	19	0.000	0.014
LCA 3	-1461	3088	3117	29	0.000	0.044
LCA 4	-1430	3082	3121	39	0.044	0.180
LCA 5	-1419	3117	3166	49	0.020	0.142
mixed Rasch 3	-1438	3058	3090	32	0.046	0.306
mixed Rasch 4	-1418	3080	3123	43	0.010	0.094
hybrid (2RM1LCA)	-1437	3050	3081	31	0.050	0.284
hybrid (1RM2LCA)	-1445	3061	3091	30	0.026	0.126

*p-values were obtained by bootstrap procedure. 500 bootstraps were calculated.

Mean probabilities for class membership were high for all three classes (Tab. 6.3). Class 1 represents a homogeneous subpopulation (LCA).

Tab. 6.3: Mean probabilities of class membership for all three classes in the hybrid model

Class	size [%]	mean probability
1 (LCA)	44.4	0.881
2 (RM)	29.1	0.848
3 (RM)	26.5	0.855

Item fit was analyzed for all nine items for the two Rasch classes separately. All indices were in the range of 0.00-0.30, except for the item trade-offs2 (sporting activity) in class 2 (Tab. 6.4). The Q-index value of 0.50 for this item is problematic, but can be ascribed to the fact that all individuals in this class scored zero on this item. A 100% probability of scoring zero might have caused deterministic results with regard to this item. Zq values for all other items showed that they do not deviate significantly from the applied model. The Q-index for the item trade-off1 in class 3 can still be considered as acceptable (Bühner, 2006). Additionally, person fit was analyzed for every individual. Z-scores revealed that 13 individuals showed extreme response behaviour with regard to the items measured. Consequently, they were excluded from final analysis (see sample).

Tab. 6.4: Q-indices for both Rasch classes

task	Class 2 ("mainly weighing")			Class 3 ("compensatory")		
	Items	Q-index	Zq	Items	Q-index	Zq
Class trip	weighing1	0.140	-0.04	weighing1	0.239	1.13
	trade-off1	0.110	-0.72	trade-off1	0.312	1.14
	cut-off1	0.086	-0.26	cut-off1	0.121	-0.36
Sporting activity	weighing2	0.096	0.29	weighing2	0.082	-0.68
	trade-off2	0.500	-0.02	trade-off2	0.117	0.16
	cut-off2	0.108	-0.63	cut-off2	0.113	-0.33
River development	weighing3	0.153	0.67	weighing3	0.096	-0.51
	trade-off3	0.000	-0.75	trade-off3	0.112	-0.28
	cut-off3	0.206	-0.30	cut-off3	0.080	-0.72

Note: Zq values should not deviate significantly from 0; ($|Zq| < 1.96$)

Class profiles generated by plotting (mean) item response probabilities for all items across all decision making tasks in the three class hybrid model are depicted in fig. 6.1. Each class profile represents the (mean) item response probabilities for individuals within that specific class.

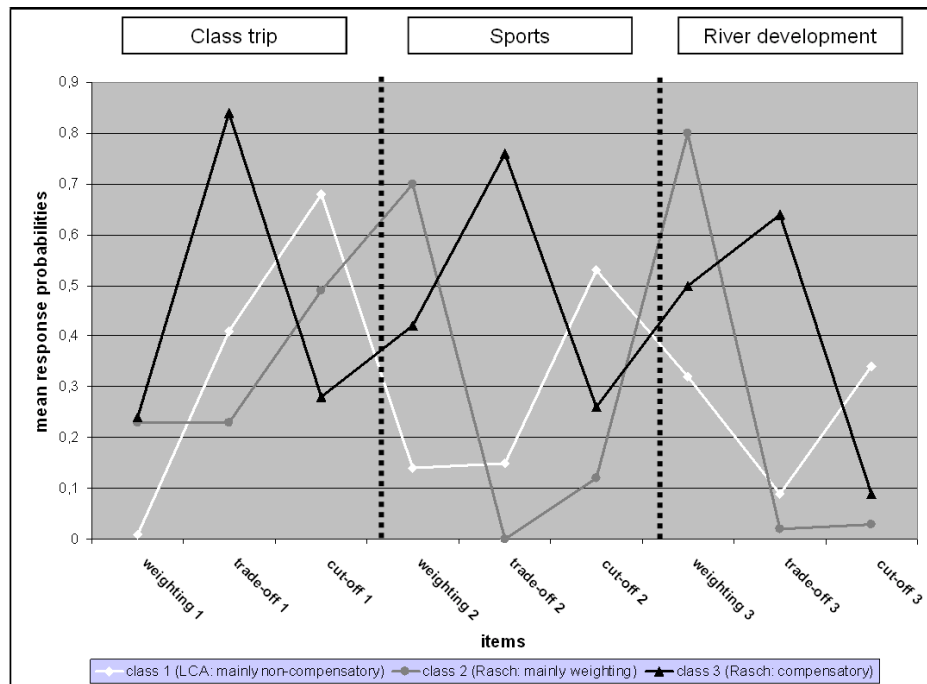


Fig. 6.1: Mean item response probabilities for all items across all three decision making tasks

Class 1 (LCA: “mainly non-compensatory”)

Class 1 (44.4 %) represented the largest subpopulation in the analyzed sample. In comparison with the other two classes, this class was characterized by the highest item response probabilities for all cut-off items across all decision making situations (0.68-0.34; Fig. 6.1). Performance on these items was highest in the class trip task and lowest in the river development task. Response probabilities for the item weighing were low across all tasks. While response behaviour was dominated by using cut-offs in the sporting activity and in the river development task, it was slightly different in the class trip situation. In this situation, individuals used a combination of cut-offs and trade-offs to reach a decision. Typically, individuals in this class eliminated two of the four given options by using a cut-off and then traded-off the remaining two options. The application of this mixed strategy is likely to be a result of the structure of the decision making task, as one criterion - the aspect of free time activities - implicitly classified the four given options into two groups. However, as the use of cut-offs was still the dominating strategy across all decision making tasks, we described this class as “mainly non-compensatory”. Typical examples of response patterns for the sports and the river development task are given in Tab. 6.5. In the sports decision making task, two “cut-offs” are used (see REMA13). Firstly, all options that have some restriction with regard to participating times are eliminated. Secondly, all options that have a high risk of injury are eliminated. Consequently, the option “cycling” is chosen. The second response example for the sporting activity task (see ANME29) shows that students did not necessarily

rely only on the provided criteria and the corresponding information but also integrated prior experiences as an important criterion when reaching a decision. A similar decision making strategy was used in the river development task (see REMA13), where the degree of “potential harm to surrounding vegetation” and “effectiveness” was used as a cut-off. The second example response pattern (see ULJO24) for this task was very similar, because the individual also used cut-offs. Contrary to the first strategy, the weighing of the provided criteria is explicitly stated. If one compares these two exemplary responses, it is evident that both individuals used the same criteria “efficiency” and “potential harm to surrounding vegetation” to reach a decision. The only difference was that the second response explicitly placed the provided criteria into a hierarchy, whereas the first response does that only implicitly in the order of the considered criteria.

Tab. 6.5: Example responses for class 1: "mainly non compensatory"

<p><u>Sporting activity task</u></p> <p>"I don't like swimming because I have to consider opening hours for the swimming pool. Besides, I don't really like swimming. Concerning Volleyball, I also don't like to be restricted to participation times and the risk of injury is very high. The same holds for Inline skating. That's why I would choose cycling, because you can do it anytime and the risk of injury is very low."</p> <p>(REMA 13, grade 8)</p> <p>"I wouldn't choose cycling because I have to cycle to school everyday. I wouldn't choose swimming either, because I already do swimming in my free time and I would rather try something new. As I'm not really interested in Volleyball, I would choose Inline-skating" (ANME 29, grade 6)</p> <p><u>River development task</u></p> <p>"If you used the pesticide other plants would be affected, too. This cannot be the intention of a method against the Japanese knotweed because then you would destroy other plants. The same holds for the sheep to some extent. They eat other plants that usually grow there. Mowing is not effective enough. Planting reed and alder is efficient and it does not affect other plants in the surrounding. Thus, I'd choose reed and alder." (REMA 13, grade 8)</p> <p>"If you look at the aspect of efficiency, the method should be either "high" or "very high" in efficiency. That's why I wouldn't choose mowing. Duration and frequency is unimportant to me. Besides, it is very important that the method does not harm any surrounding plants. Thus, I wouldn't choose the pesticide or the sheep. I would choose reed and alder, because it does not harm other plants and efficiency is "high". (ULJO24, grade 7)</p>
--

Class 2 (Rasch: "mainly weighing")

The second class (29.1 %) was characterized by an extensive use of "weighing criteria" in the sporting activity and in the river development task. Performance on the trade-off and cut-off items was extremely low in these two situations (0.00-0.02; 0.12-0.03). A typical example of a response pattern for the sporting activity as well as for the river development task is given in Tab. 6.6 (see CLSE02 and CHMA25). Responses show that individuals stated which provided criteria were most important and which were less important to them.

The focus is on the criteria and their importance in relation to one another. The options and the given information were not explicitly processed or compared to each other.

Concerning the class trip task, response behaviour was different. In this context, individuals were able to apply cut-offs as well as trade-offs to reach a decision. The item profile for this situation resembles the one of the mainly “non-compensatory” class; only with lower response probabilities for these items (Fig. 6.1). The aspect of weighing criteria has the lowest mean item response probability when compared to the other two situations. Two aspects are most relevant for this difference in response behaviour. One is the aspect of context. Responses show that students were influenced by previously made decisions or experiences with a sporting activity. For example, students that already play volleyball in their free time were less inclined to consider other options and their corresponding information. Besides, students who disliked sports in general or had negative experiences with one of the sporting activities did also not engage in a decision process. The second aspect is the format of the decision making tasks itself. While students were not explicitly asked to weigh given criteria in the class trip situation, they were requested to do so in the sporting activity and river development situation. Consequently, the aspect of “weighing criteria” dominated over the other items, because students spent most of their time on considering the given criteria and reflecting on their relation.

Tab. 6.6: Example responses for class 2: „mainly weighing“

Sporting activity task

“The cost of equipment is not important to me. If you want to do sports, you have to spend some money first. More important to me are costs per month. The fewer, the better! Participation times are rather important, because you want to do a sport whenever you like. The aspect of risk of injury is only marginally important to me, because you can buy protective equipment or be more cautious. Thus, I would choose Inline-skating.” (CLSE02, student)

River development task

“I think effectiveness and effects on other plants are very important aspects. I would like to apply a method which is effective but does not disturb the ecosystem. Besides, work load should be relatively little. That's why duration and frequency of application is important to me. I would neglect the aspect of costs. Thus, I choose reed and alder.” (CHMA25, grade 8)

Class 3 (Rasch: “compensatory”)

In comparison with the other two classes, individuals in the third class (26.5 %) could be characterized by highest mean item response probabilities for all trade-off items across all three decision making tasks (0.64-0.84). In contrast, mean item response probabilities for all items concerning the use of cut-offs were comparably low (0.09-0.28). Mean item probabilities for all items that indicated weighing of provided criteria ranged from 0.24-0.50. Typical examples of response patterns for the sporting activity and the river development tasks are illustrated by REWE14 and DOSA27 in Tab. 6.7. Both students compared given options with one another. Positive and negative features of the provided options were highlighted and contrasted. No option was immediately eliminated during the decision making process because of one aspect alone. According to individuals' ability to weigh pros and cons of the given options and to compensate negative with positive aspects, we described this class as “compensatory”. Neither the different contexts nor the different formats had a strong influence on the application of the decision making strategy, as students showed the ability to use trade-offs across all tasks. However, performance on most items was lowest in the river development task, which may be a result of time pressure or general decrease in motivation towards the end of the test time.

Tab. 6.7: Example responses for class 3: „compensatory“

<p><u>Sporting activity task</u></p> <p>“Concerning cost of equipment it would be best to choose swimming. However, costs per month are highest for swimming. Concerning participation times, swimming also has advantages because you only need to consider opening hours. Moreover, “swimming” has an extensive effect on all body muscles - the same as Volleyball. Inline-Skating and cycling are very good, because you are not restricted to participation times at all. This fits my daily routines well. Concerning costs per month, six and three Euros are also affordable. Costs of equipment are the same for both sporting activities and comparably high, but costs per month are lowest. Main training effects are also similar, but I think that Inline-skating has a more intensive effect on your legs. That’s a positive effect for me personally. Comparing Volleyball with all other activities, it is the only team sport, which I would like to do. However, you are restricted to participation times and the risk of injury is high. So eventually, I would choose Inline-skating.” (REWE14, student)</p> <p><u>River development task</u></p> <p>“Concerning the pesticide efficiency is very high and duration and frequency of application is very low. But costs are high and the method affects other plants. Planting reed and alder does not have an effect on other plants but efficiency is “only” high. Grazing has to be applied several times and other plants are affected by the sheep, too. However, costs are low and efficiency is very high. Costs for mowing are also low, but efficiency, when compared to all other methods, is only moderate and thus inferior. So all in all, I would choose to plant reed and alder, because it does not have an effect on other plants and efficiency is still high.” (DOSA 27, student)</p>
--

6.6 General Discussion

The participants in the study employed three different types of decision making strategies in the presented decision making situations: “mainly non-compensatory”, “compensatory” and “mainly weighing”. The identification of these three different types was stable over all models that we fitted to the data. The most dominant strategy was the non-compensatory decision making strategy. Students focused on one aspect at a time and a decision could be reached by examining only the most important aspects. This “elimination-by aspect” strategy supports prior research in decision theory (Tversky, 1972; Abelson & Levi, 1985; Jungermann et al., 2004) as well as science education research (Hong & Chang, 2004). It

also contributes to findings that students concentrate only on a few most important aspects in a decision making situation (Kortland, 1996; Ratcliffe, 1997). Interestingly, in the class trip situation this group employed a combination of cut-offs and trade-offs in a consecutive process. Such a combination of strategies has been previously described by other researchers (e.g. Svenson, 2003; Jungermann et al., 2004). However, the use of such a combined decision making strategy could very well have been partially triggered by the structure of the given options in this decision making situation.

The compensatory decision making type showed the most consistent response behavior across all decision making situations and thus showed greatest context-independence. Individuals in this class were able to trade-off given options in a systematic way which confirms previous findings in science education (Seethaler & Linn, 2004; Yang & Anderson, 2003; SEPUP, 1995; Wilson & Sloane, 2000). Individuals were most flexible in applying trade-offs in different situations. Moreover, the SEPUP project showed that students can be trained in using trade-offs which corresponds to the fact that most of the individuals in the compensatory class were those students who already had previous experience with decision making tasks in science education.

The "weighing criteria" type is of special interest to further research and also points out the limitations of the current study. From an educational perspective, it is important to consider central criteria that describe possible options with respect to one's individual importance. Thus, it is crucial to be aware of underlying value considerations. However, individuals assigned to the "weighing criteria" type concentrated only on this aspect and hardly applied any cut-offs or trade-offs at all. The dominance of weighing criteria in this subgroup was partly a result of the format of the decision making tasks, because students were explicitly requested to weigh given criteria in the sporting activity and river development tasks. The influence of task format is also supported by the notion that students in this group did apply cut-offs and trade-offs in the decision making task, that did not explicitly request any weighing of criteria. Although the intention was to support students in explicitly weighing provided criteria, they had apparent difficulties in combining their explicit weighing with using cut-offs or trade-offs. Such a mixed strategy might have been too complex. This finding is in agreement with research that identified students' difficulties in weighing trade-offs (Seethaler & Linn, 2004). The research also supports evidence from decision theory that individuals often apply equal weights to given criteria and "add up" positive criteria when reaching a decision (e.g. Gigerenzer et al., 1999).

6.7 Conclusion

We aimed to analyze how students and university undergraduates elaborate on given options in decision making situations where every option represented a possible solution to the problem. The types of decision making strategies that were identified differed - amongst others - in their degree of processing the given information. In the light of competence assessment one might ask the question which decision making strategy might be the most elaborate one. But can one really deduce a hierarchy of decision making strategies from these findings? From a normative educational perspective, one might argue that the use of a compensatory strategy, which allows all given options to be traded-off, would be the most elaborate strategy. The ability to take positive and negative aspects into account and to elaborate on decision making situations from multiple perspectives is regarded as the highest level of performance in situations with equally legitimate options (e.g. SEPUP, 1995). However, in other decision making situations where some of the given options may have extremely negative aspects or the number of given options is extremely high, a mixed strategy of cut-offs and trade-offs would be the most successful one. Thus, for future research it will be necessary to define a characteristic set of situations in which different combinations of decision making strategies are most appropriate.

This adds to the question whether the identified strategies can represent different ability levels with respect to specific decision making situations. Focusing on the use of elaborated decision making strategies, one might argue that intuitive decision making behaviour can be a starting point in these situations. In the current study, a non-compensatory strategy was the most frequently applied strategy. It can be employed by focusing on a few most important aspects only. Such a strategy might be most appropriate when some of the given options in a decision making situation are not adequate and should be eliminated. However, if options are equally legitimate and characterized by inherent conflicts in terms of ecological, economical or socio-economical aspects, as was the case in the river development situation, this might be an intermediate level of performance. This strategy alone would probably not lead to the most successful solution. In this situation, a mixed strategy or a compensatory strategy may be most successful.

Furthermore, the study indicates that more research needs to be done to analyze students' difficulties in weighing given decision criteria and options as well as their difficulties with processing given information on the basis of this weighing. Several authors suggest the development of teaching approaches that can help to induce reflection on these strategies and processes that take place in decision making situations (e.g. Bögeholz & Barkmann, 2005; Kolsto, 2006). The ability to reflect on the structure of decision making processes and

different decision making strategies is assumed to accompany an increase in decision making competence (ibid.). Thus, it should be included in future considerations of defining different competence levels. Consequently, assessment should also take both aspects into account.


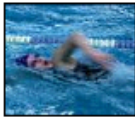


Although the present study focused only on cognitive aspects of decision making strategies, it can present a valuable starting point for quantitative assessment of decision making competence as described in the National Educational Standards. There, decision making competence is defined as the ability to make informed decisions in order to be able to participate in societal discussions. The ability to choose and to apply a decision making strategy that is appropriate for a specific situation as well as the ability to reflect on such strategies represents one crucial element of decision making competence. Together with relevant motivational and social aspects, the presented measure can contribute to the assessment of decision making competence as defined in the National Educational Standards.

6.8 Appendix A/B

Appendix A

Which sporting activity would you choose?

Imagine that you would like to participate in more sport activities during your spare time. You found some information on four different sporting activities on the internet, which is outlined as follows:

Options	Inline Skating 	Swimming 	Cycling 	Volleyball 
Criteria				
Cost of equipment	approximately 150 €	approximately 80 €	approximately 150 €	approximately 80 €
Cost per month	approximately 3 €	approximately 20 €	approximately 6 €	approximately 15 €
Participation times	any time	restricted to opening hours of pool facility	any time	playing time is dependent on availability of participants
Risk of injury	high	low	medium	high
Main training effects	<ul style="list-style-type: none"> stimulation of cardio-vascular system improved fitness stimulation especially of fundament, thigh and calf as well as lower back muscles 	<ul style="list-style-type: none"> stimulation of cardio-vascular system improved fitness stimulation of all body muscles 	<ul style="list-style-type: none"> stimulation of cardio-vascular system improved fitness stimulation especially of thigh and calf as well as back muscles 	<ul style="list-style-type: none"> stimulation of cardio-vascular system improved fitness stimulation of all body muscles improved coordination

Test task a:

How important are the provided criteria to you? Please tick only one box per criterion!

Criteria	very important	important	less important	not important
Cost of equipment	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Cost per month	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Participation times	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Risk of injury	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Main training effects	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Test task b:





Compare the four options with regard to the information provided in the table. Decide which option you would choose and explain your decision in as much detail as possible!

Appendix B

What can be done to deal with the Japanese knotweed along river banks?

Since the beginning of the 20th century Japanese knotweed has been spreading heavily along river banks. The knotweed causes severe damage, because it overrides native plants. In addition, knotweed does not provide any stabilisation to river banks which causes landslides. What is more, various animals, which depend on native plants that are typical for river banks, vanish, too.

Imagine that you are a member of the biology club at your school and you want to do something against the spread of knotweed. [...] The following information about possible removal procedures is available:

Options	Use of pesticide 	Replacement with alder and reed species 	Grazing (sheep) 	Mowing 
Criteria				
Efficiency	very high	high, if replacement plants grow well (option hinders growth and spread of Japanese knotweed)	very high	moderate
Duration and frequency of procedure	once a year, for two years	once	several times per year for 4-5 years	several times per year for 4-5 years
Cost of procedure per year	high	medium	low	low
Potential harm to surrounding vegetation	affects other plant species in proximity	no effect	existing plants will be consumed as well	low, although plants that are sensitive to mowing will also be damaged

Test task a:

How important are the provided criteria to you? Please tick only one box per criterion!

Criteria	very important	important	less important	not important
Efficiency	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Duration and frequency of procedure	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Costs per procedure per year	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Potential harm to surrounding vegetation	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>

Test task b:

Compare the four options with regard to the information provided in the table. Decide which option you would choose and explain your decision in as much detail as possible!

6.9 Weiterführende Erörterungen

Aus den Ergebnissen der beschriebenen Studie lassen sich mehrere Konsequenzen für die folgenden Studien im Hinblick auf eine Optimierung und Veränderung des eingesetzten Testinstruments ableiten. Die wichtigsten Konsequenzen ergeben sich für die Auswahl der Kontexte, die Darstellungsform der offenen Aufgaben sowie die Vorgehensweise in der Auswertung der Schülerinnen- und Schülerantworten.

Die Ergebnisse in Bezug auf das Drei-Klassen-Hybridmodell zeigten, dass eine der drei Klassen in der Auswahl und Anwendung einer Entscheidungsstrategie stark vom Kontext der Entscheidungssituation beeinflusst wurde. Dies war v.a. bei der stark durch Vorerfahrungen geprägten Sportwahlsituation der Fall. Darüber hinaus war diese Aufgabe auch dafür verantwortlich, dass einige Probandinnen und Probanden keinen Entscheidungsprozess durchführten und somit aus der psychometrischen Modellierung herausgenommen werden mussten. Aus diesem Grund wurden für die Weiterentwicklung des Testinstruments ausschließlich Situationen gewählt, die den Einfluss von Vorerfahrungen sowie die emotionale Involviertheit gering hielten. Des Weiteren wurde auf Kontexte Nachhaltiger Entwicklung fokussiert, um die Relevanz und damit auch die Akzeptanz des entwickelten Testinstruments für den Biologieunterricht zu stärken.

In Bezug auf die verschiedenen offenen Aufgabenformate konnte die Studie 1 zeigen, dass eine Gewichtungstabelle, in der Schülerinnen und Schüler vor dem eigentlichen Bewertungs- und Entscheidungsprozess die Entscheidungskriterien ihrer persönlichen Präferenz nach hierarchisieren sollten, eher eine Barriere als eine Unterstützung für einen elaborierten und nachvollziehbar dokumentierten Entscheidungsprozess darstellten. Dies wurde besonders in der Klasse der Gewichter deutlich, die nach dem Ausfüllen der Gewichtungstabelle auf eine weitere Dokumentation vollständig verzichteten. Aus diesem Grund wurde die Idee der Gewichtungstabelle für die optimierten und neu entwickelten Aufgaben verworfen. Konsequenz daraus war, dass der Aspekt der Gewichtung von Entscheidungskriterien ausschließlich an den Schülerantworten nachvollzogen wurde.

Der dritte Aspekt für die Weiterentwicklung des Testinstruments betraf die Vorgehensweise in der Auswertung der offenen Aufgaben. Damit verbunden war die Frage nach einer optimierten psychometrischen Modellierung. Die Ergebnisse der Studie 1 ließen sich mit Hilfe des Drei-Klassen-Hybridmodells am besten modellieren. Dennoch ist ein derartiges Modell, welches aus skalierbaren sowie unskalierbaren Teilpopulationen besteht, für eine Kompetenzmessung sowie für eine systematische Förderung von Kompetenzen schwierig.

Darüber hinaus stellt ein derartiges Modell eine Barriere für eine Kommunikation zentraler Strukturen sowie Entwicklungsverläufe von Bewertungskompetenz besonders unter schulpraktischen Aspekten dar.

Somit wurden die identifizierten Entscheidungsstrategien in einem optimierten Scoring Guide in ein polytomes Scoring übersetzt. Dabei wurde die non-kompensatorische Entscheidungsstrategie, die in der Klassenfahrtsituation identifizierte Mischstrategie aus einer non-kompensatorischen und einer kompensatorischen Vorgehensweise sowie die kompensatorische Entscheidungsstrategie im Hinblick auf die Intensität der Informationsverarbeitung in verschiedene Antwortkategorien umgesetzt (vgl. Anhang B. 2). Des Weiteren wurden neue Aufgaben entwickelt, die die Reflexionsfähigkeit von Schülerinnen und Schülern adäquater abtesten sollten, da ein Großteil der in der Studie 1 eingesetzten Reflexionsaufgaben entweder zu einfach war oder unterwartetes Antwortverhalten provozierte.

7 Students' Use of Decision Making Strategies with regard to Socio-Scientific-Issues - An Application of the Partial Credit Model³⁰

7.1 Abstract

Decision making about socio-scientific issues is an important aspect of modern science education worldwide. Among the many topics that represent socio-scientific issues, issues relating to the sustainable development of our environment are one crucial topic. However, difficulties exist with respect to the assessment of teaching outcomes related to these issues. This paper presents results from two quantitative studies that were conducted to develop a test instrument that measures students' use of decision making strategies in situations relating to sustainable development. Data was analysed using the Partial Credit model. Analyses concerning reliability and validity showed that the developed instrument provides an adequate measure of students' decision making strategies. Analyses indicated that the test instrument can be used among different age groups and that decision making competence increases with respect to years of education. Most elaborate strategies were characterized by the use of trade-offs, the ability to weigh important decision criteria and the ability to reflect on the structure of the decision making processes. In contrast, base level strategies were characterized by the use of cut-offs and an absence of elaborate meta-reflection. The results can provide a valuable starting point for analysing and fostering students' decision making competence in the science classroom.

³⁰ Eggert, S. & Bögeholz, S. (2008). Students' Use of Decision Making Strategies with regard to Socio-Scientific-Issues – An Application of the Partial Credit Model. Manuscript resubmitted for publication. Die Arbeit wird gefördert vom Bundesministerium für Bildung und Forschung.

7.2 Introduction

Science educators and researchers worldwide have highlighted the notion that socio-scientific issues are important for modern science education and should be attributed an important part in today's science classrooms (c.f. AAAS, 1991; National Committee on Science Standards and Assessment 1996; KMK, 2005). Socio-scientific issues are complex in nature and typically do not have a clear cut solution. While they have their basis in science, they cannot be solved by referring solely to scientific knowledge. Rather, they involve various societal aspects and have to be resolved through the integration of different, often competing, perspectives. Typically, socio-scientific issues confront students with situations in which they have to engage in argumentation or decision making situations. They are regarded as real world problems that can prepare students for becoming informed citizens. For Germany, the curricular integration of socio-scientific issues into the National Standards was initiated only in 2004, where socio-scientific decision making was defined as one competence area of science education (Standing Conference of the Ministers of Education and Cultural Affairs of the Länder: KMK, 2005). However, development of most state curricula is still under way. Implementation of socio-scientific issues into science classrooms still faces the normal restrictions of any new curricula, such as limited classroom time, dominance of teaching scientific knowledge or science teachers' reluctance to teach about issues that do not always correspond to their university education. Moreover, it is not only unfamiliarity with teaching these topics that poses difficulties with respect to acceptance; it is also the shift from an input-orientated to an output-orientated curriculum in the German National Standards. This "standardization movement", which has only recently begun, is similar to changes in science education in other countries (Scantlebury, Boone, Butler-Kahle & Fraser, 2001). At the moment, German science teachers are under pressure to teach according to these standards and to prepare students for outcome-orientated competence tests. This is complicated to achieve, given that at present new National Standards coexist together with traditional state curricula that do not correspond to the National Standards. Test instruments for measuring students' competencies are currently being prepared by the Institute for Educational Progress (IQB) in Berlin, which is in charge of developing and conducting National Assessments. With respect to science education, test instruments that focus on scientific knowledge and scientific inquiry were already developed. Development of test instruments for decision making competence are still subject to debate, because science education research on this competence area is still comparably new and measurement procedures are more intricate in comparison to test instruments for scientific knowledge for example. Because of this, more emphasis must be placed on socio-

scientific decision making, necessitating a more detailed look at how to design test instruments and how to measure such competencies. If research can give more answers to the question, “What do students really gain by engaging in Socio-Scientific Inquiry?” (Sadler, 2007, p. 371) and moreover, provide data as to how such gains can be measured, it will then also be helpful to support and encourage teachers to integrate these topics into their science classroom. The research presented in this article presents one approach to the development of test instruments concerning competencies in socio-scientific decision making. As contexts relating to socio-scientific issues are highly diverse and often demand different competencies, the study focuses on issues relating to the sustainable development of our environment and aims at getting a deeper insight into central aspects of decision making competence within the scope of this context.

Socio-scientific decision making and the curriculum

Science Standards worldwide have called for changes in the way science is taught at schools (AAAS 1991; National Committee on Science Standards and Assessment 1996; QCA, 1999). Among the most important changes are demands that science should no longer be discussed without its personal and societal implications (*ibid.*; cf. Seethaler & Linn, 2004). For students to understand the local, national and global challenges of science and technology (cf. National Science Education Standards, 1996), they do not only need to understand basic scientific concepts, but also need to understand and apply cost and benefit analyses as well as the concept of trade-offs in order to acquire important procedural competencies to reach informed decisions about these complex issues (*ibid.*; QCA, 1999). While America and England integrated socio-scientific decision making into their national curricula some time ago, this was only recently realized in the German National Standards (KMK, 2005). In Germany, decision making concerning issues of applied science was defined as one key competence area for science education and described as a prerequisite for preparing students to becoming informed citizens:

“If students have developed decision making competence, they will be able to evaluate and judge new topics in areas of applied science and will be able to take part in societal - sometimes controversial – discussions.” (KMK, 2005, p. 14)

Typical examples of applied science issues are topics of conservation biology or sustainable development such as the sustainable use of biodiversity, human impact on local and global ecosystems and problems of globalization (*ibid.*). Concerning problems for the sustainable development of our planet, students should be able to compare multiple options and to develop solutions that correspond to the aims of sustainable development (KMK, 2005).

These aspects are also central to decision making competence in those new state curricula that are already formulated (cf. Niedersächsisches Kultusministerium, 2007).

7.3 Theoretical background - Decision making competence and socio-scientific issues

Dealing with socio-scientific issues, such as sustainability issues, typically confronts students with problem or decision making situations that are both factually and ethically complex (Bögeholz & Barkmann, 2005). Contrary to routine decisions such as waste separation, decision making situations about socio-scientific issues cannot be solved intuitively or spontaneously only. Instead, students have to engage in various reasoning or argumentation processes. They have to generate options (= possible solutions), evaluate and judge relevant information and compare generated options to be able to make a choice (c.f. Eggert & Bögeholz, 2006; Jimenez-Alexandre, 2002; SEPUP, 1995; Zohar & Nemet, 2002). One aspect, which has been identified by prior science education research as crucial to informed decision making is the ability to use so called trade-offs (c. f. SEPUP, 1995; Seethaler & Linn, 2004; Siegel, 2006). The use of trade-offs is characterized as the ability to consider and compare the advantages and disadvantages of multiple options. Individuals' ability to use trade-offs is not only a research interest in science education but also in various psychological research fields (c.f. Payne, Bettman & Luce, 1998; Jungermann, Pfister & Fischer, 2004). There, the use of trade-offs is described as a compensatory decision making strategy, because positive aspects of an option can compensate for negative aspects (c.f. Abelson & Levi, 1985; Jungermann et al., 2004).

Results from science education research, with respect to students' use of trade-offs are diverse. While some researchers supported the notion that students are able to use trade-offs, others revealed shortcomings. Ratcliffe (1997) found that students are able to compare a set of given options, but only in an unsystematic way. In addition, she found that systematic reasoning, over the time span of an intervention on decision making about a controversial issue, improved only to a limited extent. Hogan (1999) as well as Hong and Chang (2004), found that students focused only on a limited number of aspects in a decision making situation and did not use trade-offs to compare possible options. Instead, most students established cut-off levels that options had to meet and thus reduced the number of possible options through an elimination process (Hong & Chang, 2004). This strategy is, in contrast to compensatory strategies, described as non-compensatory (c.f. Jungermann et al., 2004).

The Science Education for Public Understanding Project (SEPUP) showed that students' competencies with respect to the use of trade-offs especially improved after long-term educational measures on controversial issues (Wilson & Sloane, 2000). A more recent study (Siegel, 2006), which also used SEPUP teaching material as well as assessment procedures, showed that before an intervention students were able to make use of trade-offs on the basis of some relevant pieces of evidence but only to a limited extent. However, performance after the intervention improved significantly in some of the analysed decision making situations. Seethaler and Lynn (2004) analysed eighth grade students' decision making competence in a teaching sequence about different agricultural methods of crop growing, among them genetically modified crops. Students were requested to choose one of four given options, to give evidence to support their position and to give evidence for rejected options. Results showed that students were able to trade-off given options. Moreover, they not only elaborated on the chosen option, but could also support evidence for rejected options. However, students' answers fell short of explicitly weighing the trade-offs they presented. The ability to explicitly weigh important decision criteria or trade-offs in a decision making situation seems to pose difficulties for students in general. Several authors attribute this to the fact that weighing decision criteria or trade-offs includes prioritizing conflicting values (Bögeholz & Barkmann, 2005; Jimenez-Aleixandre, 2002; Kolsto, 2006). To respond to these difficulties and to foster students' decision making competence, Kolsto (2006), among others, suggests that presenting different reasoning patterns can be a means to induce meta reflection about decision making processes and inherent value conflicts and thus, can be one way of fostering students' decision making competence.

In regards to the assessment of such competencies, of interest would be to know which of the described elements of decision making strategies, that being the use of trade-offs, cut-offs, weighing decision criteria or reflecting on argumentation and reasoning processes, are more difficult to perform than others. This is important for identifying students' competencies, but also with respect to the development of such competencies. To summarize prior research: the use of trade-offs is a crucial element necessary to reach an elaborate and informed decision, because students do not only need to elaborate on the chosen option but also to elaborate on rejected options. As several pieces of information must be simultaneously integrated, this can be regarded as a more difficult strategy than the use of cut-offs (\triangleq non-compensatory strategy), when information can be consecutively analysed (c.f. Abelson & Levi, 1985). Prior research in science education (cf. Hogan, 1999; Hong & Chang, 2004) showed that students use a non-compensatory decision making strategy in order to decide between multiple options. However, in decision making situations with multiple, equally legitimate options, such a strategy is not to be favoured, as several important decision criteria might be

neglected. Thus, the transition from a non-compensatory to a compensatory decision making strategy might be one aspect of increasing decision making competence (fig. 1).

Moreover, prior research in decision theory showed that individuals tend to reduce the amount of information given by using cut-offs as a first step and then use trade-offs to compare remaining options (e.g. Jungermann et al., 2004; Payne et al., 1998). Such a decision process is characterized by a combination of both a non-compensatory and a compensatory decision making strategy and might well represent an intermediate level of performance. The ability to explicitly weigh decision criteria or weigh trade-offs in a decision making situation is, according to research difficult to perform and thus might only be found at a high ability level (c.f. Kolsto, 2006; Seethaler & Linn, 2004).

Moreover, according to Kolsto (2006) and others, one can hypothesize that the transition from one level of performance to the next must be accompanied by an understanding of the processes that precede a decision. Thus, some sort of meta-reflection on the structures of decision making strategies might be found at different performance levels (fig. 1). Referring to the construct of problem solving in the OECD/PISA studies (Programme for International Students Assessment; OECD, 2004), an elaborate level of meta-reflection would be characterized by monitoring the decision making process and detecting possible flaws while working on the problem.

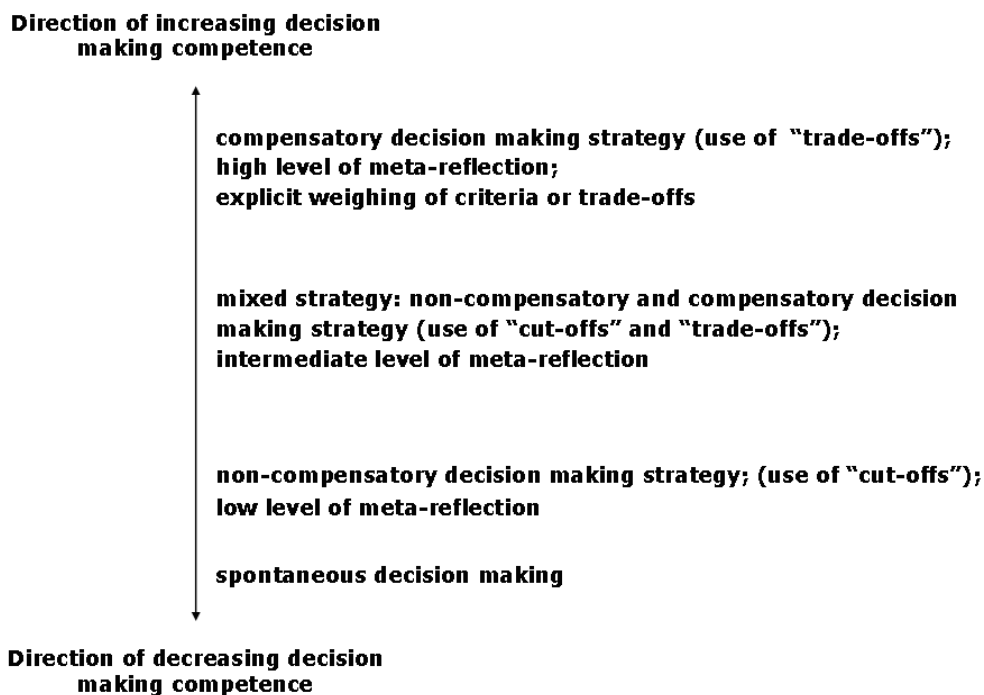


fig. 7.1: Assumed increase in decision making competence with respect to the use of and the reflection on decision making strategies

The Measurement Model

To analyse whether the assumed continuum of increasing decision making competence (cf. fig. 1) accurately pictures students' competence, the Partial Credit model, which is an extension of the traditional Rasch model for polytomous data, was chosen for data analysis. The rationale behind this model is that partial success on a specific test item is expressed by partial credits. Moreover, responses that are assigned partial credits are of hierarchical nature. This means that a response, which is given a higher partial credit, is qualitatively better than a response, which was assigned a lower partial credit. This procedure is extremely useful for the assessment of procedural competencies, such as argumentation or problem solving skills, where students' answers are usually not only marked as correct or incorrect (cf. OECD/PISA, 2004). Thus, it can provide a more detailed picture of students' competencies. In the present study, we expected students to apply different decision making strategies and to show different qualities of monitoring and reflection processes. By using the Partial Credit model, we aimed to picture students' competencies on a continuum from a base level of competence to an elaborate level of competence. Moreover, we chose the Partial Credit model over the Rating Scale model (cf. Wright & Masters, 1982), which is the pendant to Likert Scales in Rasch modelling. The Partial Credit is less restrictive than the Ratings Scale model, because the former allows for different response categories in different items. In addition, the relative difficulties of partial credits can vary from item to item.

In view of the above, the presented study addresses the following research questions:

- Can the Partial Credit model adequately explain students' use of and reflection on decision making strategies on one latent continuum (cf. fig. 1)?

This main research question is closely connected to the quality, i.e. reliability and validity, of the developed test instrument. Contrary to most prior research, which focused on students' decision making competence of a particular age group, we needed to picture a continuum of increasing competence. Thus, we needed to analyse decision making competence across a wider range of students. Consequently, our second research question was:

- Can the developed test instrument provide a measure for the use of and reflection on decision making strategies across age groups?

In addition to these two research questions, we analysed other factors that could contribute to the explanation of students' decision making competence. Thus, the third research question was:

- In what way can students' performances in other school subjects and verbal skills contribute to the explanation of students' decision making competence?

7.4 Methods

The research reported here was conducted in two studies. In the first study, the focus was on students' decision making competence. A paper-and-pencil test was administered to 370 German secondary school students (195 female) from 10 grammar and comprehensive schools. Participants were selected from grades 6 (n=105), 8 (n=100), 10 (n=82) and 12 (n=83) and ranged from 12 to 19 years. These grades represent important learning stages in the National Standards. Grade 12 represents the final year at secondary schools in Germany. Additionally, 78 second-year biology university undergraduates (53 female) were tested. The complete sample consisted of 448 individuals. Final analyses were conducted with 436 individuals, because data analysis showed that 12 individuals had to be excluded because of aberrant response behavior (cf. results, p. 19).

In the second study, we analysed 83 students from grade 10 (mean age =15.86) from one grammar school. These students were not the same grade 10 students who took part in the first study. At the end of grade 10, students typically receive their *mittlere Reife*. This degree is equivalent to the intermediate high school certificate in the US or the GCSE in Britain. Students can at this point decide to either leave school or enrol for another two years to complete their *Abitur*, which is an equivalent to the high school diploma in the US or the A-levels in Britain. The students in the second study had to complete the same decision making questionnaire that was used in the first study. In addition, they were given test items concerning verbal skills. Subject grades in German, mathematics and biology were also collected. The aim of the second study was to back up results from the first and to receive more information on the decision making questionnaire in terms of its validity. Data from both studies was analysed with the Partial Credit model (Masters, 1982), using the software programs WINMIRA (von Davier, 2001) and Conquest (Wu, Adams & Wilson, 1998). Correlations and regression analyses were conducted, using SPSS 12 (SPSS Inc., 2003).

Test Instrument - Decision making questionnaire

Item construction

To measure the use of and reflection on decision making strategies with respect to sustainable development issues, a questionnaire with open-ended items was developed. Four main steps with respect to the development of test items were carried out.

First, National Science Standards (especially standards for biology education), newly developed state curricula and relevant research literature were analysed. After that, central criteria for item construction were defined.

Items should...

be of relevance to the National Standards as well as new state curricula.

present ...

a) real-world decision making situations that have multiple possible solutions and

b) topics related to conservational biology or sustainable development.

enable ...

students to make use of decision making strategies,

students to reflect on and monitor decision making processes and

enable students of various age groups to work on the presented decision making situations.

On this basis, preliminary test items were developed. Secondly, these test items were discussed with researchers in the research group as well as a group of selected teachers with expertise in decision making competence (n=10), who took part in the national curriculum project "biology in context" (Bayrhuber et al., 2007). Teachers were asked to evaluate the constructed test items with respect to the topics' relevance, the potential to measure decision making competence and the test items wording and phrasing. Teachers' recommendations were then used to create an optimized version of the test items.

Thirdly, we pre-piloted optimized test items with both junior high school students (n=25) and university undergraduates (n=20). These two groups were selected in order to check whether the test items were comprehensible to both ends of the competency spectrum.³¹ After students had responded to the test items, they were asked to comment on the test items' content, difficulty and comprehensibility. Results gave useful advice to optimize the test items in terms of wording, phrasing and presentation of options.

Fourthly, a quantitative pilot study (n=291) was conducted to analyse the test items' functioning within a larger sample. The study was used to verify that decision making strategies described in the literature could actually be found with respect to decision making situations of sustainable development. Moreover, we used different contexts and task formats as well as different items to trigger students' reflection and monitoring processes. We also checked whether it was possible for students to complete the questionnaire within 45 minutes (one school lesson). One central result of this pilot study was the identification of typical situations that were appropriate and effective to measure students' competencies with regard to sustainable development issues. Moreover, results showed that it was not useful to ask students to reflect on their decision making processes shortly after they had reached a decision. In particular, we did not get the whole range of qualitatively different reflection

³¹ As open ended items were used for measurement, it was possible to use the same test items across such a wide range of participants. Contrary to closed items, an individual's competency was judged by the quality of the given answer and not by solving test items that represent different competency levels (cf. item construction in the SEPUP project: Roberts et al., 1997).

processes as described in our working hypothesis (cf. fig. 1). Items that requested students to reflect on decision making processes reached by other students proved to be more effective. For the main study, final items were again discussed with both researchers in the research group and teachers with expertise in decision making competence.

*Final test items*³²

Final items were separated into two booklets. In the first booklet students were confronted with two decision making situations. Here, students were requested to compare all given options in as much detail as possible and finally to reach an informed decision. The first situation was concerned with the problem of overfishing (overfishing of codfish (*Gadus morhua*) in the Baltic Sea; see Appendix A), the second with the topic of neophyte invasion (a neophyte plant (*Fallopia japonica*) that causes land slides along river banks). These tasks represent typical decision making situations with multiple possible options. They have their basis in ecology, are part of lower as well as upper secondary school curricula, but are suitable to broaden the topics with respect to the sustainable development of our environment.

The topic of overfishing of Codfish in the Baltic Sea is a typical example of the general problem of overfishing, as most Baltic Sea fish are threatened by extinction. All four options presented in the test item are being executed by the European Commission for Fisheries (Europäische Kommission, 2005). Covering all dimensions of sustainable development, ecological, economical and socio-economic aspects were included to describe the four options in this test task (c.f. Appendix A).

The topic of the neophyte plant *Fallopia japonica* is one typical example of negative impacts of invasive species on local ecosystems. Next to *Fallopia japonica* there are various other *Fallopia* species or other neophyte plants such as the Himalayan Balsam that have similar impact on river ecosystems. In many regions of Germany these plants are now a frequently observed species along river banks. All options that are presented to treat *Fallopia japonica* are described and discussed by the Federal Agency for Nature Conservation (BfN, 2003). In order to choose one option, students had to consider the damage that a neophyte plant can cause to the river ecosystem. On the basis of this information, students then had to analyse four different options with respect to efficacy and possible consequences to the surrounding vegetation. In addition, costs as well as duration and frequency of application of the different options had to be taken into account. The completion of the first booklet took approximately 25 minutes.

³² Der komplette Fragebogen kann bei den Autorinnen angefragt werden. Eine gekürzte Version des Fragebogens ist in Appendix A / B (S. 116-118) gegeben.

The second booklet was designed to analyse how students reflected on the structure of different decision making strategies that were presented to them. Here, students did not have to reach a decision themselves but were confronted with decision making processes made by other students. The decision making processes that were presented were based on real students' answers from the quantitative pilot study. The context of the second booklet was a consumer choice situation about buying chocolate bars (a shortened version of the task is shown in Appendix B). Consumer choice situations are typical examples for education of sustainable development at German schools. They are suitable to teach about global interrelations and the impact of personal decisions on global trade. The decision making processes that were presented to students represented different decision making strategies: a non-compensatory ("Paul's strategy"), a compensatory ("Martina's strategy") and an intuitive or justifying strategy ("Stefan's strategy"). "Paul" focuses on only one aspect at a time and eliminates options that do not fulfil his cut-off levels. "Martina" compares all given options and trades off both negative and positive aspects. Likewise, she explicitly weighs the given criteria that are used to compare the different chocolate bars. "Stefan" spontaneously reaches a decision and then justifies his decision afterwards by highlighting positive and ignoring negative aspects. Thus, these decision making strategies represent different levels of decision making competence as proposed in fig.1. Moreover, all the strategies that were presented included a shortcoming or possible weakness. This aspect had to be identified by students and a suggestion for improvement had to be made. In summary, all students were requested to describe the presented decision making strategies, to compare them with one another and to propose suggestions for improvement. The completion of the second booklet took approximately 20 minutes. Students who finished early gave the test to the test administrators and were allowed to leave the room.

In the second study, we supplemented this final decision making questionnaire with two subtests from a cognitive abilities test (Heller & Perleth, 2000), analysing students' verbal skills. In the verbal classification subtest, students had to match one word to another from five possibilities that has a similar meaning. In the verbal analogies subtest, students were given one pair of words and had to complete a second pair from five possibilities. Cronbach's alpha for the two subtests were found to be $\alpha=0.72$ and $\alpha=0.56$. The second alpha value is low, but confirms alpha values of the reference, representative sample of the cognitive abilities test (cf. Heller & Perleth, 2000, p. 20: subtest 1: $\alpha=0.69$, subtest 2: $\alpha=0.56$).

Background variables were also collected, these being grades in biology, mathematics and German. Grade in biology was also collected in the first study.

Scoring and Data Analysis

Students' written answers for booklet 1 and 2 served as the basis for our analyses. Every answer was analysed using a scoring guide, which consisted of central indicators for a) the use of and b) the reflection on decision making strategies (cf. tab. 1 and tab. 2). These indicators were established on the basis of our theoretical framework, on results of the pilot study as well as inductively derived from the data using a deductive-inductive approach (Mayring, 2000). To develop the scoring guides, 20% of students' answers were transferred into a software program for qualitative data analysis (MAXqda2). Students' answers were coded by two researchers independently who used a first version of the category system. More indicators and/or categories were added to the coding system when needed. After completion of data analysis, the two researchers compared and discussed their developed coding systems and agreed on a final version of the scoring guides (tab. 1 and tab. 2). After that, the whole sample was analysed according to this scoring guide by the two researchers independently. Inter-rater reliability was found to be 95%.

Tab. 7.1: Scoring Guide for booklet 1

Decision making tasks: Codfish (1, 2, 3) and Neophyte (4, 5, 6)			Assigned Scores	Adjusted Scores after Data Analyses
No.	Indicator	Description		
		Students and Undergraduates...		
1, 4	strategy for chosen option	• do not state anything	0	0
		• state <u>only</u> positive aspects	1	0
		• state positive <u>and</u> negative aspects (→ use of "trade-offs")	2	1
2, 5	strategy for rejected options	• do not state anything	0	0
		• state only negative aspects (→ use of "cut-offs")	1	1
		• state negative aspects for some of the rejected options and negative as well as positive aspects for other rejected options (→ use of "cut-offs" first and then use "trade-offs" for remaining options)	2	2
		• state negative and positive aspects for all rejected options (→ use of "trade-offs")	3	3
3, 6	weighing criteria	• do not explicitly weigh given criteria to own individual importance	0	0
		• explicitly weigh given criteria to own individual importance	1	1

Tab. 7.2: Scoring Guide for booklet 2

Decision making task: Consumer choice situation (chocolate)			Assigned Scores	Adjusted Scores after Data Analysis
No.	Indicator	Description		
		Students and Undergraduates...		
7	Paul's strategy (“non-compensatory”)	• do not respond to the task	0	0
		describe Paul's strategy...		
		• by referring to the <u>information</u> that is given for the chosen chocolate (no. 3)	1	1
		• by referring to <u>some aspects</u> of the decision making strategy	2	1
		• by referring to <u>all aspects</u> of the decision making strategy	3	2
		• <u>accomplish Score 3</u> and can explain the <u>conflict</u> that was inherent in this decision making strategy (trade versus cultivation of cacao plants)	4	3
8	Martina's strategy (“compensatory”)	• do not respond to the task	0	0
		describe Martina's strategy...		
		• by referring to the <u>information</u> that is given for the chosen chocolate (no. 4)	1	0
		• by referring to <u>some aspects</u> of the decision making strategy	2	1
		• by referring to the <u>main aspects</u> of the decision making strategy	3	2
		• by referring to <u>all aspects</u> of the decision making strategy	4	3
9	choice of chocolate	• do not select the chocolate that suits Paul's and Martina's preferences best	0	0
		• select the chocolate that suits Paul's and Martina's preferences best	1	1
10	suggestion for improvement of Martina's strategy	• do not give a suggestion for improvement	0	0
		• give a suggestion for improvement by referring to the given information	1	0
		• give a suggestion for improvement by referring to aspects of the decision making strategy	2	1
11	Stefan's strategy (spontaneous / justifying)	• do not respond to the task	0	0
		describe Stefan's strategy...		
		• by referring to the information that is given for the chosen option (no. 3)	1	1
		• by referring to aspects of the decision making strategy	2	2

Conducted Analyses

Study 1

Data was analysed using the Partial Credit model. Rasch analysis has several advantages when measuring students' competencies. Among the most important is the transformation of raw scores and, as a consequence, the representation of both persons and items on the same equal linear scale. The difficulty with raw scores is that they represent nominal or ordinal data. They are quantitative observations that are based on counting observed events (Wright & Linacre, 1989). With these observations, one can define an origin ("no counts"), but one cannot tell anything about the distances between one and two counts, two and three counts and so forth. In other words, the addition of raw scores does not define the distances - intervals - between these counts. Consequently, raw scores have to be transformed in a way that distances can be explained in terms of interval measures. This is exactly done with the help of the Rasch model, which converts raw scores to logarithmic units, namely logits. This logarithmic transformation makes it possible to analyse students' abilities and item difficulties on an equal interval linear scale (c.f. Boone, 2005; Wilson, 2005). Persons are plotted as a function of their ability; items are plotted as a function of their difficulty on the same logit scale (Wilson, 2005; see fig. 2). More able persons and more difficult items are at the top of the map, less able persons and less difficult items are at the base of the map. A person who is plotted on the same level as an item has a 50% chance of getting that item correct. In fig. 2, a person who has an ability of 0 logits has a 50% chance of getting item 4 correct. A person, who has an ability of -1 logits, has a chance lower than 50% of getting item 4 correct. A person who has an ability of 1 logits has a chance of more than 50% of getting item 4 correct.

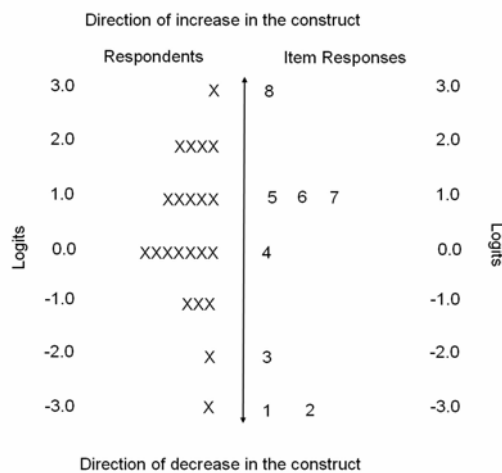


fig. 7.2: Person-Item map (adapted from Wilson, 2005, p. 96)

Person-item maps provide a very useful starting base to evaluate the functioning of a test instrument. They give valuable information whether a theoretically proposed increase on a construct can actually be identified empirically. Moreover, several statistical tests can be conducted to judge the functioning of a test instrument in terms of reliability and validity. The following tests were applied in the present study.

Reliability indices for the developed test instrument were calculated both for persons and items, namely person separation reliability and item separation reliability. Person separation reliability is an estimate of how well one can differentiate persons on the measured variable (Wright & Masters, 1982). The estimate refers to the observed response variance that is reproducible and thus, is based on the same construct as Cronbach's alpha (Bond & Fox, 2000). In addition to person separation reliability, expected a posteriori reliability (EAP/PV) was calculated. EAP/PV reliability can be favoured over person separation reliability, because the latter can sometimes underestimate reliability (cf. Walter, 2005). Item separation reliability was calculated in the same way as for person separation reliability, substituting item variance for person variance. Item separation reliability expresses "how well the sample of subjects has spread out the items along the measure of the test" (Fisher, 1992, p. 238).

Secondly, fit statistics were evaluated both for persons and items (Boone, 2005). These fit statistics, also described as fit validity or response performance validity (Wright & Stone, 1999), contribute to the test instruments' validity. Item fit values (using item Q-indices) allow conclusions to be drawn about how well an item fits to the theoretical construct that is to be measured. Person fit statistic was analysed in order to depict any individuals who responded to the items in the questionnaire in an unexpected way. Thus, this fit statistic (using z-scores) provides a technique which can highlight an unexpected response behavior which can then be analysed in more detail (Boone, 2005). Persons who were assigned extreme z-scores were checked in detail and - in some case - eliminated from further analysis.

Thirdly, all polytomous items were analysed with regard to their threshold parameters. Here, Thurstonian thresholds were used. More specifically, we analysed whether the assigned scores 0, 1, 2, 3 and 4 (see tab. 1 and tab. 2) were meaningful steps with regard to difficulty, that is to say whether threshold parameters were ordered. If analyses showed that the assigned scores are not ordered in terms of difficulty, the Partial Credit model would not hold for explaining the measured construct.

After detailed analysis of the developed test instrument, person ability parameters³³ were analysed using classical test theory. Simple linear regression analysis was conducted to investigate the influence of years of education on students' abilities.

Data obtained from study 2 was also analysed with the Partial Credit model. Again person ability parameters were analysed using classical test theory. Correlations were conducted to analyse the influence of subject grades and verbal skills on decision making competence.

7.5 Results

Development of test instrument using the Partial Credit model

The Partial Credit model was fitted to the whole sample (n=436) as well as to each subgroup (=different years of education) separately. It was found that the Partial Credit model represented the best solution not only for the whole sample but also for each subgroup. The following results are based on the whole sample.

As a first step, the Partial Credit model was compared with other probabilistic test models such as Mixed Rasch models or Latent Class analyses (c.f. Rost, 2004). These test models assume that the latent trait that is to be measured, cannot be explained on one latent continuum. This preliminary analysis was a prerequisite to use the Partial Credit model for data analysis. All applied models were compared by means of Information Criteria, such as the Bayesian Information Criterion (BIC) and the Consistent Akaikes Information Criterion (CAIC) (c.f. Rost, 2004). Results showed that, the Partial Credit model had the lowest – and thus, best – values for BIC and CAIC and consequently was identified as the model with best explanatory power. Results that are presented in the following section only refer to the Partial Credit model.³⁴

Concerning reliability, item separation reliability was found to be 0.99. This high item separation reliability indicates that great confidence can be placed in the replicability of item placement across future samples. Person separation reliability was found to be moderate with 0.66. However, EAP/PV reliability showed a better result with 0.74. Inter-rater reliability was high with 95%.

³³ In Rasch analysis, person ability is estimated for every person. In this study, a special method, called Warm Likelihood estimation, was used. As a consequence of the Rasch model, WLE's represent linear measures. They can thus be favoured over raw scores for parametric tests in classical test theory, which assume linearity.

³⁴ Eine Zusammenstellung der Werte der Informationskriterien sowie der P-Werte des Bootstrapverfahrens konnte hier aus Gründen der Schwerpunktsetzung des Artikels nicht dargestellt werden. Alle Werte sind in den ergänzenden Ausführungen dargestellt (vgl. Kap. 7.9).

Item fit analyses showed that all items fit the Partial Credit model. All Q-indices were in the acceptable range of 0.00 to 0.30 (cf. Appendix C). Zq values supported these results (ibid.). Analyses of person fit values showed that only few students (n=12) responded to the questionnaire in an unexpected way. These students were eliminated from the final analysis. In most of these cases, students hardly responded to the items for the first booklet, but only responded to the items in the second booklet. All other students showed regular person fit, which supports the appropriate construction of the developed test items. Analyses of threshold parameters indicated that most of the assigned scores were ordered with respect to item difficulty (see tab. 1 and tab. 2).

No adjustments had to be made for items 2, 5, 9 and 11. Items 3, 6 and 9 were dichotomous items and thus were assigned only one threshold. Adjustments had to be made for items 1, 4, 7, 8 and 10. Here two categories had to be collapsed together as thresholds did not increase in difficulty (c.f. Bond & Fox, 2000). After these adjustments, all thresholds parameters were ordered. A more detailed description for all items and the corresponding thresholds is given in a “smoothed” version of the Person-Item map (fig. 3). Each case (x) represents 5.2 cases.

		Respondents	Items Codfish task	Items Neophyte task	Items Chocolate task
logits	2			6: WC (2.45)	
	1		3: WC (1.26)		10: suggestion for improving compensatory strategy (1.70)
		X			8.3: D. compensatory strategy (1.05) 11.2: D. intuitive, justifying strategy (0.88) 7.3: D. non-compensatory strategy (0.73)
	0	XXX	2.3: RO (0.30)	5.3: RO (0.43)	
		XX			8.2: D. compensatory strategy (0.16)
	-1	XXXXXX			7.2: D. non-compensatory strategy (-0.20)
		XXXXXX			
		XXXXXXXX		4: CO (-0.66) / 5.2: RO (-0.70)	
		XXXXXX	2.2: RO (-0.88)		8.1: D. compensatory strategy (-0.95)
	-2	XXXXXXXX	1: CO (-1.20)		9: choice of chocolate (-1.30)
XXXXXXXX				11.1: D. intuitive, justifying strategy (-1.50) 7.1: D. non-compensatory strategy (-1.54)	
XXXXXXXX		2.1: RO (-1.95)			
-3			5.1: RO (-2.21)		
	XXXX				
...					
	-4	XX			

Note. CO= chosen option; RO=rejected options; WC=weighing criteria; D=description. Item Thresholds are given in brackets.

fig. 7.3: Person-Item map: Partial Credit model

Items 1 and 4 indicate response behavior concerning the chosen option. For these two items we had to adjust prior assigned scores. Unsurprisingly, students found it relatively easy to highlight positive aspects for the chosen option. As a consequence, we could not separate zero performance on these items from stating only positive aspects (c.f. tab. 1). Thresholds for items 2 and 5 indicate how students processed the information for the rejected options in both tasks. The hypothesized increase in difficulty from a non-compensatory, to a mixed and to a compensatory strategy for the rejected options could be confirmed as all thresholds for both tasks were ordered (2.1, 2.2, 2.3 as well as 5.1, 5.2, 5.3).

Comparing thresholds between item sets 1, 2, 3 (codfish) and 4, 5, 6 (neophyte) one can also see that both item sets are relatively comparable with respect to item difficulty despite of the different context (cf. fig. 3). In addition, the person-item map indicates that the ability to use trade-offs is similarly difficult for the chosen and the rejected options (items 1, 2.2 and 4, 5.2). Items 3 and 6 indicate students' ability to explicitly weigh the given criteria according to personal importance. Thresholds show that these items, together with one item from the second booklet (item 10) were most difficult. In particular, these items contributed to the fact that the developed test instrument was, in general, slightly too difficult for the analysed sample, as no students were found who had a 50% chance of giving an adequate response to these items. Consequently, standard errors for these three items were highest (0.071; 0.076; 0,073).

Answering items 7, 8 and 11, students had to describe and analyse a non-compensatory, a compensatory and an intuitive, justifying decision making strategy. First thresholds (7.1, 8.1, 11.1) for all three items show that it was easiest for students to describe and comment on the presented strategy by referring to the content information that was given. Typically, students referred to the characteristics of the different chocolates that were chosen by "Paul", "Martina" or "Stefan". Concerning Paul and Martina, students frequently focused on the aspect of fair trade vs. cultivation of cacao plants.

As the following exemplary response shows, the strategies that were used by Paul and Martina are not described:

"Martina and Paul didn't choose chocolates number one and two, because of its genetically modified ingredients (no. 1) and because of the price (no. 2). Paul finally chose chocolate no. 3, because it is a fair trade chocolate, which was important for him. Martina chose chocolate no. 4, because it was grown organically and besides, this chocolate was cheaper than chocolate no. 3."

(SETH 28, 8 years of education, adjusted score 1 on item 7, adjusted score 0 on item 8; cf. tab. 2).

Similarly, descriptions of Stefan's strategy also focus on the content information:

“For Stefan, it was most important to be able to buy the chocolate everywhere. All other aspects were rather unimportant to him, e.g. whether the cacao plants are organically grown or not.”

(NEAX05, 8 years of education, score 1 on item 11; cf. tab. 2)

In contrast to this mainly content-focused description, it was assumed to be more difficult for students to explain the strategies that were used by the different persons. For example, students had to identify the elimination process as one characteristic of the non-compensatory strategy:

“Paul reached a decision by carrying out an elimination process.”

(BEST28, 8 years of education; adjusted score 1 on item 7; cf. tab. 2)

However, we could not separate the difficulty of such an answer from answers that merely focussed on the content information (see example SETH28 above) and had to merge categories 1 and 2.

However, answers that explained the elimination process as a consequence of the established value hierarchy, increased with respect to item difficulty:

“Paul considered all given criteria and decided that cultivation of cacao plants and genetically modified ingredients are very important aspects for him. Because of that, he eliminated chocolates no. 1 and 2. Important to him was that the chocolate is a fair trade chocolate. Because chocolate no. 4 does not meet this criterion, he eliminates this one as well and chooses chocolate no. 3, because this is the only chocolate that remains.”

(DOGU14, 8 years of education, adjusted score 2 on item 7; cf. tab. 2)

Concerning the compensatory strategy, students had to explain that this strategy made it possible to consider all options and all criteria at the same time:

“First, Martina assigns points for every criterion and every chocolate. The better a chocolate does on one criterion, the more points it gets. After having assigned points to all chocolates, she sums all scores up and chooses the chocolate that got the most points.”

(BEST28, 8 years of education, adjusted score 2 on item 8; cf. tab. 2)

To explain that this strategy made it possible to consider all given criteria according to their importance was found to be more difficult:

“Martina’s decision making process is different to Paul’s. She assigns points for every criterion and every chocolate. Besides, she orders the given criteria according to personal importance by assigning a multiplication factor to each criterion. Taking this multiplication factor into account, she adds up all the points for each chocolate and chooses the chocolate that got most points.”

(DOKA03, 8 years of education, adjusted score 3 on item 8; cf. tab. 2)

Most difficult was to identify the strategy weaknesses (items 7.3 and 10):

“Paul eliminated those chocolates that did not fulfil his expectations. However, all of the presented chocolates had a disadvantage, so that eventually, he would have had to eliminate all chocolates. But he chooses chocolate no. 3 and neglects the negative aspect of “cultivation of cacao plants”.

(SIHE27, 12 years of education, adjusted score 3 on item 7; cf. tab. 2)

“[Concerning Martina’s strategy] It was inappropriate to include the aspect “availability” into the decision making process, because this aspect was unimportant to Martina in the first place. Thus, she should have ignored this criterion.”

(MADE29, 12 years of education, adjusted score 1 on item 10; cf. tab. 2).

The influence of years of education on decision making competence

Concerning research question two, we analysed the influence of years of education on students’ use of decision making strategies using linear regression analysis. Person ability parameters (WLEs) were exported into SPSS, transformed and z-standardized (mean ability = 500, SD = 100). The regression graph shows that years of education has a strong effect on students’ decision making competence ($r^2=0.284$, $p<0.001$; c.f. fig. 4). Additionally, t-tests were calculated to analyse the difference in mean person ability between the different age groups. T-tests showed that there was a significant increase in person ability from 6 to 8 and 10 to 12 years of education. However, no significant increase in person ability could be found between 8 and 10 years of education as well as 12 years of education and university undergraduates.

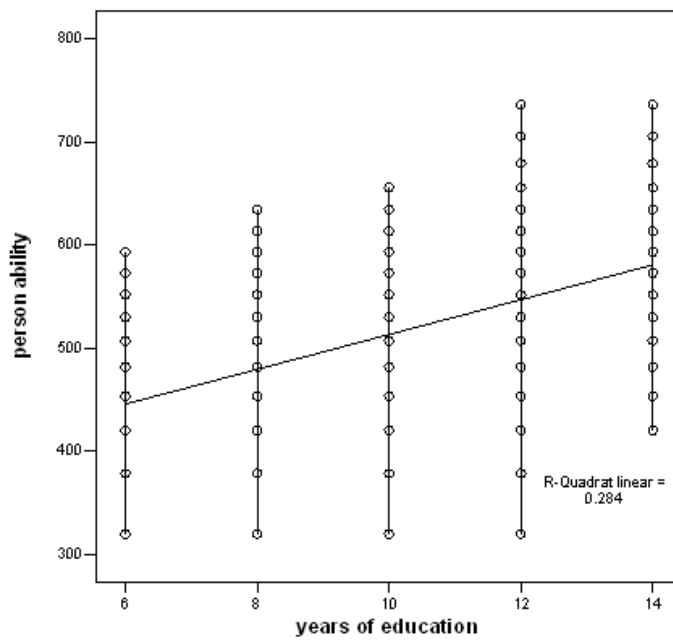


fig. 7.4: Influence of years of education on individual's performances on the decision making items

The results from regression as well as Rasch analysis indicate that the same test instrument can be used across different age groups. However, this assumption can only be sustained if the test items function in a similar way across the different age groups. In terms of item response theory, this means that item difficulties need to be comparable between the different subgroups. A way to deal with this problem is to analyse “differential item functioning (dif)” (cf. Bond & Fox, 2001; Wilson, 2005). For this reason, we divided the sample into two major subgroups that represent German lower secondary (Sekundarstufe 1: 6, 8 and 10 years of education) and German upper secondary students (Sekundarstufe 2: 12 years of education) and university undergraduates (14 years of education), because these two groups are most likely to react to the given test items in a different way. Although a Pearson correlation of item estimates proved to be highly significant ($r=0.87$, $p<.001$), an analysis of logit differences showed large dif for items 7 and 10. Both items refer to the identification of the possible weaknesses in the presented non-compensatory and compensatory decision making strategies. Logit differences show that it was comparably easier for upper secondary students and university undergraduates to identify the weaknesses in the presented strategies, in comparison to lower secondary students.

The influence of other school subjects and verbal skills

The aim of the second study was to analyse possible relationships between students' decision making competence and performance in subjects other than biology as well as in relation to verbal skills. As a first step of data analysis, person ability parameters concerning the decision making questionnaire had to be obtained. As the sample size in the second study was relatively small, data sets from the first and the second study were merged in order to estimate person ability. Following this, all calculations were conducted only for the second sample (n=83). Results of Pearson correlations showed a significant relationship between decision making competence (person ability parameters) and grades in German ($r=.23$, $p<.05$) and Mathematics ($r=.24$, $p<.05$). Correlations however, were low. In contrast to the first study, no significant relationship could be found between decision making competence and grades in biology ($r=.04$). However, the correlation was only very small in the first study, too ($r=.20$, $p<.05$). Moreover, no relationships could be found between decision making competence (person ability parameters) and students' verbal skills (subscale verbal classification: $r=.06$, subscale verbal analogies: $r=-.03$).

7.6 Summary and Discussion

Development of test instrument using the Partial Credit model

Concerning the first research question, we could show that the Partial Credit model can adequately explain students' use of and reflection on decision making strategies on one latent continuum. The Partial Credit model was the best model to explain response behavior, not only for the whole sample, but also for each subgroup (grade). In addition, it was the best model in comparison to other item response models. Test reliability was high concerning item separation reliability and moderate concerning person separation reliability and EAP/PV reliability. For competence assessment tests, values for person separation reliability and EAP/PV reliability are in an acceptable range and are comparable to other competence assessment studies, such as the 2003 PISA national tests for several subdimensions of scientific literacy (Senkbeil et al., 2003). Here, reliability indices for subdimensions, such as inductive or deductive reasoning range between 0.66 and 0.74. These results are also in line with the SEPUP project, where person separation reliability for the dimension "evidence and trade-offs", corresponding to the way decision making competence was assessed in the present study, was found to be 0.74 (Briggs & Wilson, 2003). At present, more decision making tasks are being developed by the research group. The focus is on the development of easier items that represent a base level of decision making competence (cf. fig. 3). The

integration of such easier items into the test instrument will likely boost reliability indices (Boone, 2005). Fit statistics supported the developed test instrument in terms of fit validity (Wright & Linacre, 1989). Item fit statistics showed that all items exhibited a good item fit and thus can adequately measure decision making competence on one latent continuum. Analysis of person fit identified only few persons that answered to the test items in an unexpected way, which again supports the quality of the developed test instrument. If this was not the case, the use of the Partial Credit model would not be appropriate.

Analyses of Thurstonian thresholds showed that the assumed increase in difficulty (c.f. fig. 1) could be confirmed with some minor adjustments. In particular, three aspects proved to be most central for characterizing increasing person ability and item difficulty:

- An increase in ability from a non-compensatory to a mixed strategy and finally to a compensatory decision making strategy
- A general difficulty in explicitly weighing relevant decision criteria to make a choice
- A change in response quality concerning the reflection on decision making strategies from a mainly content-focused description to a mainly strategy-focused explanation

Concerning the first aspect, it was easier for students to use cut-offs, focusing only on one aspect at a time, rather than to use trade-offs and thereby to consider both positive and negative aspects for several options at the same time. A mixed strategy of using both cut-offs and trade-offs was found to be of intermediate difficulty, as such a strategy reduced the amount of information that had to be simultaneously integrated.

Results from regression analysis in this study identified a strong effect of years of education on decision making competence. Increase in person ability was significant from 6 to 8 years of education and 10 to 12 years of education, with the former showing a large effect size. Thus, the first two years at secondary school seem to be crucial for fostering decision making competence with respect to the use of decision making strategies.

Moreover, the identified hierarchy in the use of decision making strategies is also a result of the type of decision making situations that were presented in the questionnaire. The codfish, as well as the neophyte decision making situation, presented options that were all equally legitimate. Thus, a compensatory decision making strategy was identified as the most elaborate strategy. However, in other decision making situations concerning sustainable development, there may be options which are inadequate solutions to a problem and consequently, should be eliminated as a first step in the decision making process. Thus, for future research, a bigger item pool will be needed to vary the decision making situations and also to integrate situations with more options, that demand the use of a mixed strategy of using cut-offs and trade-offs.

Contributing to prior research on students' ability to explicitly weigh decision criteria or trade-offs in a decision making situation (e.g. Jimenez-Aleixandre, 2002; Kolsto, 2006; Seethaler & Linn, 2004), we found that responses to these items proved to be most difficult. Actually, no student was found who had a 50% chance of presenting an adequate answer to these items. One explanation may certainly be that students found it hard to explicitly weigh the criteria given in these decision making situations, because they represented different ecological and economical dimensions of the problem. Thus, a choice could have induced conflicts in choosing one aspect over the other. Another explanation is that individuals in general do not explicitly weigh criteria, but do this rather implicitly or give equal weight to the presented criteria (Payne et al., 1998). This might especially be the case for non-compensatory decision making strategies. Here, options are eliminated by establishing cut-offs and the order of cut-offs is likely to be influenced by their importance (ibid.). However, situations of the sustainable development of our environment are typically characterized by the existence of inherent conflicts that cannot be resolved easily. Consequently, students have to learn to develop solutions that represent a compromise between economic, ecological and socio-economic dimensions, which often includes establishing a value hierarchy. The ability to weigh decision criteria and to disclose underlying value considerations may be an elaborate way to work with these multifaceted situations.

Of particular interest to our study was students' ability to reflect on the underlying structure of decision making strategies. Prior to data analysis, we had only a vague idea of how meta-reflection might increase with respect to item difficulty. Analyses indicated that one major difference in students' answers was that they either referred to the information that was given for the different options ("content-focused description") or that they directly referred to the decision making process that was presented ("strategy-focused explanation"). Responses to the compensatory strategy showed that an explanation of the presented strategy was about 1 logit more difficult than a response that focused only on the presented content. Concerning the description of the non-compensatory strategy, results were slightly different. An answer that described the elimination process on a basic level was found to be equally difficult as a mere description of the given content information. However, an answer that showed all aspects of the decision making strategy was again found to be about 1 logit more difficult. Thus, the study indicates that the transition from a content-focused to a strategy-focused explanation can be regarded as one major aspect in increasing decision making competence. Another important step in increasing decision making competence is students' ability to identify and analyse the flaws that were integrated into the presented decision making strategies and to make a suggestion for improvement. This was particularly evident for the compensatory strategy. These findings contribute to the assessment of problem

solving in the PISA studies (OECD, 2004). Here, the ability to monitor a problem solving process and to correct possible mistakes while working on a problem is part of the most elaborated competence level. To summarize these results, we were able to support findings from prior research, that being that high school students are able to use cut-offs (Hong & Chang, 2004; Hogan, 1999) and trade-offs (Seethaler & Linn, 2004; Wilson & Sloane, 2000) in decision making situations and that students sometimes find it difficult to use trade-offs (ibid.; Jimenez-Aleixandre, 2002). The present study enabled a merging of these different findings into one analysis and moreover, to establish a hierarchy of different strategies in terms of increasing difficulty. The observed effect, attributable to years of education, allows the conclusion that decision making competence develops over years and thus, can be fostered in the science classroom. Defining distinct ability levels could help teachers to specifically foster students of different age groups. Before such levels can be defined for decision making competence, a greater variety of different decision making situations, as well as different items for meta-reflection are needed to have a precise idea of what constitutes and distinguishes competency levels.

The influence of years of education

Concerning the second research question, the developed test instrument could successfully be used across different age groups. All students, from lower secondary school to university undergraduates, were able to respond to the developed test items. This conclusion is supported by analyses for person fit, which revealed that only a few persons were either bored or overstrained by the presented items. Moreover, participants were not limited to the use of the given information, but were free to integrate additional facts or opinions, according to their individual prior knowledge, into their answers. Thus, the idea of measuring students' ability by measuring the quality of their responses to a specific item proved to be successful in the present study. The hypothesized increase in decision making competence with respect to years of education completed this picture. This result also contributes to the quality of the test instrument in terms of validity. With regard to analyses of differential item functioning, two test items proved to be disproportionately easier for upper secondary students and university undergraduates than for lower secondary students. These items (items 7 and 10) were concerned with the description of the compensatory decision making strategy and the identification of the integrated weakness in this strategy. Unsurprisingly, upper secondary students and university undergraduates received higher scores on these items relatively easier than lower secondary students. As a consequence, it would be better to analyse and to apply a Partial Credit model for the two subgroups separately.

The role of other school subjects and verbal skills

Concerning grades in other school subjects, significant, but rather low, correlations could be found between decision making competence and grades in German and Mathematics. Regarding biology grades, a significant, but again low, relationship was observed in the first study, but not in the second. These results are not surprising, given the fact that National Standards for science education in Germany were only formulated in 2005 and new state curricula are only momentarily being implemented. For the state of Lower Saxony, where the study was conducted, the new state curriculum was only implemented in 2007. Before, students' grades in biology were traditionally dominated by their performance on scientific knowledge tests. Not surprisingly, the items in the decision making questionnaire were relatively novel to students. This aspect accounts for the small relationship between grades in biology and students' performances in the present study.

The relationship between decision making competence and grade in German was assumed to be stronger. According to the National Standards for German literacy education, students should be able to present well-founded arguments for a certain topic, to form and integrate counter-arguments and to weigh these arguments to finally define and justify one's individual position. The formulation of arguments and counter-arguments can be assumed to be fairly similar to working on the pros and cons of different options in a decision making situation. A significant, but low correlation could only partly confirm this assumption. This result can partly be due to the relatively small sample size in the second study. In addition, variance in the analysed sample was only low, given the fact that only 10th graders were analysed.

Furthermore, students' verbal skills were assessed. Results showed that verbal skills had no effect on students' decision making competence. This is in contrast to the assessment of reading literacy in the PISA studies, for example. There, verbal skills as part of general cognitive abilities have an impact on students' reading competence (Deutsches PISA-Konsortium, 2001). One explanation for the lack of relationship may well be that verbal skills, as measured in this study, represented a construct relatively distant to decision making competence. As described before, students had to classify words and to find word analogies and thus, to decode the meaning of words. In contrast, the decision making competence items were constructed in a way that lower as well as upper secondary students could understand the given texts and work on the task that they were given. Difficult words or words that lower secondary students were not familiar with were avoided or reduced to a minimum. Thus, difficulties in decoding the meaning of special words were not expected. Besides, the reliability of the verbal analogies subtest was only mediocre, but confirmed the official reliability indices for this test. These results indicate that for future research, it may

be more fruitful to investigate the relationship to students' argumentation skills in more detail, because socio-scientific decision making and argumentation skills as described in the National Standards for German literacy education represent more closely connected competence areas. Therefore, a study is momentarily being conducted to shed more light on this relationship.

7.7 Conclusion

This paper aimed to unite the importance of teaching and implementing socio-scientific decision making into the science classroom with an innovative, quantitative approach to the measurement of such a competence. Although quantitative assessment has its limitations, the value and impact of socio-scientific decision making as part of science education will greatly be influenced by the outcomes of teaching interventions and consequently, outcome orientated measures. Thus, we need assessment tests that can tell us something about the learning outcomes of students' engagement with socio-scientific issues.

The study showed that using the Partial Credit model can be one successful way of measuring students' competencies in a quantitative and, at the same time, detailed way. The Partial Credit model does not only describe students' competencies with respect to "pass" or "fail", but can empirically picture different qualities of students' decision making competence. Thus, it can give a more detailed picture of students' competencies than traditional multiple choice tests. Moreover, different qualities of students' competencies were not analysed by using different test items, but by the quality of students' responses to the same open ended test items. The use of open ended items can account for the demands of decision making competence as a procedural competence more adequately than closed item formats. In contrast to most other competence assessment studies, such as the PISA studies for scientific literacy, this procedure also allowed us to use the same test instrument across different age groups. Thus, the developed test instrument makes it possible to directly compare students' performances on the same test instrument.

Moreover, a crucial aspect of using the Partial Credit model is to empirically confirm a theoretically hypothesized increase in a competence. In our study, we assumed different qualities of decision making strategies and - along with these strategies - different qualities of reflection and monitoring processes. This hypothesized increase was translated into the scoring guide that we used for analysing students' answers. The analysis of person ability, item difficulty and especially item thresholds makes it possible to judge whether such a hypothesized increase really pictures students' decision making competence. This is one of

the major advantages that Rasch modelling has in contrast to classical tests such as Likert scales for example.

In addition, the identification of different ability levels on one latent continuum, which can be visualized as a Person-Item map, is very useful to report students' abilities. It can serve as a basis of communication among researchers and teachers and it can provide a basis for a detailed, criterion-orientated description of different competency levels. If one can describe what constitutes different competence levels of decision making competence, one can then develop specific teaching interventions to foster students' competencies at different competence levels. Especially this aspect will bring quantitative assessment and the science classroom more closely together.

7.8 Appendix

Appendix A

Scenario: What can be done to stabilize the population of codfish in the Baltic Sea?				
<p>Codfish is a popular fish for consumption. However, the population of codfish in the Baltic has dramatically decreased over the last few years. One reason for this decrease is that young fish are being caught in great numbers. If no action is taken, it is likely that codfish will become extinct in the future. There are several possibilities that can be applied in order to stop this from happening</p>				
Options	Description			
1. Larger meshes in fishing nets	Meshes of fishing nets are enlarged so that the younger fish can get through. Only fish that have an adequate size for consumption will then be caught.			
2. Ceasing fishing for three years	There will not be any fishing for the coming three years.			
3. Protected areas and times	Fishermen and fishing companies are prohibited to catch any fish in specially protected areas at certain times, so that fish are able to lay their eggs.			
4. Annual adjustment of fishing quota to codfish population	The annual fishing quota will be adjusted according to the codfish population. In a good year, fishermen can catch many codfish, but in a bad year, they can only catch a small amount.			
<p>Imagine you had to decide on one of the four possible options! The following information can help you to make a choice:</p>				
Option	Larger meshes	Ceasing fishing for 3 years	Protected areas and times	Annual adjustment of fishing quota
Criteria				
Offer of codfish after implementation	high	no offer for three years	marginal	fluctuant
Offer of codfish in three years time	sufficient	sufficient	sufficient	fluctuant
Threat to codfish population	over fishing is possible	codfish population can recover completely within three years	over fishing is not possible	over fishing is not possible
Employment in the fishing industry	initially assured, insecure in case of over fishing	under threat	assured	insecure, depends on codfish population

Appendix B

To be spoilt for choice

Martina and Paul want to buy chocolate bars for a lottery at the next school party. They discuss which chocolate they would like to buy and they decide for milk chocolate. Four different kinds of milk chocolate bars are on offer. Paul and Martina read the information on the chocolate bars and this is the information that they get:

	Chocolate 1	Chocolate 2	Chocolate 3	Chocolate 4
Cultivation of cacao plants	conventional	organic	conventional	organic
Trade	conventional	fair trade	fair trade	conventional
Price	0,69 €	1,39 €	1,09 €	0,89 €
Provider	many supermarkets	Organic food stores	many supermarkets and some organic food stores	many supermarkets
Genetically modified ingredients	milk from cows that have been feed with genetically modified fodder	none	none	none

Firstly, Martina and Paul consider which criteria are the most important to them:

"A very important aspect for us is that the chocolate does not contain any genetically modified ingredients.

Another very important aspect is that the chocolate is not too expensive.

An important aspect is that the chocolate is a fair trade chocolate.

Another important aspect is that the cacao plants are being cultivated organically.

The provider is rather unimportant to us."

Task: Look at the way Paul has made his decision and then at the way Martina has made her decision:

Appendix B

Continued:

Paul's decision:

"Firstly, I considered the very important aspect of "genetically modified ingredients". Because I wanted to buy a chocolate without any genetically modified ingredients, I didn't choose chocolate number 1.

In addition, it is very important that the chocolate isn't too expensive which is why chocolate number 2 is out.

An important aspect for me is that the chocolate comes from fair trade. Therefore, I chose chocolate number 3."

	Chocolate 1	Chocolate 2	Chocolate 3	Chocolate 4
Cultivation of cacao plants			conventional	
Trade			fair trade	conventional
Price		1,39 €	1,09 €	
Provider			many supermarkets and some organic food stores	
Genetically modified ingredients	milk from cows that have been feed with genetically modified fodder		none	

Martina's decision:

„I drew up a chart in order to compare all given chocolates. I assigned 1 to 3 'smilies' (☺) per criterion per chocolate. The better the chocolate is with respect to a specific criterion, the more smilies it gets (c.f. the information on the first page). I also rated the decision criteria according to importance:

	factor / importance	Chocolate 1	Chocolate 2	Chocolate 3	Chocolate 4
Cultivation of cacao plants	2 (important)	☺ x 2 : 2	☺☺☺ x 2 : 6	☺ x 2 : 2	☺☺☺ x 2 : 6
Trade	2 (important)	☺ x 2 : 2	☺☺☺ x 2 : 6	☺☺☺ x 2 : 6	☺ x 2 : 2
Price	3 (very important)	☺☺☺ x 3 : 9	☺ x 3 : 3	☺☺ x 3 : 6	☺☺☺ x 3 : 9
Provider	1 (not important)	☺☺ x 1 : 2	☺ x 1 : 1	☺☺☺ x 1 : 3	☺☺ x 1 : 2
Genetically modified ingredients	3 (very important)	☺ x 3 : 3	☺☺☺ x 3 : 9	☺☺☺ x 3 : 9	☺☺☺ x 3 : 9
sum		18	25	26	<u>28</u>

„According to the result, I would choose chocolate no. 4. “

Appendix C

Item fit values for the 11 Items

Task	No.	item	Q-index	Zq
Codfish	1	chosen option	0.20	-0.49
	2	rejected options	0.15	0.62
	3	weighing criteria	0.24	0.47
Neophyte	4	chosen option	0.25	0.76
	5	rejected options	0.15	0.32
	6	weighing criteria	0.23	-0.03
Chocolate	7	Description: non-compensatory strategy	0.10	-1.32
	8	Description: compensatory strategy	0.10	-1.55
	9	Choice of chocolate	0.24	0.45
	10	Suggestion for improvement	0.21	-0.37
	11	Description: intuitive, justifying strategy	0.24	1.19

Note. Zq values should not deviate significantly from 0; ($|Zq| < 1.96$).

7.9 Ergänzende Ausführungen

Im Folgenden sind die Ergebnisse der im Methodenteil beschriebenen Analysen zum globalen Modell Fit dargestellt. Die Analysen anhand der Informationskriterien sowie des Bootstrapverfahren stellten die Voraussetzung für die im Ergebnisteil dargestellten Analysen im Hinblick auf das Partial Credit Modell dar.

Modell	log-likelihood	BIC	CAIC	Anzahl Parameter	Pearson*	Cressie Read*
Partial Credit Modell	-3600	7406	7440	34	0,060	0,187
LCA 2	-3583	7573	7640	67	0,167	0,487
LCA 3	-3531	7676	7777	101	0,113	0,393
Mixed Rasch 2	-3530	7479	7548	69	0,093	0,120
Mixed Rasch 3	-3529	7690	7794	104	0,140	0,380
Hybrid (1RMILCA)	-3544	7501	7569	68	0,100	0,320

*Die Pearson χ^2 - und Cressie-Read-Prüfgrößen stellen die empirischen p-Werte aus dem Bootstrapverfahren dar.

Anhand der Informationskriterien BIC und CAIC lässt sich das Partial Credit Modell als dasjenige mit dem besten Modellfit identifizieren. In Bezug auf die Pearson- und Cressie-Read-Statistik sind die Ergebnisse weniger eindeutig. Alle eingesetzten Modelle werden im Hinblick auf die 5% Signifikanzgrenze bestätigt. Ist dies der Fall, sollte das einfachste Modell, d.h. das Modell mit der geringsten Klassenanzahl, in diesem Fall das Partial Credit Model, gewählt werden (vgl. Davier, 1997).

8 Zusammenfassung und Diskussion

Im Rahmen der Zusammenfassung und Diskussion wird zunächst das im Forschungsschwerpunkt I entwickelte Modell für Bewertungskompetenz zusammengefasst (8.1). Anschließend werden die empirischen Ergebnisse der drei Studien des Forschungsschwerpunkts II resümiert und diskutiert (8.2). In einem dritten Schritt wird das theoretisch formulierte Kompetenzstrukturmodell anhand der empirischen Ergebnisse kritisch beleuchtet und erweitert (8.3). Im abschließenden Kapitel der Dissertation wird ein Fazit gezogen und ein Ausblick sowohl auf zukünftige Forschungsperspektiven als auch auf die Relevanz der vorliegenden Ergebnisse für eine Förderung von Bewertungskompetenz im Biologieunterricht gegeben (Kap. 9).

8.1 Forschungsschwerpunkt I

Ziel des Forschungsschwerpunkts I war es, den Begriff der Bewertungskompetenz zu präzisieren und auf Basis fachdidaktischer sowie psychologischer Forschung zentrale Strukturen für ein Kompetenzstrukturmodell zu identifizieren. Unter Berücksichtigung kompetenzdiagnostischer Anforderungen für eine anschließende Operationalisierung wurden vier Teilkompetenzen identifiziert, die sich zumindest theoretisch sinnvoll voneinander trennen lassen.

Zusammenfassend bezeichnet Bewertungskompetenz die Fähigkeit von Schülerinnen und Schülern, sich in komplexen Bewertungs- und Entscheidungssituationen begründet und systematisch bei unterschiedlichen Handlungsoptionen zu entscheiden, um kompetent am gesellschaftlichen Diskurs um die Gestaltung von Nachhaltiger Entwicklung teilhaben zu können (vgl. Bayrhuber et al., 2007).

Das entwickelte Kompetenzstrukturmodell beschreibt zwei prozedurale und zwei konzeptuelle Teilkompetenzen, die für die Bewältigung von Bewertungs- und Entscheidungssituationen mit mehreren Handlungsoptionen relevant sind. Zu den prozeduralen Kompetenzen gehört die Fähigkeit, Handlungsoptionen in einer

Entscheidungssituation zu entwickeln, wobei Informationssuch- und -verarbeitungsprozesse notwendig sind. Diese Informationen müssen im Hinblick auf ihre Adäquatheit und Relevanz reflektiert werden (Teilkompetenz „Sachinformationen generieren und reflektieren“). Die erarbeiteten Handlungsoptionen werden dann in einem zweiten Schritt unter Anwendung einer Entscheidungsstrategie miteinander verglichen. Der Entscheidungsprozess sowie die getroffene Entscheidung müssen wiederum reflektiert werden (Teilkompetenz „Bewerten, Entscheiden und Reflektieren“). Beide prozeduralen Teilkompetenzen fassen mit Bewertungs- und Entscheidungssituationen den Bereich der Domäne sehr weit und stellen, vergleichbar den Kompetenzen beim Problemlösen in den PISA Studien (vgl. Kap. 2), eher fächerübergreifende Kompetenzen dar.

Für Gestaltungssituationen Nachhaltiger Entwicklung sind neben biologischem Fachwissen die zwei konzeptuellen Teilkompetenzen des Kompetenzstrukturmodells relevant. Dazu gehört zum einen Wissen darüber, was das Konzept Nachhaltiger Entwicklung im Kern ausmacht (Teilkompetenz „Kennen und Verstehen von Nachhaltiger Entwicklung“) sowie zum anderen ein Wissen darüber, welche persönlichen und gesellschaftlichen Werte und Normen in einer bestimmten Gestaltungssituationen berührt und im Entscheidungsprozess berücksichtigt werden können (Teilkompetenz „Kennen und Verstehen von Werten und Normen“).

Neben der Beschreibung dieser vier Teilkompetenzen wurden für die im Fokus des Forschungsschwerpunkts II stehende Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ mögliche Graduierungen postuliert, d.h. Kompetenzentwicklungsverläufe in Form von Kompetenzniveaus formuliert. Eine steigende Kompetenz zeichnet sich durch eine Anwendung elaborierter Entscheidungsstrategien aus. Des Weiteren können auf höheren Kompetenzniveaus Bewertungs- und Entscheidungsprozesse auf Strategieebene reflektiert, Fehler analysiert und korrigiert werden bzw. die Grenzen in der Anwendung von Entscheidungsstrategien erkannt werden (vgl. Kap. 5).

8.2 Forschungsschwerpunkt II

Ziel der drei empirischen Studien des Forschungsschwerpunkts II war es, die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ zu operationalisieren und ein reliables und valides Messinstrument für die Anwendung von Entscheidungsstrategien in Bewertungs- und Entscheidungssituationen Nachhaltiger Entwicklung zu entwickeln.

Studie 1

Die Studie 1 (n=291) hatte v.a. explorativen Charakter. Zunächst sollte geklärt werden, inwiefern sich spontanes Entscheidungsverhalten sowie non-kompensatorische oder kompensatorische Entscheidungsstrategien bei der Bearbeitung von Bewertungs- und Entscheidungssituationen mit unterschiedlichen, aber gleichwertigen, Handlungsoptionen beobachten lassen. Aufgrund der sehr breit gefassten Domäne für „Bewerten, Entscheiden und Reflektieren“ wurden dabei biologische als auch nicht biologische Themen für die Entscheidungssituationen gewählt, um den Einfluss des Kontextes zu untersuchen. Außerdem wurde die Darstellungsweise der Bewertungs- und Entscheidungssituationen variiert, um Aufschlüsse über die Eignung verschiedener offener Aufgabenformate zu erhalten. Im Bezug auf die psychometrische Modellierung kamen, aufgrund der theoretischen Annahme unterschiedlicher Entscheidungsstrategien, neben dem für kompetenzdiagnostische Studien üblichen ein- oder mehrdimensionalen Raschmodellen, weitere IRT-Modelle zum Einsatz.

Zentrale Forschungsfragen der ersten empirischen Studie waren:

- Welche unterschiedlichen Entscheidungsstrategien lassen sich bei Schülerinnen und Schülern sowie Studierenden in der Bearbeitung von Entscheidungssituationen beobachten?
- Inwiefern beeinflussen der Kontext und die Darstellungsweise einer Entscheidungssituation die Auswahl einer Entscheidungsstrategie?
- Welches IRT-Modell beschreibt das Antwortverhalten in Bezug auf die Anwendung von Entscheidungsstrategien am besten?

In Bezug auf die erste Forschungsfrage konnten bei den untersuchten Probandinnen und Probanden drei Klassen³⁵ von Entscheidungsstrategien identifiziert werden: eine non-kompensatorische und eine kompensatorische Entscheidungsstrategie sowie eine Strategie, die sich durch ein explizites Gewichten der gegebenen Entscheidungskriterien anhand von persönlichen Präferenzen auszeichnete (im Folgenden „Gewichter“ genannt)³⁶. Die non-kompensatorische Entscheidungsstrategie machte mit 44,4 % den größten Anteil der

35 Der Begriff Klasse wird hier im Sinne von probabilistischen Klassen im Rahmen des Drei-Klassen-Hybridmodells verstanden und nicht im Sinne von Schulklassen.

36 Außerdem wurden weitere Personen mit spontanem Entscheidungsverhalten identifiziert, die lediglich ihre Entscheidung dokumentierten, ihren Entscheidungsprozess jedoch nicht. Da diese Personen auf allen Items mit Null scorten, konnte sie nicht in die probabilistische Auswertung integriert werden.

Stichprobe aus. Charakteristisch für diese Strategie ist ein Anlegen von Schwellenwerten, die eine Option mindestens erreichen muss. Schneidet eine Option im Hinblick auf ein Entscheidungskriterium zu schlecht ab, wird diese eliminiert (→ Anwendung von *cut-offs*). Dieses Ausschlussverfahren wurde solange angewandt bis letztendlich nur noch eine Option übrig blieb, die dann gewählt wurde. Diese Entscheidungsstrategie wird in der Entscheidungstheorie auch als *elimination-by-aspects* Prozess beschrieben (vgl. u.a Tversky, 1972). Individuen, die eine derartige Entscheidungsstrategie anwenden, werden auch als *Satisficer* bezeichnet, da sie eine zufriedenstellende, aber nicht optimale, Option auswählen (vgl. u.a. Payne et al., 1998). Da bei dieser Entscheidungsstrategie immer nur ein Entscheidungskriterium pro Zeitpunkt bearbeitet wird, ist die Intensität der Informationsverarbeitung vergleichsweise gering.

26,5% der Stichprobe verwendeten eine kompensatorische Entscheidungsstrategie. Diese Strategie zeichnete sich durch ein Anwenden von *trade-offs* aus. Mitglieder dieser Klasse konnten die gegebenen Optionen in den präsentierten Entscheidungssituationen miteinander vergleichen und dabei die Vor- und Nachteile gegeneinander abwägen. Bei dieser Entscheidungsstrategie werden mehrere Informationen zum gleichen Zeitpunkt verarbeitet. Deswegen liegt die Intensität der Informationsverarbeitung bei dieser Strategie im Vergleich zur non-kompensatorische Strategie höher. Individuen, die eine kompensatorische Strategie anwenden und somit die im Vergleich aller Optionen beste auswählen, werden in der Entscheidungstheorie im Gegensatz zu *Satisficern* als *Optimizer* bezeichnet (vgl. u.a. Payne et al., 1998).

Die dritte Klasse der Gewichter (29,1%) war dadurch charakterisiert, dass Probandinnen und Probanden in dieser Klasse weder *trade-offs* noch *cut-offs* anwendeten. Vielmehr konzentrierten sie sich darauf, die gegebenen Entscheidungskriterien mit Hilfe einer vorgegebenen Gewichtungstabelle anhand ihrer persönlichen Präferenzen explizit zu gewichten und dadurch in eine Hierarchie von „weniger wichtig“ bis „sehr wichtig“ zu bringen. Diese Gewichtung wurde dann im Folgenden jedoch nicht als Basis für ein Abwägen der Vor- und Nachteile von Optionen oder ein Ausschließen von Optionen verwendet. Vielmehr folgte auf die Gewichtung sofort die Auswahl einer Option.

Besonders die Klasse der Gewichter machte den Einfluss unterschiedlicher Darstellungsweisen in Bezug auf das Aufgabenformat der Entscheidungssituationen deutlich. Während die Probandinnen und Probanden in der Klassenfahrtsituation nicht explizit aufgefordert wurden, die gegebenen Entscheidungskriterien nach der persönlichen Präferenz zu ordnen, mussten sie diese Gewichtung in der Sportauswahl und in der Neophytenaufgabe durch Ausfüllen einer Gewichtungstabelle vornehmen. Diese als Unterstützung gedachte Tabelle führte jedoch dazu, dass die Individuen in dieser Klasse ausschließlich gewichteten,

diese vorgenommene Gewichtung jedoch anschließend nicht in Form eines weiterführenden Entscheidungsprozesses dokumentierten.

Im Vergleich zu den Gewichtern wurden die Klassen der kompensatorischen sowie der non-kompensatorischen Entscheidungsstrategien nicht bzw. nur gering von der Darstellungsweise der Entscheidungssituationen beeinflusst. So wendete die Klasse der kompensatorischen Entscheider diese Strategie über alle drei Kontexte hin an. Bei der Klasse der non-kompensatorischen Entscheider konnte für die Klassenfahrtsituation ein Einfluss der Darstellungsweise identifiziert werden. Während die Individuen in dieser Klasse in der zweiten und dritten Entscheidungssituation ausschließlich *cut-offs* anwendeten, wendeten sie in der Klassenfahrtsaufgabe eine Mischstrategie aus *trade-offs* und *cut-offs* an. Diese Mischstrategie ist zurückzuführen auf die Auswahl der präsentierten Optionen, da diese sich in zwei kulturelle und zwei sportliche Angebote unterscheiden ließen. Je nachdem, welche Präferenz bei den Probandinnen und Probanden vorlag, wurden zunächst zwei Optionen aufgrund des unterwünschten Freizeitangebots ausgeschlossen. In einem zweiten Schritt wurden dann die beiden verbliebenen Optionen durch ein Abwägen der Vor- und Nachteile miteinander verglichen³⁷. Der Einfluss des Kontextes wurde zudem in der Entscheidungssituation Sportauswahl deutlich. In der Klasse der non-kompensatorischen Entscheider wurde vielfach das Kriterium Spaß in den Entscheidungsprozess integriert und Optionen anhand dieses Kriteriums eliminiert. Eine generelle Abneigung gegen Sport konnte auch dazu führen, dass die Entscheidungssituationen nicht als solche interpretiert und somit kein Entscheidungsprozess initialisiert wurde.

In Bezug auf die psychometrische Modellierung konnte ein Drei-Klassen-Hybridmodell das Antwortverhalten am besten modellieren. Das Hybridmodell wurde dabei im Rahmen der Modellgeltungstests sowohl durch die Informationskriterien BIC und CAIC als auch durch das Bootstrapverfahren als das beste Modell identifiziert. Die mittlere Zuordnungswahrscheinlichkeit der Individuen zu den drei Klassen lag bei 85% und 88%. Die Itemfit Analysen für die beiden Raschklassen zeigten, dass lediglich 1 Item einen problematischen Fit aufwies, was darauf zurückgeführt werden konnte, dass alle Individuen in dieser Klasse das Item lösen konnten. Personfit Analysen identifizierten 13 Personen mit einem unerwarteten Antwortmuster, die für die Berechnung der berichteten Ergebnisse herausgenommen wurden.

³⁷ Eine derartige Mischstrategie wird oftmals auch dann angewendet, wenn die Komplexität einer Entscheidungssituation, d.h. entweder die Anzahl der Entscheidungsoptionen oder -kriterien zu hoch ist und die gegebenen Informationen dementsprechend nicht gleichzeitig bearbeitet werden können (Jungermann et al., 2004).

Studien 2 und 3

Die Studien 2 (n=448) und 3 (n=83) dienten der Weiterentwicklung und Optimierung des Testinstruments. Dabei standen zwei Aspekte im Zentrum. Zunächst wurden die in der Studie 1 identifizierten Entscheidungsstrategien im Hinblick auf ihren Komplexitäts- und Elaborationsgrad unter Berücksichtigung der im Kompetenzstrukturmodell postulierten Niveaustufen hierarchisiert (vgl. Kap. 5, Tab. 5.4). Des Weiteren wurde das Testinstrument um Aufgaben zur Reflexion über Bewertungs- und Entscheidungsprozesse erweitert, um damit alle Aspekte der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ messbar zu machen. Im Zentrum der dritten Studie standen vertiefende Analysen hinsichtlich der Validität des Testinstruments.

Die für die Studien 2 und 3 zentralen Forschungsfragen waren:

- Kann die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ durch ein eindimensionales polytomes Raschmodell modelliert und damit die Anwendung von und Reflexion über Entscheidungsstrategien auf einem latenten Kontinuum dargestellt werden?
- Inwiefern genügt das entwickelte Testinstrument den Gütekriterien der Reliabilität und Validität?

Die Analysen der verschiedenen eingesetzten IRT-Modelle zeigten, dass das polytome Raschmodell (Partial Credit Modell) das Antwortverhalten im Bezug auf die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ am besten erklären kann. Ein Vergleich der informationstheoretischen Maße BIC sowie CAIC konnte zeigen, dass alle anderen IRT-Modelle das Antwortverhalten schlechter modellierten. Diese Ergebnisse konnten mit Einschränkung durch die Ergebnisse des Bootstrapverfahrens bestätigt werden. Das Bootstrapverfahren identifizierte zwar mehrere Modelle, die auf die Daten passten, jedoch war im Vergleich dieser Modelle das Partial Credit Modell das sparsamste und wurde aus diesem Grund als das relativ beste Modell identifiziert (vgl. von Davier & Rost, 1997).

Das Partial Credit Modell wurde dann im Weiteren im Hinblick auf Reliabilität und Validität genauer untersucht. Personfit Analysen zeigten, dass lediglich 11 Personen ein modellinkonformes Antwortverhalten zeigten und somit aus den Analysen herausgenommen werden mussten. Typischerweise bearbeiteten diese Personen nur die Reflexionsaufgaben nicht aber die Aufgaben, in denen sie selbst eine Entscheidung treffen mussten. Eine derart geringe Anzahl modellinkonformen Antwortverhaltens stärkt die Güte der entwickelten Testitems. Analog zu den Person-fit Indices wurden Item-fit Indices analysiert. Die Item-Q-Indices sowie Zq-Werte und p-Werte zeigten, dass die Items nicht signifikant von dem unter

Annahme des Partial Credit Modells angenommenen Antwortmusters abwichen. Die Raschreliabilität des Tests lag mit 0.74 in einem akzeptablen Rahmen. Die Analysen der Schwellenparameter zeigten, dass die Antwortkategorien einiger Items nicht in ihrer Schwierigkeit anstiegen und somit für einige Items zwei Antwortkategorien zusammengeschmolzen werden mussten. Danach waren alle Schwellenparameter in ihrer Schwierigkeit geordnet.

Des Weiteren wurde die untersuchte Stichprobe auf Personenhomogenität hin untersucht. Das Partial Credit Modell konnte auch für die einzelnen Klassenstufen als das beste Modell identifiziert werden. Pearson Korrelationen zeigten hohe Zusammenhänge zwischen den mittleren Personenfähigkeiten der einzelnen Klassenstufen. Um darüber hinaus genauere Aussagen über ein mögliches unterschiedliches Antwortverhalten in den Subgruppen treffen zu können, wurde die Stichprobe in Schülerinnen und Schüler der Sekundarstufe I und Schülerinnen und Schüler der Sekundarstufe II sowie Studierende geteilt. Zwischen diesen beiden Subgruppen ist ein heterogenes Antwortverhalten am wahrscheinlichsten. Ein Grafischer Modelltest der Itemparameter dieser beiden Subgruppen konnte dabei zwei Items identifizieren, die in den beiden Subgruppen unterschiedlich schwierig waren. Es handelt sich dabei um zwei Items, die die Reflexionsfähigkeit über Bewertungsprozesse abtesten. Dabei war das Erreichen höherer Scores in Bezug auf diese Items für die Schülerinnen und Schüler der Sekundarstufe II sowie für Studierende relativ leichter als für Schülerinnen und Schüler der Sekundarstufe I. Die Antworten der Probandinnen und Probanden der Sekundarstufe I in Bezug auf die Reflexionsaufgaben waren dabei v.a. durch eine Reflexion auf Inhaltsebene charakterisiert, während die Antworten der Probandinnen und Probanden der Sekundarstufe II v.a. durch eine Reflexion auf Strategieebene charakterisiert waren.

Als Konsequenz der unterschiedlichen Itemschwierigkeiten sollte eine psychometrische Modellierung zukünftig für die Sekundarstufe I und die Sekundarstufe II bzw. Studierende getrennt vorgenommen werden. Für die beschriebene Studie 2 wurden die Analysen dennoch für die gesamte Stichprobe vorgenommen, da a) die Gesamtstichprobe ($n=370$) für Raschmodellierungen relativ gering war und b) Analysen nach Klassenstufen getrennt andere Schwierigkeiten, wie beispielsweise höhere Fehler für die Personen- und Itemschätzer hervorgerufen hätten.

Hinweise auf die Validität des Testinstruments lieferten zum einen die bereits beschriebenen Personfit-Analysen. Zum anderen konnte ein starker Effekt der Schulzeit auf die mittlere Personenfähigkeit im Hinblick auf „Bewerten, Entscheiden und Reflektieren“ die Validität des Testinstruments bestärken. Des Weiteren konnten in der Studie 3 Zusammenhänge zwischen Bewertungskompetenz und Schulleistungen in den Fächern Deutsch und Mathematik nachgewiesen werden. Ein Zusammenhang besonders zur Deutschnote war

insofern wahrscheinlich, als dass Argumentationsfähigkeit und die Anwendung von Bewertungs- und Entscheidungsstrategien ähnliche Kompetenzkonzepte repräsentieren. Ein Zusammenhang zur Biologienote konnte in der Studie 2, nicht jedoch in der Studie 3 nachgewiesen werden. Auch der Zusammenhang in der Studie 2 war sehr gering. Diese Ergebnisse waren insofern zu erwarten, als dass das entwickelte Testinstrument keine für den Biologieunterricht typischen Kompetenzbereiche abtestete. Der nicht vorhandene Zusammenhang in der Studie 3 ist darüber hinaus durch die geringe Stichprobengröße sowie die Zusammensetzung der Stichprobe zu erklären.

8.3 Kritische Betrachtung des theoretischen Modells auf Basis der empirischen Ergebnisse

Im Folgenden sollen die für die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ in Kapitel 5 postulierten Niveaustufen im Hinblick auf die empirischen Ergebnisse kritisch beleuchtet und ggf. erweitert werden. Dabei sollen die folgenden Aspekte diskutiert werden: die Graduierung der angewendeten Entscheidungsstrategien, die Anzahl der berücksichtigten Optionen und Entscheidungskriterien, die Fähigkeit zur Reflexion über Bewertungs- und Entscheidungsprozesse sowie die Fähigkeit, Entscheidungskriterien explizit zu gewichten.

Im Hinblick auf die Graduierung der angewendeten Entscheidungsstrategien konnte die theoretisch angenommene Struktur zu großen Teilen bestätigt werden. Erweiterungen der theoretischen Annahmen resultieren v.a. durch eine optimierte Vorgehensweise in der Auswertung der Schülerinnen- und Schülerantworten in den Studien 2 und 3 im Vergleich zur Studie 1. So wurde die Auswertung im Hinblick auf die Anwendung einer Entscheidungsstrategie nicht mehr über alle Entscheidungsoptionen integrierend vorgenommen, sondern für die gewählte Option und die nicht gewählten Optionen getrennt. Für die *nicht gewählten Optionen* konnte ein Anstieg in der Schwierigkeit der angewendeten Entscheidungsstrategien von intuitivem bzw. rechtfertigendem Entscheidungsverhalten, über eine non-kompensatorische, zu einer Mischstrategie hin zu einer kompensatorischen Entscheidungsstrategie durch das Partial Credit Modell abgebildet werden. Die Antwortkategorien, welche die verschiedenen Entscheidungsstrategien repräsentierten, waren in ihrer Schwierigkeit geordnet. Der Anstieg in der Itemschwierigkeit betrug dabei jeweils 1 logit und mehr (vgl. Kap. 7, Abb. 7.3). Für den Umgang mit der *gewählten Option* konnten zwei Antwortqualitäten voneinander getrennt werden. Ein basales Niveau stellte intuitives bzw. rechtfertigendes Entscheidungsverhalten dar. Probandinnen und Probanden, die intuitiv entschieden, dokumentierten lediglich welche Option sie ausgewählt hatten.

Rechtfertigende Antworten konnten auf demselben basalen Niveau verortet werden. Typischerweise wurde in diesen Antworten zunächst die ausgewählte Option benannt und anschließend positive Aspekte zur Bestätigung bzw. Rechtfertigung angeführt. Negative Aspekte für die gewählte Option wurden ignoriert. Diese Antworten bestätigen das von Haidt im *social intuitionist model* postulierte rechtfertigende Entscheidungsverhalten (Haidt, 2001). Beide Entscheidungsstrategien wurden auf demselben Niveau verortet, da die Schwellenparameter im Hinblick auf diese beiden Antwortkategorien („nichts“ oder „nur positive Aspekte benannt“) nicht geordnet waren. Von diesem basalen Niveau konnte dann aber eine kompensatorische Vorgehensweise, d.h. die Benennung positiver und negativer Aspekte für die gewählte Option, unterschieden werden.

Ein Anstieg in der Anzahl der bearbeiteten Entscheidungskriterien, wie im theoretischen Modell formuliert, war empirisch nicht umzusetzen. Dies lag v.a. daran, dass eine Verarbeitung von Kriterien sowohl inhaltlich als auch statistisch mit der Anwendung einer Entscheidungsstrategie im Hinblick auf die *nicht gewählten Optionen* zusammenhing. Somit konnte das Item „Anzahl der bearbeiteten Kriterien“ nicht in die Raschmodellierung integriert werden. Neben der Verletzung statistischer Voraussetzungen, sind darüber hinaus auch inhaltliche Aspekte anzuführen, die einer Berücksichtigung entgegenstehen. So kann eine non-kompensatorische Entscheidungsstrategie unter Berücksichtigung lediglich eines Kriteriums angewendet werden und in einer Entscheidung münden. Eine Mischstrategie aus einer non-kompensatorischen und kompensatorischen Vorgehensweise erfordert jedoch die Berücksichtigung von mindestens zwei Kriterien. Auch eine kompensatorische Entscheidungsstrategie erfordert bei den zum Einsatz gekommenen Entscheidungssituationen eine Berücksichtigung von mindestens zwei Kriterien. In der Regel wurden bei einer kompensatorischen Vorgehensweise - bedingt durch das Aufgabenformat - mehr als zwei Kriterien berücksichtigt. Zusammenfassend lässt sich somit feststellen, dass diese beiden Items sowohl statistisch als auch inhaltlich einander bedingen und somit letztendlich redundante Informationen im Hinblick auf die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ liefern würden.³⁸ Im Hinblick auf die Fähigkeit zur Reflexion über Bewertungs- und Entscheidungsprozesse konnten die theoretischen Annahmen bestätigt und darüber hinaus erweitert werden. Im theoretischen Modell wurde die Fähigkeit zur Reflexion, darunter die Fähigkeit Fehler in Entscheidungsprozessen zu identifizieren, auf den Niveaus 3 und 4 verortet. Auch die empirischen Ergebnisse der Raschskalierung zeigten, dass die Identifikation von Fehlern in den Reflexionsaufgaben eine hohe Schwierigkeit darstellte. Darüber hinaus waren Antworten, die Vorschläge für eine

38 Stochastische Abhängigkeiten ließen sich auch für weitere Items des ursprünglichen Scoring Guides für die Studien 2 und 3 finden. Diese Items mussten in gleicher Weise reduziert werden.

Verbesserung von Entscheidungsstrategien gaben, für die Probandinnen und Probanden im Bezug auf die Reflexionsaufgaben am schwierigsten. Diese beiden Aspekte sollten für eine Präzisierung der Kompetenzniveaus 3 und 4 herangezogen werden. Des Weiteren konnten die Ergebnisse der Raschskalierung zeigen, dass sich eine elaborierte Reflexionsfähigkeit durch eine Reflexion auf *Strategieebene* auszeichnet, während sich eine basale Reflexionsfähigkeit durch eine Reflexion auf *Inhaltsebene* auszeichnet und damit eher eine Beschreibung als eine Reflexion darstellt. Derartige inhaltsbezogene Beschreibungen von Reflexion sollten in die Beschreibung des basalen und mittleren Kompetenzniveaus integriert werden.

Die Fähigkeit Entscheidungskriterien explizit zu gewichten wurde in der Studie 1 zwar von 29,1% der Stichprobe gezeigt, jedoch war dies ein Resultat des Aufgabenformats. In den Studien 2 und 3 stellte diese Fähigkeit die höchste Schwierigkeit aller Items dar, da keine Unterstützung in der Gewichtung gegeben wurde. Keine Probandin und kein Proband lösten mit einer 50 % Wahrscheinlichkeit diese Items. Dies hat mehrere Ursachen. Zum einen konnte Forschung im Rahmen der Entscheidungstheorie zeigen, dass Individuen in der Regel Entscheidungskriterien nicht gewichten oder ihnen eine gleiche Wichtigkeit zuordnen (vgl. u.a. Gigerenzer et al., 1999; Jungermann et al., 2004). Zum anderen werden Gewichtungen oftmals auch implizit vorgenommen, aber nicht dokumentiert. Dies ist insbesondere bei non-kompensatorischen Entscheidungsstrategien der Fall, wo die Reihenfolge von Ausschlusskriterien von ihrer (subjektiven) Wichtigkeit beeinflusst wird. Die Schlussfolgerung zu ziehen, dass ein explizites Gewichten aus zukünftigen Modellierungen entfernt werden sollte, ist aber zu voreilig. Besonders Entscheidungssituationen Nachhaltiger Entwicklung sind durch Konflikte zwischen den ökologischen, ökonomischen und sozialen Dimensionen gekennzeichnet. Diese Konflikte können in der Regel nur gelöst werden, indem ein Aspekt dem anderen untergeordnet wird. In solchen Situationen wird explizites Gewichten von Entscheidungskriterien der drei Dimensionen notwendig.

9 Fazit und Ausblick

Die vorliegende Arbeit hatte aus Forschungsperspektive zunächst das Ziel, ein Modell für Bewertungskompetenz theoretisch herzuleiten, d.h. zentrale Strukturen von Bewertungskompetenz zu identifizieren. Ziel des Forschungsschwerpunkts II war die empirische Überprüfung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“. Neben dieser Forschungsperspektive hatte die Projektarbeit im Rahmen des Gesamtprojekts „Biologie im Kontext“ auch das Ziel, Schulentwicklungsprozesse in Gang zu setzen und dabei Lehrkräfte in der Integration von Bewertungskompetenz im Biologieunterricht zu unterstützen. Aus diesem Grund soll im Folgenden sowohl ein Fazit und Ausblick auf Forschungsebene als auch auf Unterrichtsentwicklungsebene gegeben werden.

Forschungsperspektive

Im Zentrum der Forschungsarbeit stand die Operationalisierung der Teilkompetenz „Bewerten, Entscheiden und Reflektieren“. Die empirischen Ergebnisse konnten zeigen, dass die entwickelten Testitems raschskalierbar sind und dass das entwickelte Testinstrument hinreichend reliabel und valide ist. Außerdem konnte gezeigt werden, dass ein offenes Aufgabenformat geeignet ist für die Messung. Es gibt den Probandinnen und Probanden die Möglichkeit, je nach Vorwissen sowie individuellen Einstellungen und Wünschen, ihren Bewertungs- und Entscheidungsprozess darzulegen. Im Gegensatz zu anderen kompetenzdiagnostischen Studien, wo oftmals auf eine Klassenstufe fokussiert wird, kann das entwickelte Testinstrument in mehreren Jahrgangsstufen eingesetzt werden. Das Testinstrument stellt einen ersten Ansatz dar, Bewertungskompetenz mit Hilfe eines Modells der Item-Response-Theorie zu modellieren. Es zeigt Möglichkeiten auf, wie Bewertungskompetenz von Schülerinnen und Schülern quantitativ erhoben werden kann, und leistet damit einen wichtigen Beitrag zur Umsetzung und Überprüfung der im Jahr 2004 verabschiedeten Bildungsstandards für den mittleren Schulabschluss. Dies ist vor allem unter dem Gesichtspunkt zu sehen, dass das nationale Assessment für die Bildungsstandards in den naturwissenschaftlichen Fächern nur auf die Kompetenzbereiche Fachwissen und

Erkenntnisgewinnung fokussieren wird. Gerade aus diesem Grund ist es notwendig, auch für die Kompetenzbereiche Bewertung und Kommunikation, Möglichkeiten für quantitative Erhebungen zu erarbeiten.

Dennoch ist anzumerken, dass das vorgestellte Testinstrument zukünftig weiter optimiert werden muss. Zentrale Aspekte einer Optimierung sind v.a. weiterführende Untersuchungen im Hinblick auf die Validität sowie eine Vergrößerung des Itempools - und damit auch Verlängerung des Testinstruments - zur Verbesserung der Reliabilität. Im Hinblick auf eine Stärkung der externen Validität sollte untersucht werden, inwiefern „Bewerten, Entscheiden und Reflektieren“ mit der Kompetenz des Argumentativen Schreibens im Deutschunterricht zusammenhängt (vgl. Kap. 7). Vergleicht man die Bildungsstandards für das Fach Deutsch mit der Operationalisierung von „Bewerten, Entscheiden und Reflektieren“ in dieser Arbeit, sind positive Zusammenhänge zu vermuten. Des Weiteren wäre in Zukunft ein gemeinsamer Einsatz des entwickelten Testinstruments mit Testaufgaben des SEPUP Projekts zu überlegen, um die externe Validität zu erhöhen. Einschränkend muss jedoch geprüft werden, inwiefern Barrieren in einem Einsatz dieser Aufgaben bestehen, da diese Aufgaben zum einen stärker vorwissenabhängig sind und zum anderen kulturelle Besonderheiten aufweisen. Neben der Optimierung des entwickelten Testinstruments stellt sich zukünftig die Frage, wie die verbleibenden drei Teilkompetenzen operationalisiert und wie eine gemeinsame Modellierung aller Teilkompetenzen gestaltet werden könnte. Da zu Beginn dieser Arbeit noch keine kompetenzorientierten Testaufgaben für Bewertungskompetenz vorlagen, fokussierte diese Arbeit auf die Operationalisierung einer Teilkompetenz. Für den empirischen Zugang wurde die Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ ausgewählt, die bislang im Rahmen von Bildung für eine Nachhaltige Entwicklung und Biologieunterricht in Deutschland in der Breite kaum Berücksichtigung findet, jedoch zentral für die Lösung von Problemsituationen moderner Naturwissenschaften ist.

Daraus folgt, dass im Hinblick auf die psychometrische Modellierung ein *between item dimensionality* Ansatz gewählt wurde, d.h. dass die entwickelten Items ausschließlich auf eine Teilkompetenz, d.h. eine latente Dimension, laden. Zukünftig sollen Testitems jeweils getrennt für die anderen Teilkompetenzen entwickelt werden. Längerfristiges Ziel muss es sein, alle Aufgaben in einem Testinstrument gemeinsam einzusetzen und eine psychometrische Modellierung nach dem *between item dimensionality* Ansatz durchzuführen. Dies ist, nach Meinung der Autorin, eine angemessene und v.a. praktikable Lösung in Anbetracht der vorhandenen Testlage im Bezug auf Bewertungskompetenz. Aus den genannten Gründen ist ein *within item multidimensionality* Ansatz in Bezug auf Bewertungskompetenz nach diesem Entwicklungsstand des Kompetenzbereichs nicht möglich gewesen. Es bleibt darüber hinaus abzuwarten, inwiefern hohe positive

Zusammenhänge zwischen den postulierten Teilkompetenzen gemessen werden, die es dann notwendig machen könnten, die bisher im theoretischen Modell getrennten Teilkompetenzen empirisch zusammenzufassen.

Unterrichtsperspektive

Das Gesamtprojekt Biologie im Kontext sowie das Göttinger Teilprojekt haben drei Schwerpunkte im Hinblick auf Unterrichtsentwicklung: Kompetenzorientierung, Kontextorientierung und Unterstützung von Lehrkräften (Bayrhuber et al., 2007). In Anlehnung an Chemie im Kontext wird nach der Philosophie des symbiotischen Implementationsansatzes (Gräsel & Parchmann, 2004) bei *bik* in Lerngemeinschaften, *learning communities* (Brown, 1997), aus Lehrkräften und Fachdidaktikern bzw. Fachdidaktikerinnen verschiedener Universitäten zusammengearbeitet. Ziel dieser Lerngemeinschaften ist es, auf Basis der Bildungsstandards für die naturwissenschaftlichen Fächer neue Unterrichtskonzeptionen für einen outputorientierten, kompetenzfördernden Unterricht zu entwickeln, sowie Lehrkräfte in der Umsetzung dieser neuen Konzepte in ihrem Unterricht zu unterstützen. Diese Lerngemeinschaften stellen die Basis für kontinuierliche Entwicklungs- und Austauschprozesse zwischen Lehrkräften und Fachdidaktikern bzw. Fachdidaktikerinnen, und somit zwischen Unterrichtspraxis und fachdidaktischer Forschung dar (Bayrhuber et al., 2007).

Im Rahmen der vorliegenden Forschungsarbeit lassen sich besonders drei dieser Entwicklungs- und Austauschprozesse feststellen. Dies ist zum einen die Entwicklung von Unterrichtsmaterialien zur Förderung von Bewertungskompetenz am Beispiel von Entscheidungssituationen Nachhaltiger Entwicklung, zum anderen der Einsatz des in den empirischen Studien entwickelten Testinstruments zur Evaluation von Biologieunterricht sowie eine Dissemination der Ergebnisse der Lerngemeinschaften z.B. in Form von Lehrerfortbildungen bzw. zur Entwicklung neuer Curricula.

Im Rahmen des ersten Aspekts wurden zahlreiche Unterrichtsmaterialien entwickelt, die die im theoretischen Modell der Bewertungskompetenz formulierten Teilkompetenzen umsetzen. Eine typische Lernaufgabe für den Biologieunterricht stellt die Entscheidungssituation „Mühlenbach“ dar³⁹. In dieser Aufgabe, die den Schwerpunkt auf der Förderung von „Bewerten, Entscheiden und Reflektieren“ hat, müssen Schülerinnen und Schüler den ökologischen Zustand eines Fließgewässers beurteilen. In einem zweiten Schritt

39 Eggert, S., Barfod-Werner, I. & Bögeholz, S. (2008). Entscheidungen treffen – wie man vorgehen kann. In U. Harms (Hrsg.). Unterricht Biologie kompakt. Fächerübergreifend unterrichten. Jg. 32, Heft 336, S. 13-18.

müssen sie mehrere Entscheidungsoptionen unter ökologischen, ökonomischen und sozialen Aspekten miteinander vergleichen und eine Option für die Entwicklung des Mühlenbachs auswählen. Sie erhalten dabei ein Bewertungssystem zur Unterstützung eines systematischen Vergleichs von Entscheidungsoptionen, welches die Anwendung einer kompensatorischen Entscheidungsstrategie fördert. Weitere bereits entwickelte Lernaufgaben sind so konstruiert, dass Schülerinnen und Schüler eine Mischstrategie aus einer non-kompensatorischen und einer kompensatorischen Entscheidungsstrategie anwenden müssen, da von den präsentierten Entscheidungsoptionen eine aufgrund schlechter ökologischer Ausprägungen ausgeschlossen werden muss⁴⁰. Ziel der Entwicklung unterschiedlicher Entscheidungssituationen ist dabei, dass Schülerinnen und Schüler Entscheidungsstrategien flexibel je nach Entscheidungssituationen anwenden können. Darüber hinaus enthalten alle entwickelten Aufgaben Unterstützungshilfen für eine Reflexion von Bewertungs- und Entscheidungsprozessen sowie der Anwendung von Entscheidungsstrategien.

Die beschriebenen Lernaufgaben wurden in den Lerngemeinschaften von bik in Niedersachsen und Thüringen auch in Unterrichtseinheiten zur kumulativen Förderung von Bewertungskompetenz eingesetzt. Dabei wurde im Rahmen eines Pre-Posttest Designs die Wirkung dieser Einheiten mit Hilfe des entwickelten Testinstruments gemessen. Die Ergebnisse dieser Interventionsstudien werden im Anschluss an dieses Vorhaben „Vom Modell zur empirischen Überprüfung“ veröffentlicht. Auch wird es in Zukunft durch eine Vergrößerung des Aufgabenpools sowie eine anschließende Identifikation von Items gleicher Schwierigkeit möglich sein, verschiedene Aufgabensets in Vor- und Nachtests einzusetzen. Damit können eventuelle Lerneffekte oder sinkende Motivation durch die Verwendung gleicher Aufgaben in wiederholten Testungen vermieden werden.

Der dritte Aspekt spricht die Dissemination der Ergebnisse von Biologie im Kontext an. Zum einen wurden die entwickelten Lernaufgaben im Rahmen von Lehrerfortbildungen und dabei speziell von Multiplikatorenfortbildungen für die Umsetzung der Bildungsstandards an niedersächsischen Schulen eingesetzt. Dabei waren viele der an der Lerngemeinschaft zu Bewertungskompetenz beteiligten Lehrkräfte involviert. Nicht zuletzt flossen die theoretischen und empirischen Erkenntnisse der vorliegenden Arbeit in die Entwicklung des niedersächsischen Kerncurriculums, welches die alten Rahmenrichtlinien ersetzte, ein (Niedersächsisches Kultusministerium, 2007).

Mit dieser Arbeit liegt ein Modell für Bewertungskompetenz vor, was es zukünftig möglich macht, Schülerkompetenzen im Biologieunterricht systematisch auf Basis der identifizierten Teilkompetenzen zu fördern. Damit kann das Modell einen Beitrag zur Reduktion von

40 Lücken, M. & Schröter, B. (Hrsg.) (2008): Biologie im Kontext. Aufgaben-CD-ROM. Leibniz-Institut für die Pädagogik der Naturwissenschaften (IPN): Kiel.

Heterogenitäten im Bezug auf den Begriff der Bewertungskompetenz leisten. Außerdem kann das Modell Lehrkräften als Orientierung dienen, was Bewertungskompetenz für den Biologieunterricht im Kern ausmacht und welche Strukturen bei der Entwicklung von Unterricht berücksichtigen werden sollten. In Hinblick auf eine psychometrische Modellierung konnte die Arbeit einen ersten Ansatz für eine zukünftige umfassendere Messung von Bewertungskompetenz liefern und dabei einen wertvollen Beitrag zur Stärkung des Kompetenzbereichs Bewertung im Vergleich zu den beiden traditionelleren Kompetenzbereichen Fachwissen und Erkenntnisgewinnung leisten.

10 Literaturverzeichnis

- Abelson, R. P. & Levi, A. (1985). Decision making and decision theory. In G. Lindzey & E. Aronson (eds.), *Handbook of Social Psychology (Vol 1)* (pp. 231-309). Reading, Mass.: Addison-Wesley.
- American Association for the Advancement of Science (1991). *Science for all Americans, Project 2061*. New York: OUP.
- Adams, R. & Wu, M. (2002). *PISA 2000. Technical Report*. Paris:OECD.
- Ahlf-Christiani, Chr. et al. (2003). Förderung der Urteilskompetenz im Fachunterricht der Sekundarstufe I und II. *Werkstattmaterialien BLK-Programm „21“ Nr. 9, Partizipatives Lernen – Nachhaltigkeitsindikatoren entwickeln* (online), <http://www.transfer-21.de> (abgerufen am 11. Juni, 2006).
- American Association for the Advancement of Science (1991). *Science for all Americans, Project 2061*. New York: Oxford University Press.
- Barkmann, J. & Bögeholz, S. (1999). *Ecosystem Assessment of Three Peatland Sites in Northern Germany as an Environmental Education Project*. Paper presented at "Case Studies in Environmental Education and Research"; Association of University Departments of Environmental Sciences in Europe (auDes) Conference; Zürich. [verfügbar über: http://www.ipn.uni-kiel.de/blk21-sh/Boege_Bark_ecosys.htm, Abrufdatum 12.11.2006]
- Barkmann, J. & Bögeholz, S. (2003). Kompetent gestalten, wenn es komplexer wird: Eine kurze Einführung zur ökologischen Bewertungs- und Urteilskompetenz. Zeitschrift „21“, 3, 49-52.
- Baumert, J., Bos, W. & Lehmann, R. (Hrsg.) (2000). *TIMSS/III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie. Mathematische und naturwissenschaftliche Bildung am Ende der Schullaufbahn. Band 1: Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit*. Opladen: Leske + Budrich.
- Bayrhuber, H. (1992). Ethische Analyse der Gentherapie von Keimbahnzellen im Unterricht. *Friedrich-Jahresheft*, 128-131.

- Bayrhuber, H., Bögeholz, S., Elster, D., Hammann, M., Hößle, C., Lücken, M., Mayer, J., Nerdel, C., Neuhaus, B., Prechtel, H. & Sandmann, A. (2007). Biologie im Kontext. *MNU* 60(5), 282-286.
- Beach, L. R. & Mitchell, T. R. (1978). A contingency model for the selection of decision strategies. *Academy Management Review*, 3, 439-449.
- Beaton, A. E. & Allen N. L. (1992). Interpreting Scales through Scale Anchoring. *Journal of Educational Statistics*, 17(2), 191-204.
- Beck, B. & Klieme, E. (2007). *Sprachliche Kompetenzen*. Weinheim, Basel: Beltz Verlag.
- Betsch, T., Haberstroh, S. & Höhle, C. (2002). Explaining Routinized Decision Making. A Review of Theories and Models. *Theory & Psychology*, 12(4), 453-488.
- Betsch, T. & Haberstroh, S. (2005). *The Routines of Decision Making* (S. 359-376). Mahwah, NJ: Erlbaum Associates.
- Biggs, J. B. & Collis, K. F. (1982). *Evaluating the Quality of Learning: The SOLO Taxonomy*. New York: Academic Press.
- Bögeholz, S. (2000). Naturerfahrung: Ein Baustein der Bildung für Nachhaltige Entwicklung – Natur erleben und gestalten. In ANU Bayern (Hrsg.), *Politische Ökologie. Sonderheft 12: Bildung für nachhaltige Entwicklung als neue Lernkultur – Schnittmenge Mensch*, 17-18.
- Bögeholz, S. & Barkmann, J. (2003). Ökologische Bewertungskompetenz für reale Entscheidungssituationen: Gestalten bei faktischer und ethischer Komplexität. *DGU-Nachrichten, Jahresheft 27/28*, 44-53.
- Bögeholz, S., Hößle, C., Langlet, J., Sander, E. & Schlüter, K. (2004). Bewerten - Urteilen - Entscheiden im biologischen Kontext: Modelle in der Biologiedidaktik. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 89-115.
- Bögeholz, S., & Barkmann, J. (2005). Rational choice and beyond: Handlungsorientierende Kompetenzen für den Umgang mit faktischer und ethischer Komplexität. In R. Klee, A. Sandmann & H. Vogt (Hrsg.), *Lehr- und Lernforschung in der Biologiedidaktik, Bd. 2* (S. 211-224). Innsbruck: Studien Verlag.
- Bögeholz, S. (2006). Explizites Bewerten und Urteilen. Beispielkontext Streuobstwiese. *Praxis der Naturwissenschaften – Biologie in der Schule*, 55(1), 17-24.
- Boone, W. J. & Scantlebury, K. (2005). The Role of Rasch Analysis When Conducting Science Education Research Utilizing Multiple Choice Tests. *Science Education*, 90, 253-269.

- Bond, T. G. & Fox, C. M. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. Mahwah, NJ: Erlbaum Associates.
- Briggs, D.C. & Wilson, M. (2003). An introduction to multidimensional measurement using Rasch models. *Journal of Applied Measurement*, 4(1), 87-100.
- Brown, A. L. (1997). Transforming schools into communities of thinking and learning serious matters. *American Psychologist* 52, 399-413.
- Bundesamt für Naturschutz (2003). Neoflora. Invasive gebietsfremde Arten in Deutschland. (<http://www.floraweb.de/neoflora/handbuch.html>; abgerufen am 8. September 2007).
- Bühner, M. (2006). *Einführung in die Test- und Fragebogenkonstruktion*. München: Pearson Studium.
- Bybee, R. (1997). *Achieving Scientific Literacy: From Purposes to Practices*. Portsmouth, NH: Heinemann.
- Davier, v. M. & Yamamoto, K. (2007). Mixture Distribution and HYBRID Rasch models. In M. v. Davier & C. Carstensen (Hrsg.), *Multivariate and mixture distribution Rasch models* (S. 99-115). New York: Springer.
- Davier, v. M. (1997). *Methoden zur Prüfung probabilistischer Testmodelle*. IPN: Kiel.
- Davier, v. M. & Rost, J. (1997). Self Monitoring - A Class Variable? In J. Rost & R. Langeheine (Hrsg.), *Applications of latent trait and latent class models in the social sciences* (S. 296-305). Münster: Waxmann.
- Davier, v. M. (2001). Winmira - A program system for analyses with the Rasch model, with the Latent Class Analysis and with the Mixed Rasch Model. Kiel: IPN Software.
- Deutsches PISA-Konsortium (Hrsg.) (2001). *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich*. Opladen: Leske + Budrich.
- Deutsches PISA-Konsortium (Hrsg.) (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster, New York, München, Berlin: Waxmann.
- Dulitz, B. & Kattmann, U. (1990). *Bioethik. Fallstudien für den Unterricht*. Stuttgart: Metzler.
- Eckensberger, L., Breit, H. & Döring, Th. (1999). Ethik und Barriere in umweltbezogenen Entscheidungen: Eine entwicklungspsychologische Perspektive. In V. Linneweber & E. Kals (Hrsg.), *Umweltgerechtes Handeln* (S. 165-189). Berlin: Springer.

- Eggert, S. & Bögeholz, S. (2006). Göttinger Modell der Bewertungskompetenz – Teilkompetenz „Bewerten, Entscheiden und Reflektieren“ für Gestaltungsaufgaben Nachhaltiger Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 177-199.
- Eggert, S. & Hößle, C. (2005). Bewertungskompetenz im Biologieunterricht. Ein Überblick. *Praxis der Naturwissenschaften – Biologie in der Schule*, 55(1), 1-10.
- Eigner-Thiel, S. & Bögeholz, S. (2004). Bildung für Nachhaltige Entwicklung aus Sicht von MultiplikatorInnen außerschulischer Bildungsträger. *Umweltpsychologie*, 8(2), 80-100.
- Ernst, A. (1997). *Ökologisch-soziale Dilemmata. Psychologische Wirkungsmechanismen des Umweltverhaltens*. Weinheim: Beltz Verlag.
- Europäische Kommission (2005). *Fischerei und Aquakultur in Europa*, 27. (http://ec.europa.eu/fisheries/publications/magaz/fishing/mag27_de.pdf; abgerufen am 8. Oktober 2008)
- Gigerenzer, G., Todd, P. M. & the ABC Research Group (1999). *Simple heuristics that make us smart*. New York: Oxford University Press.
- Gräsel, C. & Parchmann, I. (2004). Implementationsforschung – oder der steinige Weg, Unterricht zu verändern. *Unterrichtswissenschaft* 32(3), 196-214.
- Große, F. & Bögeholz, S. (2005). Explizite Bewertung am Beispiel der Streuobstwiese – Zur Bedeutung von Entscheidungs- und Ökosystemkontext. In M. Schrenk & W. Holl-Giese (Hrsg.), *Bildung für eine nachhaltige Entwicklung* (S. 159-173). Hamburg: Dr. Kovac.
- Haan, G. de & Harenberg D. (1999). Expertise „Förderprogramm Bildung für eine Nachhaltige Entwicklung“, verfasst für die Projektgruppe „Innovation im Bildungswesen“ der BLK im Auftrage des Bundesministeriums für Bildung, Wissenschaft, Forschung und Technologie. Freie Universität Berlin.
- Haan, G. de (2006). The BLK ‘21’ programme in Germany: a ‘Gestaltungskompetenz’-based model for Education for Sustainable Development. *Environmental Education Research*, 12(1), 19-32.
- Haidt, J. (2001). The Emotional Dog and its Rational Tail: A Social Intuitionist Approach to Moral Judgement. *Psychological Review*, 108(4), 814-834.
- Hammann, M. (2004). Kompetenzentwicklungsmodelle. *MNU*, 57(4), 196-203.
- Hammann, M. (2006a). PISA und Scientific Literacy. In R. Messner & U. Steffens (Hrsg.), *PISA macht Schule – Konzeptionen und Praxisbeispiele zur neuen Aufgabenkultur* (pp. 127-179). Wiesbaden: Institut für Qualitätsentwicklung.

- Hammann, M. (2006b). Kompetenzförderung und Aufgabenentwicklung. *MNU*, 59(2), 85-95.
- Hammann, M., Phan, T., Ehmer, M. & Bayrhuber, H. (2006c). Fehlerfrei Experimentieren. *MNU*, 59(5), 292-299.
- Harms, U., Mayer, J., Hammann, M., Bayrhuber, H. & Kattmann, U. (2004). Kerncurriculum und Standards für den Biologieunterricht in der gymnasialen Oberstufe. In H. E. Tenorth (Hrsg.), *Kerncurriculum Oberstufe 2 – Biologie, Chemie, Physik, Geschichte, Politik* (S. 22-85). Weinheim, Basel: Beltz Verlag.
- Hartig, J., Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127-143). Heidelberg: Springer Medizin Verlag.
- Hartig J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen* (S. 83 – 99). Weinheim, Basel: Beltz Verlag.
- Heller, K. A. & Perleth, C. (2000). Kognitiver Fähigkeitstest für 4. bis 12. Klassen. Beltz Test. Göttingen: Hogrefe.
- Helmke, A. & Hosenfeld, I. (2004) Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärungen und Perspektiven. In: R. S. Jäger & A. Frey (Hrsg.), *Lernprozesse , Lernumgebung und Lerndiagnostik. Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert*. Landau: Empirische Pädagogik.
- Hogan, K. (1999). Thinking Aloud Together: A Test of an Intervention to Foster Students' Collaborative Scientific Reasoning. *Journal of Research in Science Teaching*, 36(10), 1085-1109.
- Hogarth, R. M. (1987). *Judgement and choice: the psychology of decision*, 2nd Ed. Chichester: Wiley & Sons.
- Hong, J. L. & Chang, N. K. (2004). Analysis of Korean High School Students Decision-Making Process in Solving a Problem Involving Biological Knowledge. *Research in Science Education*, 34, 97-111.
- Höble, C. (2001). *Moralische Urteilsfähigkeit. Eine Interventionsstudie zur moralischen Urteilsfähigkeit von Schülern zum Thema Gentechnik*. Dissertation. Innsbruck: Studienverlag.
- Höble, C. & Bayrhuber, H. (2006). Sechs Schritte moralischer Urteilsfindung – Aktuelle Beispiele aus der Bioethikdebatte. *Praxis der Naturwissenschaften. Biologie in der Schule*, 4(55), 1-6.

- Jimenez-Aleixandre, M. P. (2002). Knowledge Producers or Knowledge Consumers? Argumentation and Decision Making about environmental management. *International Journal of Science Education*, 24(11), 1171-1190.
- Jungermann, H., Pfister, H. R. & Fischer, K. (1998). *Die Psychologie der Entscheidung (1. Auflage)*. Heidelberg, Berlin: Spektrum Akademischer Verlag.
- Jungermann, H., Pfister, H. R. & Fischer, K. (2004). *Die Psychologie der Entscheidung (2. Auflage)*. Heidelberg, Berlin: Spektrum Akademischer Verlag.
- Klahr, D. (2000). Exploring Science. *The Cognition and Development of Discovery Processes*. Cambridge: MIT Press.
- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, 6, 10-13.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. J. (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. Berlin: Bundesministerium für Bildung und Forschung.
- Klieme, E. & Leutner, D. (2006). Neues DFG-Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“. *Zeitschrift für pädagogische Psychologie*, 20(1/2), 137-138.
- KMK (2004). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss* (online), www.kmk.org/schul/Bildungsstandards/Biologie_MSA_16-12-04.pdf (abgerufen am 15. Juni 2004).
- KMK (2005). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss*. Luchterhand.
- Kohlberg, L. (1976). Moralstufen und Moralerwerb: Der kognitiv-entwicklungstheoretische Ansatz. In W. Althof, G. Noam & F. Oser (Hrsg.) (1995). *Lawrence Kohlberg. Die Psychologie der Moralentwicklung* (S. 179-192). Frankfurt am Main: Suhrkamp.
- Kolsto, S. D. (2006). Patterns in Students' Argumentation Confronted with a Risk-focused Socio-scientific Issue. *International Journal of Science Education*, 28(14), 1689-1716.
- Kortland, J. (1996). An STS case study about students' decision making on the waste issue. *Science Education*, 88(6), 673-689.
- Kuhn, D. (1999). A Developmental Model of Critical Thinking. *Educational Researcher*, 28, 16-25.
- Kyburz-Graber, R. et al. (1997). *Sozio-ökologische Umweltbildung*. Hamburg: Krämer

- Lachmayer, L., Nerdel, C. & Prechtel, H. (2007). Modellierung kognitiver Fähigkeiten beim Umgang mit Diagrammen im naturwissenschaftlichen Unterricht. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 145-160.
- Lazarsfeld, P. F. & Henry, N. W. (1968). *Latent Structure Analysis*. New York: Houghton Mifflin.
- Masters, G. N. (1982). A Rasch model for Partial Credit Scoring. *Psychometrika*, 47, 149-174.
- Mayer, J., Harms, U., Hammann, M., Bayrhuber, H., Kattmann, U. (2004). Kerncurriculum Biologie der gymnasialen Oberstufe. *MNU* 57(3), 166-173.
- Mayring, P. (2000). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Dt. Studien-Verlag.
- Meisert, A. & Kierdorf, H. (2001). Bioethische Themen im Biologieunterricht. Fachdidaktische Umsetzung am Beispiel der operativen Trennung Siamesischer Zwillinge. *MNU* 54(6), 359-363.
- Müller, S. (2006). *Analyse von Bewertungskompetenz bei Schülerinnen und Schülern der Sekundarstufe I unter Verwendung des Lauten Denkens. Eine empirische Studie*. Erste Staatsprüfung. Georg-August-Universität Göttingen.
- National Committee on Science Standards and Assessment (1996). *National Science Education Standards*. Washington, DC: National Academic Press.
- National Research Council (1995). *National science education standards*. Washington, DC: National Academic Press.
- Niedersächsisches Kultusministerium (2007). Kerncurriculum für das Gymnasium Schuljahrgänge 5-10. Naturwissenschaften, www.cuvo.nibis.de (abgerufen am 20.12.2007)
- Organization for Economic Cooperation and Development (2004). *Problem Solving for Tomorrow's World. First Measures of Cross-Curricular Competencies from PISA 2003*. Paris: OECD.
- Payne, J. W., Bettmann, J. R. & Luce, M. F. (1998). Behavioral Decision Research: An Overview. In M. H. Birnbaum (Hrsg.), *Measurement, Judgment, and Decision Making, Handbook of Perception and Cognition*, 2nd Ed. San Diego: Academic Press.
- Phan, T. & Hammann, M. (2006). *Testing levels of competencies in biological experimentation*. Vortrag auf der Tagung "Sixth conference of European Researchers in Didactics of Biology: ERIDOB 2006", University of London (2006).
- Polya, G. (1954). *How to solve it*. Princeton, N: Princeton University Press.

- Poschmann, C., Riebenstahl, C. & Schmidt-Kallert, E. (1998). *Umweltplanung und -bewertung*. Stuttgart: Klett Perthes.
- Pfister, H. R. & Böhm, G. (1992). The function of concrete emotions in rational decision making. *Acta Psychologica*, 80, 199-211.
- Qualification and Curriculum Authority (1999). *The National Curriculum for England. Key Stages 1-4*. London: Crown.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danmarks pädagogiske Institut.
- Ratcliffe, M. (1997). Pupils' decision making about socio-scientific issues within the science curriculum. *International Journal of Science Education*, 19(2), 167-182.
- Ratcliffe, M. & Grace, M. (2003). *Science education for citizenship*. Maidenhead: OUP.
- Reitschert, K. & Hößle, C. (2007). Wie Schüler ethisch bewerten. Eine qualitative Untersuchung zur Strukturierung und Ausdifferenzierung von Bewertungskompetenz in bioethischen Sachverhalten bei Schülern der Sek. I *Zeitschrift für Didaktik der Naturwissenschaften* 13, 125-143.
- Rettinger, D. A., & Hastie, R. (2003). Comprehension and Decision Making. In S. Schneider & J. Shanteau (Hrsg.), *Emerging Perspectives on Judgment and Decision Research* (pp. 165-200). Cambridge: University Press.
- Roberts, L., Wilson, M. & Draney, K. (1997). *The SEPUP assessment system: An Overview*. BEAR Report Series, SA-97-1. Berkeley: University of California.
- Rost, J. (1990). Rasch Models in Latent Classes: An Integration of Two Approaches to Item Analysis. *Applied Psychological Measurement*, 14(3), 271-282.
- Rost, J. (2004a). *Lehrbuch Testtheorie – Testkonstruktion*. Bern: Huber.
- Rost, J. (2004b). Zählen oder Messen. In A. Diekmann (Hrsg.), *Methoden der Sozialforschung. Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 44, 75-92.
- Rost, J. (2006). Kompetenzstrukturen und Kompetenzmessung. *Praxis der Naturwissenschaften* 8(55), 5-8.
- Rychen, D. S. & Salganik, L. H. (2001): *Defining and Selecting Key Competencies*. Göttingen: Hogrefe & Huber.
- Sadler, T. D. & Zeidler, D. L. (2004). Patterns of informal reasoning in the context of socioscientific decision making. *Journal of Research in Science Teaching*, 42(1), 4-27.
- Sadler, T. D., Barab, S. A., & Scott, B. (2007). What do students gain by engaging in socioscientific inquiry? *Research in Science Education*, 37(4), 371-391.

- Scantlebury, K., Boone, W., Butler-Kahle, J. & Fraser, B. J. (2001). Design, Validation, and Use of an Evaluation Instrument for Monitoring Systemic Reform. *Journal of Research in Science Teaching*, 38(6), 646-662.
- Science Education for Public Understanding Project (1995). *Issues, Evidence, and You: Teacher's Guide*. Berkeley: University Of California, Lawrence Hall of Science.
- Seethaler, S. & Linn, M. (2004). Genetically modified food in perspective: an inquiry-based curriculum to help middle school students make sense of tradeoffs. *International Journal of Science Education*, 26(14), 1765-1785.
- Senkbeil, M., Rost, J., Carstensen, C.H. & Walter, O. (2003). Der nationale Naturwissenschaftstest PISA 2003. Entwicklung und empirische Überprüfung eines zweidimensionalen Facettendesigns. *Empirische Pädagogik*, 19(2), 166-189.
- Siegel, M. A. (2006). High school students' decision making about sustainability. *Environmental Education Research*, 12(2), 201-215.
- Simon, H. A. (1976). *Administrative behaviour: A study of decision making processes in administrative organization*, 3rd Ed. New York: Free Press.
- SPSS Inc. (2003). *Statistical Package for the Social Sciences*, Version 12.0.1 (Computer Software). Chicago: SPSS Inc.
- Svenson, O. (1990). Some propositions for the classification of decision situations. In K. Borcherding, O.I. Larichev & D.M. Messick (Hrsg.), *Contemporary Issues in decision making* (pp. 17-33). Amsterdam: Elsevier.
- Svenson, O. (2003). Values, Affect, and Processes in Human Decision Making: A Differentiation and Consolidation Theory Perspective. In S. Schneider, & J. Shanteau (Hrsg.), *Emerging Perspectives on Judgment and Decision Research* (pp. 287-326). Cambridge: University Press.
- Tversky, A. (1972). Elimination by aspects: a theory of choice. *Psychological Review*, 79, 281-299.
- Walter, O. (2005). *Kompetenzmessung in den PISA Studien – Simulationen zur Schätzung von Verteilungsparametern und Reliabilitäten*. Lengerich: Papst.
- Watermann, R. & Klieme, E. (2002). Reporting results of Large – Scale Assessment in psychologically and educationally meaningful terms. *European Journal of Psychological Assessment*, 18(3), 190 – 203.
- Weinert, F. E. (1999). *Concepts of competence (Contribution within the OECD project Definition and selection of competencies: Theoretical and conceptual foundations (DeSeCo))*. Neuchâtel: DeSeCo.

- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. E. Weinert (Hrsg.), *Leistungsmessungen in Schulen*. Weinheim und Basel: Beltz Verlag, S. 17-31.
- Westenberg, M. & Koele, P. (1994). Multi-attribute evaluation processes: Methodological and conceptual issues. *Acta Psychologica*, 87, 65-84.
- Willenberg, H. (2007). Lesen. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen* (S.107-117). Weinheim, Basel: Beltz Verlag.
- Wilson, M. & Sloane, K. (2000). From Principles to Practice: An Embedded Assessment System. *Applied Measurement in Education*, 13(2), 181–208.
- Wilson, M. (2005). *Construction Measures: An Item Response Modeling Approach*. Mahwah, New Jersey, London. Lawrence Erlbaum Associates.
- Wright, B.D. & Linacre, J.M. (1989). Observations are Always ordinal; Measurements, however, Must be Interval. *Archives of Physical Medicine and Rehabilitation*, 70(12), 857-860.
- Wright, B.D. & Masters, G.N. (1982). *Rating Scale Analysis*. Chicago: MESA Press.
- Wright & Stone (1999). *Measurement Essentials*. 2nd Edition (pp. 167-171). Wilmington, Delaware: Wide Range Inc.
- Wu, M. L., Adams, R. J. & Wilson, M. R. (1998). *Con-Quest: Multi-aspect test software*. Melbourne: Australian Council for Educational Research.
- Zeidler, D. L. (2003). *The Role of Moral Reasoning on Socioscientific Issues and Discourse in Science Education*. Dordrecht, Boston, London: Kluwer Academic Publishers.
- Yamamoto, K. (1989). *A Hybrid model of IRT and latent class models. (ETS Research report (RR-89-41)). Educational Testing Service*.
- Yang, F. Y. (2004). Exploring high school students' use of theory and evidence in an everyday context: the role of scientific thinking in environmental science decision making. *International Journal of Science Education*, 26(11), 1345-1364.
- Yang, F. Y. & Anderson, O. R. (2003). Senior high school students' preference and reasoning modes about nuclear energy use. *International Journal of Science Education*, 25(2), 221-244
- Zeidler, D. L. (2003). The role of moral reasoning and discourse on socio-scientific issues in science education. Dordrecht: Kluwer Academic Publishers.
- Zohar, A. & Nemet, F. (2002). Fostering Students' Knowledge and Argumentation Skills Through Dilemmas in Human Genetics. *Journal of Research in Science Teaching*, 39 (1), 36-62.

11 Anhang

Hinweise:

1. Die kompletten Fragebögen für die Studie 1 sowie die Studien 2 und 3 des zweiten Forschungsschwerpunkts der vorliegenden Arbeit können bei der Autorin angefragt werden. Eine gekürzte Version des Fragebogens für die Studie 1 ist im Appendix A und B des Kapitels 6 gegeben. Eine gekürzte Version des Fragebogens für die Studien 2 und 3 ist im Appendix A und B des Kapitels 7 gegeben.
2. Im Folgenden ist unter Punkt 11.1 der Scoring Guide dargestellt, der der Auswertung der Studien 2 und 3 (vgl. Kapitel 7) zugrunde liegt. In die finale Auswertung gingen 11 Indikatoren ein. Diese sind mit * gekennzeichnet.
Für die Studie 3 wurde der Fragebogen durch zwei Subskalen des Kognitiven Fähigkeitstests ergänzt. Dabei kamen die beiden Subskalen V1 und V3 zur Messung der verbalen Ausdrucksfähigkeit zum Einsatz (vgl. Heller & Perleth, 2000).
3. Unter Punkt 11.2 ist eine ausführliche Beschreibung des Scorings von Schülerantworten gegeben, die die Vorgehensweise anhand von Beispielantworten verdeutlichen soll. Exemplarisch ist das Scoring für die Dorschaufrage (Bewertungsaufgabe) sowie für die Schokoladenaufgabe (Reflexionsaufgabe) gegeben.

11.1 Scoring Guide zur Kodierung von Schülerantworten (Studien 2 und 3, vgl. Kapitel 7)

Dorschaufgabe (Bewertungsaufgabe, Heft 1)					
	Score 0	Score 1	Score 2	Score 3	Score
Verfahren mit gewählter Option (nichts - nur positiv - positiv & negativ)*					
Verfahren mit nicht gewählten Optionen (nichts - nur negativ - negativ sowie positiv & negativ - negativ & positiv)*					
Vergleich <u>gewählter Option</u> mit <u>nicht gewählten</u>					
Vergleich <u>nicht gewählter</u> Optionen miteinander					
Gewichten von Entscheidungskriterien*					
Dokumentation					
K1: Ladenangebot direkt					
K2: Ladenangebot in 3 Jahren					
K3: Bedrohung					
K4: Arbeitsplätze					
Weitergehende Fakten					
Persönliche Präferenzen					
Informationen falsch interpretiert					

Staudenknöterichaufgabe (Bewertungsaufgabe, Heft 1)					
	Score 0	Score 1	Score 2	Score 3	Score
Verfahren mit gewählter Option (nichts - nur positiv - positiv & negativ)*					
Verfahren mit nicht gewählten Optionen (nichts - nur negativ - negativ sowie positiv & negativ -negativ & positiv)*					
Vergleich <u>gewählter Option</u> mit <u>nicht gewählten</u>					
Vergleich <u>nicht gewählter</u> Optionen miteinander					
Gewichten von Entscheidungskriterien*					
Dokumentation					
K1: Wirkung					
K2: Dauer und Häufigkeit					
K3: Kosten					
K4: Auswirkung					
Weitergehende Fakten					
Persönliche Präferenzen					
Informationen falsch interpretiert					

Schokoladenaufgabe (Reflexionsaufgabe, Heft 2)						
	Score 0	Score 1	Score 2	Score 3	Score 4	Score
Non-kompensatorische Strategie „Paul“*						
Kompensatorische Strategie „Martina“*						
Angemessener Entscheidungsweg*			X	X	X	
Vorschläge kompensatorische Strategie „Martina“*			X	X	X	
Intuitiv-rechtfertigendes Entscheidungsverhalten „Stefan“*			X	X	X	
Vorschläge intuitiv-rechtfertigendes Entscheidungsverhalten „Stefan“			X	X	X	

11.2 Ausführliche Beschreibung zum Scoren von Schülerantworten

Dorschaufrage (Heft 1) (Die Dorschaufrage stellt eine der beiden <i>Bewertungsaufgaben</i> des Hefts 1 des Fragebogens dar, bei der Schülerinnen und Schüler selbst eine Entscheidung treffen müssen. Die zweite Aufgabe, die Staudenknotenaufgabe, ist von der Struktur her mit der Dorschaufrage identisch und wurde dementsprechend anhand derselben Indikatoren gescored.)		
Indikator / Item	Beschreibung Scores	Beispielantworten (Angegeben sind Auszüge aus Schülerantworten, die sich auf den jeweiligen Indikator beziehen.)
Verfahren mit gewählter Option	Score 0: Es wird kein Aspekt für die gewählte Option benannt	„Ich habe mich für Möglichkeit 3 [Schutzzeiten und Schutzzeiten] entschieden.“ (Klasse 6) „Ich entscheide mich für 3 [Schutzzeiten und Schutzzeiten], weil die anderen Sachen mir etwas blöde erscheinen.“ (Klasse 6)
	Score 1: Es werden ausschließlich positive Aspekte für die gewählte Option benannt oder negative Aspekte werden positiv dargestellt.	„Ich habe mich für Nr. 4 [Fangmenge an Dorschbestand angepasst] entschieden, weil die Dorsche dadurch nicht aussterben würden. Die Läden könnten immer Umsatz machen und die Arbeitsplätze in der Fischerei wären nicht bedroht.“ (Klasse 8) „Ich finde Nummer 3 [Schutzzeiten und Schutzzeiten] gut, da sich alles ein bisschen verbessern kann, wenn man nicht dort fischen darf, wo die Dorsche laichen. Außerdem haben wir dann in drei Jahren genug Dorsche und es kann keine Überfischung geben. Die Arbeitsplätze sind auch gesichert. Das ist gut.“ (Klasse 8) „Ich würde mich für die Möglichkeit 3 entscheiden, weil Dorsche sind auch Lebewesen, die auch mal Ruhe brauchen und sich erholen müssen. Schutzzeiten sind dafür da, damit die Dorsche ihre Eier ablegen können. Das Gute ist außerdem, die Arbeitsplätze bleiben gesichert und es wird keiner arbeitslos. In den Läden reicht es, wenn auch etwas weniger Fische verkauft werden.“ (Klasse 8)
<i>(Anmerkung: Generell war es möglich, sowohl eine Option auszuwählen, als auch eine Kombination aus mehreren Optionen zu wählen.)</i>		

<p>Fortsetzung: Verfahren mit gewählter Option</p>	<p><u>Score 2:</u> Es werden sowohl positive als auch negative Aspekte für die gewählte Option benannt.</p>	<p>„Ich bin für Maßnahme 3 [Schutzzeiten und Schutzzonen], da keine Überfischung mehr möglich ist und die Fische sich erholen können. Zwar ist das Ladenangebot zunächst gering, aber dies regeneriert sich in den folgenden Jahren. Des Weiteren sind die Arbeitsplätze trotz allem gesichert und somit ist diese Maßnahme meiner Meinung nach die sinnvollste.“ (Klasse 10)</p> <p>„Wenn ich mich entscheiden müsste, würde ich den 3. Punkt [Schutzzeiten und Schutzzonen] wählen. Der Tabelle kann ich dieser Möglichkeit hauptsächlich positive Informationen entnehmen. Das Ladenangebot nach Umsetzung der Maßnahme ist jedoch gering, während es bei Punkt 1 [Vergrößerte Maschenöffnungen] höher ist, doch dafür ist bei Punkt 3 [Schutzzeiten und Schutzzonen] das Ladenangebot in 3 Jahren ausreichend, Überfischung ist nicht möglich und im Gegensatz zu manchen anderen Punkten sind die Arbeitsplätze gesichert.“ (Klasse 10)</p> <p>„Die Maßnahme 3 [Schutzzeiten und Schutzzonen] ist meiner Ansicht nach die Beste, da eine Überfischung nicht möglich ist und die Arbeitsplätze auch nicht gefährdet sind. Bei Maßnahme 3 ist der Nachteil das geringe Angebot direkt nach der Umsetzung, aber wenn sich der Bestand erholt hat, kann auch das Angebot wieder steigen.“ (Klasse 12)</p>
<p>Verfahren mit nicht gewählten Optionen</p>	<p><u>Score 0:</u> Es wird kein Aspekt zu den negativen Optionen benannt.</p>	<p>[Anmerkung: • Bei einem Score 0 wird zu den nicht gewählten Optionen Nichts gesagt. Es wird beispielsweise ausschließlich die gewählte Option begründet oder • Es wird nur gesagt, dass man die Optionen X und Y nicht wählt, aber es wird keine Begründung dazu geliefert.]</p> <p>„Ich habe die Maßnahmen 1 [Vergrößerte Maschenöffnungen], 2 [Einstellung des Dorschfangs für 3 Jahre] und 4 [Fangmenge an Dorschbestand angepasst] nicht genommen, weil ich die nicht so gut finde.“ (Klasse 6)</p>

<p>Fortsetzung: Verfahren mit nicht gewählten Optionen</p>	<p><u>Score 1:</u> Es werden nur negative Aspekte für eine oder mehrere nicht gewählte Optionen benannt.</p>	<p>„Bei der ersten Möglichkeit [Vergrößerte Maschenöffnungen] ist die Überfischung trotzdem möglich. Also ist sie nicht sinnvoll. Den Dorschfang für drei Jahre einzustellen [Option 2], würde Arbeitsplätze bedrohen. Die Fangmenge an den Dorschbestand anzupassen [Option 4] würde das Ladenangebot stark schwanken lassen und die Arbeitsplätze gefährden.“ (Klasse 10)</p> <p>„Für nicht sinnvoll halte ich die Möglichkeiten 1 [Vergrößerte Maschenöffnungen] und 2 [Einstellung des Dorschfangs für 3 Jahre]. Bei Möglichkeit 1 besteht die Gefahr der Überfischung und bei Möglichkeit 2 denke ich, dass es auch nicht unbedingt helfen würde, da die Nachfrage an Dorsch nach Ablauf der Einstellung des Fangs nach 3 Jahren wieder steigen würde. An Möglichkeit 4 [Fangmenge an Dorschbestand angepasst] ist noch zu kritisieren, dass die Nachfrage eventuell höher sein könnte als die Menge an Dorsch, die zur Verfügung steht.“ (Klasse 10)</p>
	<p><u>Score 2:</u> Es werden</p> <ul style="list-style-type: none"> • negative Aspekte für einige der nicht gewählten Optionen benannt sowie • negative und positive Aspekte für die übrig gebliebenen nicht gewählten Optionen benannt. 	<p>„An der Maßnahme 1 [Vergrößerte Maschenöffnungen] ist das Problem, dass die Überfischung möglich ist. An die Maßnahme 3 [Schutzzeiten und Schutzzonen] werden sich meiner Meinung nach nicht alle halten, so dass die Gefahr einer Überfischung doch besteht. Maßnahme 4 [Fangmenge an Dorschbestand angepasst] wäre auch eine Lösung, da die Überfischung nicht möglich ist, jedoch glaube ich nicht, dass es möglich ist, den genauen Dorschbestand exakt ermitteln zu können.“ (Klasse 10)</p> <p>„Bei Möglichkeit 1 [Vergrößerte Maschenöffnungen] besteht die Problematik, dass eine eventuelle Überfischung entstehen kann, die dann zu einem erheblichen Rückgang der Dorsche in der Ostsee führen kann. Die zweite Möglichkeit [Einstellung des Dorschfangs für 3 Jahre] ist zwar theoretisch eine sehr gute Idee, aber wirtschaftlich kaum umsetzbar, da somit viele Arbeitsplätze bedroht wären. Allerdings würde sich der Dorschbestand wieder erholen. Ein weiteres Problem bei dieser Lösung ist, dass sich nach den drei Jahren der Dorschbestand schnell wieder verringern würde. Die 4. Lösung [Fangmenge an Dorschbestand angepasst] hat das Problem, dass die ganze Industrie vom Bestand der Fische abhängen würde und somit gefährdet wäre.“ (Klasse 12)</p>

Fortsetzung: Verfahren mit nicht gewählten Optionen	Fortsetzung Score 2	<p>Bei Maßnahme 1 [Vergrößerte Maschenöffnungen] kann trotz vergrößerter Maschen Überfischung stattfinden, was somit auch Arbeitsplatzstreichung mit sich führt. Die Dorsche für 3 Jahre ganz in Ruhe zu lassen [Option 2] klingt plausibel, damit sie sich erholen können. Das Problem jedoch ist damit nicht langfristig beseitigt.“ (Klasse 12) [Anmerkung: Es wird eine Kombination aus Maßnahmen 3 und 4 gewählt.]</p> <p>„Die erste Maßnahme [Vergrößerte Maschenöffnungen] scheint zunächst sehr geeignet, da die Jungfische nicht gefangen werden, die Adulten allerdings schon, wodurch man auch ein großes Angebot bringen kann. Doch es ist hierdurch nicht gesichert, dass wirklich alle Jungfische davorkommen. Ich würde mich gegen diese Maßnahme aussprechen, [...], schon allein, weil immer noch die Gefahr der Überfischung besteht. Die zweite Möglichkeit [Einstellung des Dorschfangs für 3 Jahre] wäre für den Dorsch innerhalb der 3 Jahre sehr gut. Was allerdings nach Ablauf der Frist nicht mehr zutrifft, da dann wieder alles von vorn beginnen könnte. Auch, dass während dieser 3 Jahre die Arbeitsplätze der Fischer bedroht sind, veranlasst mich, mich dagegen zu entscheiden. Dass man die Fangmenge an den Dorschbestand anpasst [Option 4], halte ich für keine geeignete Methode, da man sich auch mal irren könnte und die Dorsche zu stark gefangen werden. Dazu kommt, dass es im Bereich des Angebots und der Arbeitsplätze zu starken Schwankungen kommt.“ (Klasse 12)</p>
	Score 3: Es werden sowohl negative als auch positive Aspekte für alle nicht gewählten Optionen benannt.	<p>„Gegen 1 [Vergrößerte Maschenöffnungen] spricht, dass es zu einer Überfischung kommen kann und die Arbeitsplätze auch gefährdet sein können. Positiv ist hier, dass das Angebot an Dorsch groß genug bleibt. Gegen 2 [Einstellung des Dorschfangs für 3 Jahre] spricht, dass es dann 3 Jahre lang keinen Dorsch zu essen gibt und die Arbeitsplätze bedroht sind. Dafür spricht, dass sich der Dorschbestand vollständig erholen kann und nach den drei Jahren wieder genug Dorsch vorhanden ist. Gegen 4 [Fangmenge an Dorschbestand anpasst] spricht, dass die Zahl der Arbeitsplätze schwankt und das Angebot an Dorsch nicht gleich hoch bleibt. Dafür spricht, dass eine Überfischung nicht möglich ist.“ (Klasse 10)</p> <p>„Die Maßnahme zwei [Einstellung des Dorschfangs für 3 Jahre] ist zwar kurzfristig für den Dorschbestand am besten. Allerdings wird hierdurch zum einen die wirtschaftliche Situation der</p>

Fortsetzung: Verfahren mit nicht gewählten Optionen	Fortsetzung Score 3	<p>Fischer verschlechtert und vor allem ist es keine langfristige Lösung, denn diese Maßnahme bekämpft nur Auswirkungen aber auf keinen Fall die Ursachen. Die Fangmenge an den Bestand anzupassen [Option 4] ist zwar langfristig gesehen eine effektive Möglichkeit allerdings aufgrund der starken Schwankungen aus politischer Sicht unpraktikabel. Ich persönlich finde eine Kombination aus vergrößerten Netzen und Schutzzeiten am sinnvollsten [...]“ (Klasse 12)</p> <p>Bei Maßnahme 1 [Vergrößerte Maschenöffnungen] ist zwar das Ladenangebot gesichert, jedoch könnte es weiterhin zu Überfischung kommen (z.B. wenn in der Zukunft die Jungtiere größer sind als heutzutage). Maßnahme 2 [Einstellung des Dorschfangs für 3 Jahre] halte ich auch für wenig sinnvoll, denn auch wenn sich der Bestand zunächst vollständig erholen kann, so wird danach der Bestand schneller wieder zerstört, als er sich überhaupt aufgebaut hat.“ (Anmerkung: <i>Gewählt wurde eine Kombination aus den Möglichkeiten 3 und 4.</i>) (Klasse 12)</p>
Gewichten von Entscheidungs- kriterien	<p><u>Score 0:</u> Kriterien werden nicht explizit gewichtet.</p> <p><u>Score 1:</u> Kriterien werden explizit gewichtet.</p>	<p>[Anmerkung: Ein Score 0 wurde vergeben, wenn in der Schülerantwort an keiner Stelle eine explizite Gewichtung von Entscheidungskriterien erkennbar ist.</p> <p>[Anmerkung: Eine explizite Gewichtung ist in der Regel durch Signalwörter wie „wichtig / unwichtig“, „relevant / irrelevant“ oder durch Schülerantworten wie „für mich zählt v.a.“ oder „auf keinen Fall sollte Kriterium X vernachlässigt werden“ gekennzeichnet. Eine explizite Gewichtung liegt auch dann vor, wenn der Schülerantwort zu entnehmen ist, dass alle Entscheidungskriterien gleich wichtig sind.]</p> <p>„Ich finde, dass man Schutzzeiten [Option 3] einführen sollte [...] Durch diese Möglichkeit gibt es eine Zeit lang zwar nicht mehr genug Fische als Ladenangebot, doch ich finde, dies ist nicht so wichtig, denn man könnte auch auf andere Fische ausweichen und nach einer bestimmten Zeit wird es auch wieder Dorsche geben.“ (Klasse 8)</p> <p>„Ich entscheide mich für Möglichkeit 3, da keine Überfischung möglich ist und die Arbeitsplätze gesichert sind. Zwar ist das Ladenangebot anfangs gering, aber am wichtigsten ist doch das gesicherte Überleben der Art.“ (Klasse 12)</p> <p>„Für mich zählt v.a., dass eine Überfischung nicht möglich ist und damit der Bestand der Fische sich regenerieren kann.“ (Klasse 12)</p>

Schokoladenaufgabe (Heft 2) (Die Schokoladenaufgabe stellt im Gegensatz zu den Aufgaben des Hefts 1 des Fragebogens eine <i>Reflexionsaufgabe</i> dar, bei der Schülerinnen und Schüler über Bewertungsprozesse anderer Personen reflektieren müssen.)	
Indikator / Item	Beispielantworten (Angabe sind Auszüge aus den Schülerantworten, die sich auf den jeweiligen Indikator beziehen.)
Non-kompensatorische Strategie („Paul“)	<p><u>Score 0:</u> Pauls Vorgehensweise wird <u>nicht</u> beschrieben.</p> <p><u>Score 1:</u> Pauls Vorgehensweise wird <u>ausschließlich inhaltlich</u> beschrieben.</p> <p><u>Score 2:</u> Pauls Vorgehensweise wird <u>teilweise methodisch</u> beschrieben.</p>
	<p><i>[Anmerkung: Es wird entweder Nichts zu Pauls Vorgehensweise gesagt, oder es wird lediglich die Entscheidung benannt]</i></p> <p>„Paul hat Schokolade Nr. 3 genommen.“ (Klasse 6)</p> <p>„Paul findet die 3. [Schokolade] gut, denn die ist fair gehandelt“ (Klasse 6)</p> <p>„Paul entscheidet sich für eine konventionell hergestellte Schokolade.“ (Klasse 8)</p> <p><i>[Anmerkung: Gemeint ist hier der konventionelle Anbau der Kakaobohnen]</i></p> <p>„Paul hat sich für die 3 [Schokolade Nr. 3] entschieden, weil sie fair gehandelt wurde, nicht zu teuer ist und keine gentechnisch veränderten Zutaten hat.“ (Klasse 10)</p> <p>„Paul lässt sobald ein Kriterium schlecht ist, gleich eine Marke „ausscheiden“.“ (Klasse 6)</p> <p>„Pauls Entscheidungsweg ist so gewesen, dass er die [Schokoladen], die dem Kriterium nicht entsprechen einfach rausgestrichen hat.“ (Klasse 10)</p> <p>Paul hat das „Ausschlussprinzip“ angewendet. Dabei versucht er natürlich alle Punkte irgendwie zu berücksichtigen.“ (Klasse 12)</p>

<p>Fortsetzung: Non-kompensatorische Strategie („Paul“)</p>	<p><u>Score 3:</u> Pauls Vorgehensweise wird <u>vollständig methodisch</u> beschrieben.</p>	<p>„Paul hat sich angeguckt, was ihm wichtig ist und danach Schokoladen abgeschlossen, so dass er auf Schokolade 3 kommt.“ (Klasse 8)</p> <p>„Paul setzt drei Kritikpunkte, die er nicht haben möchte. Die Schokolade, die am Ende die wenigsten Kritikpunkte vorweist, wählt er aus, in diesem Fall Schokolade 3.“ (Klasse 10)</p> <p>„Paul geht nach dem Ausschlussprinzip vor. Er überprüft die versch. Schokoladensorten auf die Kriterien (angefangen beim wichtigsten Kriterium) und streicht die Schokolade raus, die ein Kriterium nicht erfüllt und das geht solange bis nur noch eine Möglichkeit übrig bleibt.“ (Klasse 12)</p>
<p><u>Score 4:</u> Pauls Vorgehensweise wird <u>vollständig methodisch</u> beschrieben <u>und</u> der in dieser Vorgehensweise vorhandene <u>Fehler / Konflikt</u> wird <u>erkannt</u>.</p>	<p>„Paul hat die Sorten ausgestrichen, die auf einen Wunsch nicht zutreffen. Jede der Schokoladen hat allerdings im Endeffekt einen unerwünschten Nachteil, so dass er alle hätte wegstreichen müssen.“ (Klasse 12)</p> <p>„Paul handelt nach dem Ausschlussverfahren und geht seine Kriterien nacheinander ab, so dass am Schluss eine übrig bleibt (bzw. zwei, was er anscheinend jedoch nicht weiter abwägt – biologisch oder fair gehandelt?)“ (Klasse 12)</p> <p>„Paul hat von seinem wichtigsten Kriterium ausgehend nach dem Ausschlussverfahren entschieden. Bei Paul ist das Problem, dass er so rigoros aussortiert. So kann er nicht unbedingt alle seine Wünsche verwirklichen, sondern nur ausklammern, was in einigen Punkten völlig den Kriterien widerspricht und muss sich dann mit dem übriggebliebenen begnügen.“ (Klasse 12).</p> <p>„Paul verfolgt seinen Prioritätenkatalog. Seine Entscheidung, zwei Sorten aufgrund einer sehr hohen (wichtigen) Priorität auszuschließen, ist nachvollziehbar. Nicht nachvollziehbar ist seine Entscheidung für Schokolade 3, da sie teurer als Schokolade 4 ist und nicht biologisch angebaut wurde, was zuvor als „wichtig“ angegeben wurde.“ (Studierender)</p>	

Kompensatorische Strategie („Martina“)	Score 0: Martinas Vorgehensweise wird <u>nicht</u> beschrieben.	[Anmerkung: Es wird entweder Nichts zu Martinas Vorgehensweise gesagt oder es wird lediglich die Entscheidung benannt.] „Martina nimmt Schokolade Nr. 4.“ (Klasse 6)
	Score 1: Martinas Vorgehensweise wird <u>ausschließlich inhaltlich</u> beschrieben.	„Martina hat sich für 4 [Schokolade Nr. 4] entscheiden weil sie billig ist, keine gentechnisch veränderten Zutaten drin sind und die [Kakao-]Pflanze biologisch angebaut wurde.“ (Klasse 8) „Martina entscheidet sich für die biologisch angebaute Schokolade, die allerdings konventionell gehandelt wurde. Sie ist nicht sehr teuer und enthält keine gentechnisch veränderten Zutaten.“ (Klasse 8)
	Score 2: Martinas Vorgehensweise wird <u>teilweise methodisch</u> beschrieben.	„Martina rechnet erstmal aus, wie viel Punkte die einzelnen Schokoladen haben und vergleicht dann.“ (Klasse 8) „Martina guckt sich die Schokolade im Ganzen an und zählt „Punkte“ in Form von Smilies zusammen, die Schokolade mit den meisten Punkten gewinnt.“ (Klasse 8) „Martina hat alle Kriterien in ihre Bewertung einfließen lassen.“ (Klasse 12)
	Score 3: Martinas Vorgehensweise wird <u>vollständig methodisch</u> , aber ohne den Aspekt der Gewichtung von Entscheidungskriterien, beschrieben.	„Martina betrachtet die vier Schokoladensorten nach jedem einzelnen Kriterium. Jede Schokolade sammelt so mehr oder weniger Punkte. Die Punkte, die für ein nicht erfülltes Kriterium verloren gehen, können durch ein gutes Kriterium wieder aufgeholt werden. So lässt sich die Entscheidung nach Betrachtung aller Kriterien fällen.“ (Klasse 10) „Martina schaut sich für jede Schokolade jeden Punkt einzeln an und verteilt Punkte in Bezug auf jedes Kriterium. Die Schokolade mit den meisten Punkten wird ausgewählt.“ (Klasse 10) „Martina hat sozusagen Punkte (hier Smilies) vergeben: erfüllt eine Schokolade ein Kriterium voll und ganz, vergibt sie 3 Smilies, erfüllt ein Kriterium die Bedingung nur unbefriedigend, vergibt sie nur einen Punkt. Am Ende wählt sie die Schokolade mit den meisten Punkten aus. Damit kommt sie auf ein anderes Ergebnis als Paul.“ (Klasse 12)

<p>Fortsetzung: Kompensatorische Strategie („Martina“)</p>	<p><u>Score 4:</u> Martinas Vorgehensweise wird vollständig methodisch beschrieben und der <u>Aspekt</u> der <u>Gewichtung</u> von <u>Entscheidungskriterien</u> wird benannt.</p>	<p>„Martinas Weg ist, dass sie jeden Kritikpunkt einzeln mit wichtig bis unwichtig bewertet. Die Schokoladen, die gut in den einzelnen Punkten sind, bekommen dementsprechend viele Smilies. Sie multipliziert dann die Gewichtung und die Smilies miteinander und die Schokolade mit der höchsten Punktzahl gewinnt. In diesem Falle Schokolade 4. (Klasse 10) „Martina Entscheidungsweg ist anders als Pauls. Sie erstellt eine Tabelle, wo sie ihre eigenen „Noten“ eintragen kann. Dabei achtet sie auf alle Kriterien und vergibt jeweils Punkte. Für sie fällt die Schokolade, die ein Kriterium nicht erfüllt, nicht sofort raus, sondern wird niedriger bewertet, außerdem ist die Wichtigkeit der einzelnen Kriterien entscheidend. Die Summe der Smilies ist dann letztendlich entscheidend: Je mehr Smilies, desto besser die Schokolade, sprich die Schokolade wird in ihrer Gesamtheit bewertet.“ (Klasse 12) „Martina hat jede Schokolade auf jedes Kriterium hin untersucht und dabei außerdem die unterschiedliche Gewichtung der Kriterien mit einbezogen, wodurch sich dann eine Gesamtpunktzahl für jede einzelne Sorte ergab, nach deren Reihenfolge man dann entscheiden konnte.“ (Studierender)</p>
<p>Angemessener Entscheidungsweg</p>	<p><u>Score 0:</u> Es wird entweder kein Entscheidungsweg oder der von Paul benannt. <u>Score 1:</u> Martinas Entscheidungsweg benannt.</p>	<p>„Pauls Entscheidung.“ (Klasse 6) „Ich denke Pauls Entscheidung.“ (Klasse 8) „Die Entscheidung von Martina.“ (Klasse 8) „Martinas Entscheidung.“ (Klasse 12)</p>

Vorschläge Kompensatorische Strategie („Martina“)	<p><u>Score 0:</u> Es wird kein Verbesserungsvorschlag für Martinas Vorgehensweise gegeben.</p>	<p><i>[Anmerkung: Es wird entweder keine Antwort gegeben oder es wird gesagt, dass die Entscheidung so in Ordnung ist.]</i></p> <p>„Sie sollte es wieder so machen.“ (Klasse 8)</p> <p>„Ich hätte es genauso gemacht.“ (Klasse 8)</p>
	<p><u>Score 1:</u> Es wird ein Verbesserungsvorschlag für Martina gegeben, der sich auf die inhaltlichen Aspekte der Kriterien bezieht.</p>	<p>„Dass sie mal drauf achten sollte, ob die Schokolade fair oder konventionell gehandelt wurde.“ (Klasse 8)</p> <p>„Ich würde ihr raten, noch einmal über den konventionellen Handel ihrer Schokolade nachzudenken.“ (Klasse 10)</p>
	<p><u>Score 2:</u> Es wird ein Verbesserungsvorgehensweise gegeben, der sich entweder auf den Aspekt „unwichtiges Kriterium“ und/oder den Aspekt „der Punktevergabe“ generell bezieht.</p>	<p>„Ich würde ihr raten, die Verfügbarkeit nicht mit einzubeziehen, da sie unwichtig ist.“ (Klasse 8)</p> <p>„Ich halte die Vorgehensweise in dem Punkt nicht sinnvoll, da sie den Angebotsfaktor mit einbezieht und dieser eigentlich kaum eine Rolle spielen sollte.“ (Klasse 12)</p> <p>„Vielleicht nicht nur bis zu 3 Punkten bei den einzelnen Kriterien vergeben, sondern mehr, um besser differenzieren zu können.“ (Klasse 12)</p>

Intuitiv-rechtfertigendes Entscheidungsverhalten („Stefan“)*	<p><u>Score 0:</u> Stefans Vorgehensweise wird nicht beschrieben.</p> <p><u>Score 1:</u> Stefans Vorgehensweise wird inhaltlich beschrieben.</p> <p><u>Score 2:</u> Stefans Vorgehensweise wird methodisch beschrieben.</p>	<p><i>[Anmerkung: Es wird entweder keine Antwort gegeben oder es wird gesagt dass die Entscheidung so in Ordnung ist.]</i></p> <p>„Stefan hat das gut so gemacht“. (Klasse 6)</p> <p><u>Score 1:</u> „Stefan ist das egal ob es konventionell angebaut ist oder biologisch.“ (Klasse 8) „Er geht danach, wo es sie [die Schokolade] zu kaufen gibt und ob sie fair gehandelt ist.“ (Klasse 10) „Stefan hat mehr darauf geachtet, wo man die Schokolade überall kaufen kann und findet es deswegen nicht so schlimm, dass sie 1,09 € kostet und konventionell angebaut ist.“ (Klasse 10)</p> <p><u>Score 2:</u> „Stefan vergleicht die Schokoladen nicht, sondern guckt einfach eine Schokolade aus, die er gut findet.“ (Klasse 8) „Er hat keine Methode, er nimmt das was sich am besten anhört.“ (Klasse 8) „Stefan entscheidet aus dem Bauch heraus und arbeitet nicht nach Prioritätenliste.“ (Klasse 8) „Stefan entscheidet eher aus dem Bauch heraus. Er schaut sich die Schokoladen an, urteilt dann ohne groß darüber nachzudenken.“ (Klasse 12)</p>
<p>*Anmerkung: Neben der non-kompensatorischen („Paul“) und der kompensatorischen Strategie („Martina“) wurde im Fragebogen ein intuitiv-rechtfertigendes Entscheidungsverhalten („Stefan“) präsentiert, das ebenfalls von den Schülerinnen und Schülern reflektiert werden sollte. Charakteristisch hierfür ist, dass eine Entscheidung zunächst spontan und eher intuitiv getroffen wird, die erst im Anschluss durch das Anführen positiver und das Ignorieren negativer Aspekte gerechtfertigt wird.</p>		

Lebenslauf

Sabina Eggert
 Untere Karspüle 6
 37073 Göttingen

Name	Sabina Eggert
Geburtsdatum und -ort	27.11.1974 in Bad Gandersheim
Staatsangehörigkeit	deutsch

1994	Abitur
1994 – 1995	Auslandsaufenthalt an der International School of English, Cambridge, GB
1995 – 1998	Studium der Fächer Biologie und Englisch für das Lehramt an Gymnasien, Georg-August-Universität Göttingen (Abschluss: Zwischenprüfung in Englisch und Biologie)
1998 – 2001	Studium der Fächer Biologie und Englisch für das Lehramt an Gymnasien, Christian-Albrechts-Universität Kiel (Abschluss: Erste Staatsprüfung für die Laufbahn der Studienräte an Gymnasien)
2002 – 2003	Vorbereitungsdienst für die Laufbahn der Studienräte an Gymnasien am IQSH Lübeck (Regionalseminar Ost) (Abschluss: Zweite Staatsprüfung für die Laufbahn der Studienräte an Gymnasien)
Feb. – Juni 2004	Lehrerin im Angestelltenverhältnis (Vollzeit) an der Auguste Viktoria Schule Itzehoe in den Fächern Biologie und Englisch (Sek. I und II.)
seit Sept. 2004	wissenschaftliche Mitarbeiterin in der Abteilung der Didaktik der Biologie, Georg-August-Universität Göttingen
seit Juni 2005	wissenschaftliche Mitarbeiterin und Promotionsstudentin im Rahmen des vom Bundesministerium für Bildung und Forschung geförderten nationalen Projekts „Biologie im Kontext“
seit Juni 2005	assoziiertes Mitglied im DFG Graduiertenkolleg 1195 „Passungsverhältnisse Schulischen Lernens: Verstehen und Optimieren“