

Hartig, Johannes; Höhler, Jana

## **Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen. Projekt MIRT**

*Klieme, Eckhard [Hrsg.]; Leutner, Detlev [Hrsg.]; Kenk, Martina [Hrsg.]: Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Weinheim ; Basel : Beltz 2010, S. 189-198. - (Zeitschrift für Pädagogik, Beiheft; 56)*

urn:nbn:de:0111-opus-34273

in Kooperation mit:

# **BELTZ**

<http://www.beltz.de>

### **Nutzungsbedingungen**

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### **Kontakt:**

**peDOCS**

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Mitglied der Leibniz-Gemeinschaft

Informationszentrum (IZ) Bildung

Schloßstr. 29, D-60486 Frankfurt am Main

eMail: [pedocs@dipf.de](mailto:pedocs@dipf.de)

Internet: [www.pedocs.de](http://www.pedocs.de)

Zeitschrift für Pädagogik · 56. Beiheft

# **Kompetenzmodellierung**

## **Zwischenbilanz des DFG- Schwerpunktprogramms und Perspektiven des Forschungsansatzes**

Herausgegeben von

Eckhard Klieme, Detlev Leutner und Martina Kenk

**BELTZ**

Die in der Zeitschrift veröffentlichten Beiträge sind urheberrechtlich geschützt. Alle Rechte, insbesondere das der Übersetzung in fremde Sprachen, vorbehalten. Kein Teil dieser Zeitschrift darf ohne schriftliche Genehmigung des Verlages in irgendeiner Form – durch Fotokopie, Mikrofilm oder ein anderes Verfahren – reproduziert oder in eine von Maschinen, insbesondere Datenverarbeitungsanlagen, verwendbare Sprache übertragen werden. Auch die Rechte der Wiedergabe durch Vortrag, Funk- und Fernsehsendung, im Magnettonverfahren oder auf ähnlichem Wege bleiben vorbehalten. Fotokopien für den persönlichen oder sonstigen eigenen Gebrauch dürfen nur von einzelnen Beiträgen oder Teilen daraus als Einzelkopie hergestellt werden. Jede im Bereich eines gewerblichen Unternehmens hergestellte oder genutzte Kopie dient gewerblichen Zwecken gem. § 54 (2) UrhG und verpflichtet zur Gebührenzahlung an die VG Wort, Abteilung Wissenschaft, Goethestr. 49, 80336 München, bei der die einzelnen Zahlungsmodalitäten zu erfragen sind.

© 2010 Beltz Verlag · Weinheim und Basel  
Herstellung: Lore Amann  
Gesamtherstellung: Druckhaus „Thomas Müntzer“, Bad Langensalza  
Printed in Germany  
ISSN 0514-2717  
Bestell-Nr. 41157

# Inhaltsverzeichnis

*Eckhard Klieme/Detlev Leutner/Martina Kenk*  
Kompetenzmodellierung. Eine aktuelle Zwischenbilanz des DFG-Schwerpunkt-  
programms. Einleitung zum Beiheft ..... 9

*Benő Csapó*  
Goals of Learning and the Organization of Knowledge ..... 12

## Mathematische Kompetenzen

*Marianne Bayrhuber/Timo Leuders/Regina Bruder/Markus Wirtz*  
Projekt HEUREKO  
Repräsentationswechsel beim Umgang mit Funktionen – Identifikation von  
Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells ..... 28

*Andreas Frey/Nicki-Nils Seitz*  
Projekt MAT  
Multidimensionale adaptive Kompetenzdiagnostik: Ergebnisse zur  
Messeffizienz ..... 40

*Nina Zeuch/Hanneke Geerlings/Heinz Holling/Wim J. van der Linden/  
Jonas P. Bertling*  
Projekt Regelgeleitete Itementwicklung  
Regelgeleitete Konstruktion von statistischen Textaufgaben: Anwendung von  
linear logistischen Testmodellen und Aufgabencloning ..... 52

*Eckhard Klieme/Anika Bürgermeister/Birgit Harks/Werner Blum/Dominik Leiß/  
Katrin Rakoczy*  
Projekt Co<sup>2</sup>CA  
Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht ..... 64

*Olga Kunina-Habenicht/Oliver Wilhelm/Franziska Matthes/André A. Rupp*  
Projekt Kognitive Diagnosemodelle  
Kognitive Diagnosemodelle: Theoretisches Potential und methodische Probleme ... 75

*Aiso Heinze*

Review

Mathematische Kompetenz modellieren und diagnostizieren: Eine Diskussion der Forschungsprojekte des DFG-Schwerpunktprogramms „Kompetenzmodelle“ aus mathematikdidaktischer Sicht ..... 86

## **Naturwissenschaftliche Kompetenzen**

*Tobias Viering/Hans E. Fischer/Knut Neumann*

Projekt Physikalische Kompetenz

Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I ..... 92

*Renate Soellner/Stefan Huber/Norbert Lenartz/Georg Rudinger*

Projekt Gesundheitskompetenz

Facetten der Gesundheitskompetenz – eine Expertenbefragung ..... 104

*Ilonca Hardy/Thilo Kleickmann/Susanne Koerber/Daniela Mayer/*

*Kornelia Möller/Judith Pollmeier/Knut Schwippert/Beate Sodian*

Projekt Science – P

Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter ..... 115

*Nina Roczen/Florian G. Kaiser/Franz X. Bogner*

Projekt Umweltkompetenz

Umweltkompetenz – Modellierung, Entwicklung und Förderung ..... 126

*Ilka Parchmann*

Review

Kompetenzmodellierung in den Naturwissenschaften – Vielfalt ist wertvoll, aber nicht ohne ein gemeinsames Fundament ..... 135

## **Sprachliche und Lesekompetenzen**

*Wolfgang Schnotz/Nele McElvany/Holger Horz/Sascha Schroeder/Mark Ullrich/*

*Jürgen Baumert/Axinja Hachfeld/Tobias Richter*

Projekt BITE

Das BITE-Projekt: Integrative Verarbeitung von Bildern und Texten in der Sekundarstufe I ..... 143

*Tobias Dörfler/Stefanie Golke/Cordula Artelt*

Projekt Dynamisches Testen

Dynamisches Testen der Lesekompetenz: Theoretische Grundlagen, Konzeption und Testentwicklung ..... 154

<i>Thorsten Roick/Petra Stanat/Oliver Dickhäuser/Volker Frederking/ Christel Meier/Lydia Steinhauer</i>	
Projekt Literarästhetische Urteilskompetenz	
Strukturelle und kriteriale Validität der literarästhetischen Urteilskompetenz .....	165

<i>Hans Anand Pant/Simon P. Tiffin-Richards/Olaf Köller</i>	
Projekt Standard-Setting	
Standard-Setting für Kompetenztests im Large-Scale-Assessment .....	175

<i>Johannes Hartig/Jana Höhler</i>	
Projekt MIRT	
Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen .....	189

<i>Albert Bremerich-Vos</i>	
Review	
Modellierung von Aspekten sprachlich-kultureller Kompetenz. Anmerkungen zu den Projektberichten .....	199

## **Fächerübergreifende Kompetenzen**

<i>Ellen Gausmann/Sabina Eggert/Marcus Hasselhorn/Rainer Watermann/ Susanne Bögeholz</i>	
Projekt Bewertungskompetenz	
Wie verarbeiten Schüler/-innen Sachinformationen in Problem- und Entscheidungssituationen Nachhaltiger Entwicklung – Ein Beitrag zur Bewertungskompetenz .....	
	204

<i>Samuel Greiff/Joachim Funke</i>	
Projekt Dynamisches Problemlösen	
Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme .....	
	216

<i>Klaus Lingel/Nora Neuenhaus/Cordula Artelt/Wolfgang Schneider</i>	
Projekt EWIKO	
Metakognitives Wissen in der Sekundarstufe: Konstruktion und Evaluation domänenspezifischer Messverfahren .....	
	228

<i>Jens Fleischer/Joachim Wirth/Stefan Rumann/Detlev Leutner</i>	
Projekt Problemlösen	
Strukturen fächerübergreifender und fachlicher Problemlösekompetenz – Analyse von Aufgabenprofilen .....	
	239

*Melanie Schütte/Joachim Wirth/Detlev Leutner*

Projekt Selbstregulationskompetenz

Selbstregulationskompetenz beim Lernen aus Sachtexten – Entwicklung und  
Evaluation eines Kompetenzstrukturmodells ..... 249

*Tobias Gschwendtner/Bernd Geißel/Reinhold Nickolaus*

Projekt Berufspädagogik

Modellierung beruflicher Fachkompetenz in der gewerblich-technischen  
Grundbildung ..... 258

*Franziska Perels*

Review

Modellierung und Messung fächerübergreifender Kompetenzen und ihre  
Bedeutung für die Bildungsforschung. Kritische Reflexion der Projektbeiträge ... 270

## **Lehrerkompetenzen**

*Simone Bruder/Julia Klug/Silke Hertel/Bernhard Schmitz*

Projekt Beratungskompetenz

Modellierung der Beratungskompetenz von Lehrkräften ..... 274

*Cornelia Gräsel/Sabine Krolak-Schwerdt/Ines Nölle/Thomas Hörstermann*

Projekt Diagnostische Kompetenz

Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der  
Übergangsempfehlung: eine Analyse aus der Perspektive der sozialen  
Urteilsbildung ..... 286

*Tina Seidel/Geraldine Blomberg/Kathleen Stürmer*

Projekt OBSERVE

„OBSERVER“ – Validierung eines videobasierten Instruments zur Erfassung  
der professionellen Wahrnehmung von Unterricht ..... 296

*Mareike Kunter*

Review

Modellierung von Lehrerkompetenzen. Kommentierung der  
Projektdarstellungen ..... 307

Johannes Hartig/Jana Höhler

# Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen

Projekt MIRT<sup>1</sup>

## 1. Einleitung

Kompetenzen werden im Rahmen des Schwerpunktprogramms 1293 als kontextabhängige, kognitive Leistungsdispositionen definiert (vgl. auch Weinert 2001; Hartig/Klieme 2006). Im Sinne dieser Definition kann eine Person mit einer spezifischen Kompetenz *Anforderungen in einem spezifischen Bereich von Situationen* erfolgreich bewältigen. Die zu erfassende Kompetenz ist hierbei ein *theoretisches Konstrukt* zur Beschreibung und Erklärung von Leistungsunterschieden in diesen Situationen. Wenn Kompetenzen in standardisierten Tests erfasst werden, stellt das Kompetenzkonstrukt die theoretische Verbindung zwischen dem Testverhalten und den daraus gezogenen diagnostischen Schlüssen dar. *Psychometrische Modelle* oder *Messmodelle* sind in diesem Zusammenhang das Werkzeug, um diese theoretische Verbindung in konkrete Messverfahren umzusetzen (vgl. z.B. Mislevy u.a. 2002; Wilson 2005).

In psychometrischen Modellen werden Annahmen über die Struktur des interessierenden Konstrukts sowie über die Zusammenhänge zwischen dem Konstrukt und dem beobachtbaren Testverhalten formuliert. Idealerweise sollte in einem psychometrischen Modell abgebildet werden, wie konkrete *situative Anforderungen* und *individuelle Ressourcen* bei der Bewältigung dieser Anforderungen interagieren (vgl. Embretson 1983; Hartig 2008). Die zu messenden Konstrukte werden in psychometrischen Modellen durch *latente Variablen* repräsentiert (vgl. Borsboom/Mellenbergh/van Heerden 2003). Diese Variablen dienen der Erklärung der Verteilungen und Zusammenhangsstrukturen beobachtbarer Variablen, die aus dem Testverhalten gebildet werden (vgl. Skrandal/Rabe-Hesketh 2004). Umgekehrt wird auf Basis des gewählten psychometrischen Modells vom beobachtbaren Testverhalten auf individuelle Ausprägungen oder Populationsverteilungen der latenten Variablen geschlossen.

Modelle mit latenten Variablen unterscheiden sich unter anderem hinsichtlich des Skalenniveaus der beobachteten Indikatoren und der latenten Variablen. Modelle mit *kategorialen Indikatoren* (vgl. z.B. Aufgabe gelöst/nicht gelöst) und *kontinuierlichen latenten Variablen*, auf die in diesem Forschungsvorhaben fokussiert wird, werden zum Beispiel traditionell unter dem Begriff *Item-Response-Theorie* (IRT) zusammengefasst. In den letzten Jahren wird die separate Behandlung dieser Modelle zunehmend zugun-

1 Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennz.: HA 5050/2-1) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).



ten einer breiteren und flexibleren Sichtweise aufgegeben. Spezifische Modelle stellen in einem solchen breiteren Rahmen Spezialfälle dar, die für bestimmte Arten von beobachteten Variablen geeignet sind (vgl. Muthén 2002; Rabe-Hesketh/Skrondal/Pickles 2004; Skrondal/Rabe-Hesketh 2004).

Die Modellierung von Kompetenzen bezieht *quantitative* und *qualitative* Aspekte der Beschreibung von interindividuellen Unterschieden mit ein. Der erste *qualitative* Aspekt ist die Differenzierung von (Sub-) Dimensionen, die durch verschiedene latente Variablen repräsentiert werden. Die interindividuellen Unterschiede in den latenten Variablen stellen den *quantitativen* Aspekt der Modellierung dar. Diese beiden Aspekte werden in den folgenden Abschnitten anhand erster Projektergebnisse näher erörtert. Zusätzlich wird oft angestrebt, Personen mit unterschiedlich stark ausgeprägten (Teil-) Kompetenzen dahingehend zu beschreiben, welche situativen Anforderungen sie bewältigen können. Diese *kriterienorientierte* Beschreibung, die oft mit der Definition von „Kompetenzniveaus“ oder „Kompetenzstufen“ erreicht wird, ist der zweite *qualitative* Aspekt bei der psychometrischen Modellierung von Kompetenzen (vgl. z.B. Hartig 2007, 2008).

Psychometrische Modelle können ein- oder mehrdimensional sein. In der pädagogisch-psychologischen Diagnostik überwiegt die Anwendung eindimensionaler Messmodelle, die zur Modellierung von Unterschieden zwischen Personen eine einzelne latente Variable beinhalten (vgl. z.B. Gonzalez 2003; Gonzalez/Galia/Li 2004; Hartig/Jude/Wagner 2008). Zur psychometrischen Modellierung der in einem Test erfassten Kompetenz können jedoch auch mehrdimensionale psychometrische Modelle, insbesondere mehrdimensionale IRT-Modelle (MIRT-Modelle) verwendet werden, in denen das Kompetenzkonstrukt differenziert hinsichtlich mehrerer zugrunde liegender Teilkompetenzen modelliert wird (vgl. z.B. Ackerman/Gierl/Walker 2003; Adams/Wilson/Wang 1997; Hartig/Höhler 2008; McDonald 1997; Reckase 2007, 2009). Häufig werden einfache eindimensionale Modelle eher aus pragmatischen als aus theoretischen Gründen gewählt. Verglichen mit eindimensionalen Modellen kann die Erfassung einer Kompetenz mittels eines mehrdimensionalen psychometrischen Modells eine differenziertere Diagnostik und zugleich eine Prüfung von Annahmen über die Struktur der erfassten Kompetenz und Teilkompetenzen ermöglichen (vgl. z.B. Ackerman/Gierl/Walker 2003; Walker/Beretvas 2003).

Der vorliegende Artikel skizziert zunächst kurz die Zielsetzung des DFG-Projekts „Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen“. Im darauf folgenden Abschnitt werden zwei wesentliche Charakteristika zur Unterscheidung von MIRT-Modellen sowie deren inhaltliche Implikationen für die Modellierung von Kompetenzen dargestellt.

## 2. Zielsetzung

Die zentrale Fragestellung des DFG-Projekts „Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen“ ist, inwieweit sich sprachliche Kompetenzen, die mit demselben Aufgabenmaterial erfasst wurden, auf unterschiedliche Weise modellie-

ren lassen und welche diagnostischen Aussagen aus den verschiedenen Auswertungen resultieren. Insbesondere interessiert, inwieweit komplexere mehrdimensionale Modelle gegenüber eindimensionalen Modellen (z.B. dem eindimensionalen Raschmodell) einen Zugewinn an diagnostischer Information liefern.

In dem Projekt wurden zunächst verschiedene IRT-Modelle auf dieselben Daten angewendet. Dafür wurde auf bereits vorliegende Leistungsdaten der DESI-Studie (Deutsch Englisch Schülerleistungen International) (vgl. Beck/Klieme 2007; Klieme u.a. 2008) zurückgegriffen. Verwendet werden Daten aus den Tests für Sprachbewusstheit (Grammatik), Lese- und Hörverstehen für Englisch als Fremdsprache zum Ende der neunten Jahrgangsstufe.

### 3. Charakteristika mehrdimensionaler IRT-Modelle

Im Weiteren werden zwei wesentliche Charakteristika mehrdimensionaler IRT-Modelle beschrieben, die weitreichende inhaltliche Konsequenzen für die Interpretation der latenten Variablen im Modell haben. Zunächst wird auf die Komplexität des Ladungsmusters eingegangen (Between- versus Within-Item-Mehrdimensionalität), im darauf folgenden Abschnitt auf kompensatorische versus nicht-kompensatorische Verknüpfungen mehrerer latenter Variablen bei komplexer Ladungsstruktur.

#### 3.1 *Between- versus Within-Item-Mehrdimensionalität*

In den meisten empirischen Anwendungen von MIRT-Modellen zur Messung von Schülerkompetenzen wird für jede zu erfassende Kompetenz jeweils eine separate Dimension (latente Variable) definiert. Im Kontext der PISA-Studien findet sich auch eine mehrdimensionale Modellierung von Subdimensionen *innerhalb* einzelner Kompetenzbereiche (vgl. z.B. Blum u.a. 2004). In beiden Fällen lädt jedes Item nur auf einer Dimension. Die Modelle sind durch eine *Einfachstruktur* oder *Between-Item-Mehrdimensionalität* (vgl. Adams/Wilson/Wang 1997) gekennzeichnet. Die latenten Variablen repräsentieren in diesen Modellen das Leistungsniveau in den durch die jeweiligen Item-Mengen definierten Bereichen.

MIRT-Modelle können jedoch auch eine komplexe Ladungsstruktur (*Within-Item-Mehrdimensionalität*) haben, bei der Items gleichzeitig durch mehrere latente Dimensionen beeinflusst werden. Bei dieser Modellierung können die latenten Variablen verschiedene Fähigkeiten repräsentieren, die bei der Leistung in bestimmten inhaltlichen Bereichen gemeinsam erforderlich sind. Für die Definition einer solchen komplexen Ladungsstruktur sind theoretisch fundierte Vorannahmen über Eigenschaften der Aufgaben und die inhaltliche Bedeutung der unterschiedlichen latenten Variablen notwendig. Walker und Beretvas (2003) analysierten zum Beispiel Daten aus einem Test zur Erfassung mathematischer Fähigkeiten mit einem zweidimensionalen IRT-Modell, in dem alle Aufgaben Indikatoren für eine allgemeine mathematische Fähigkeit darstell-

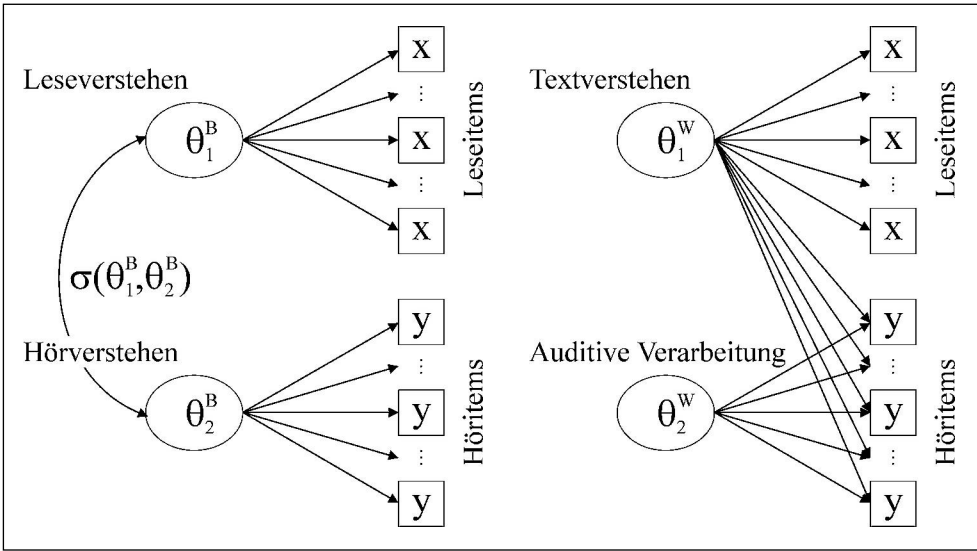


Abb. 1: Schematische Darstellung der Modellierung von Testaufgaben  $x, y$  mit (a) einem zweidimensionalen IRT-Modell mit Einfachstruktur und (b) einem zweidimensionalen IRT-Modell mit Mehrfachladungen (Quelle: Hartig/Höhler 2008).

ten. Aufgaben mit offenem Antwortformat wurden zusätzlich als Indikatoren der zweiten Dimension modelliert, die als die Fähigkeit zur Kommunikation mathematischer Inhalte interpretiert wurde. Hartig und Höhler (2008) zeigen am Beispiel von Lese- und Hörverstehensleistungen in Englisch als Fremdsprache, dass ein solches Generalfaktorenmodell in ein Modell mit Einfachstruktur überführbar ist, in dem beide Item-Gruppen auf separaten Dimensionen laden (vgl. Abbildung 1).

Eine Entscheidung zugunsten einer der beiden in Abbildung 1 dargestellten Modellvarianten auf Basis der Anpassung an die empirischen Daten ist nicht möglich, unter bestimmten Umständen können beide Modelle hinsichtlich der Anpassung sogar äquivalent sein (ebd.). Die inhaltliche Bedeutung der latenten Variablen ist jedoch grundsätzlich verschieden. Während die latenten Variablen des Modells mit Einfachstruktur direkt die Teilkompetenzen Lese- und Hörverstehen repräsentieren, ist die Interpretation der latenten Variablen im Modell mit Mehrfachladungen komplizierter. Hier repräsentiert die erste Dimension im Sinne eines Generalfaktors die Teilkompetenzen, die notwendig sind, um die *gemeinsamen Anforderungen* von Lese- und Hörverstehensaufgaben zu bewältigen. Die zweite Dimension steht hingegen für die spezifischen Kompetenzen, die *zusätzlich* für die erfolgreiche Bearbeitung von Hörverstehensaufgaben benötigt werden.

Durch die unterschiedliche Bedeutung der latenten Variablen in diesen beiden Modellierungsansätzen ergeben sich bedeutsame Implikationen für die Interpretation und Kommunikation von Fähigkeitsprofilen. So konnten Hartig und Höhler (ebd.) am Beispiel von Lese- und Hörverstehen in Englisch als Fremdsprache zeigen, dass das Modell

mit Einfachstruktur in nahezu gleichen Geschlechts- und Bildungsgangprofilen für beide Dimensionen resultierte. Dagegen sind in dem Modell mit Mehrfachladungen für die zweite, hörverstehen-spezifische Dimension die Unterschiede zwischen Bildungsgängen weniger stark ausgeprägt als für den für alle Items gemeinsamen Generalfaktor. Darüber hinaus zeigte sich für das Modell mit Mehrfachladungen, dass Mädchen im Generalfaktor zwar höhere Werte aufweisen als Jungen, dieser Effekt sich jedoch bezüglich der Leistung in der hörverstehen-spezifischen Dimension umkehrt; hier zeigen Jungen eine bessere Leistung als Mädchen.

In MIRT-Modellen mit Mehrfachladungen können Annahmen über Interaktionen zwischen den verschiedenen Teilkompetenzen und Aufgabenanforderungen überprüft werden. Solche Modelle implizieren explizite Annahmen über die Fähigkeiten, die benötigt werden, um ein Item lösen zu können und darüber, ob diese Fähigkeiten kompensatorisch verknüpft sind oder nicht (vgl. nächster Abschnitt). Besonders interessant sind Modelle mit Mehrfachladungen für die Kompetenzmodellierung bei komplexen Aufgaben, deren Lösung nicht mit einer Fähigkeitsdimension für jede Aufgabe erklärt werden kann. Auch für Fragestellungen, die sich auf spezifische Teilkompetenzen einer Kompetenzdomäne beziehen, kann dieser Modellierungsansatz eine vielversprechende Alternative darstellen.

Die geringere Komplexität und einfache Interpretation der latenten Variablen in Modellen mit Einfachstruktur kann aber auch vorteilhaft sein: In diesen Modellen stellen die geschätzten Fähigkeitswerte für die latenten Dimensionen direkte Leistungsmaße des zugrunde liegenden Items dar. In vielen Fällen werden die Leistungsmaße verschiedener Teilkompetenzen (z.B. Lese- und Hörverstehen) hoch korreliert sein, da zur erfolgreichen Bearbeitung aller Items auch gemeinsame Fähigkeiten benötigt werden. Diese Gewichtung spezifischer Fähigkeiten und deren genaue Interaktion sind jedoch nicht unbedingt von Interesse. In einem Modell mit Einfachstruktur repräsentieren die latenten Variablen die notwendige Kombination aller Fähigkeiten, die zur erfolgreichen Bearbeitung komplexer Aufgaben benötigt wird. Die Gemeinsamkeiten kommen in der Höhe der Korrelationen zwischen den latenten Variablen zum Ausdruck. Wenn das Forschungsinteresse also hauptsächlich darin besteht, deskriptive Leistungsmaße zu erhalten, unabhängig davon, wie diese miteinander interagieren, sind Modelle mit Einfachstruktur angemessener als solche mit Mehrfachladungen.

### 3.2 *Kompensatorische versus nicht-kompensatorische Verknüpfung von Dimensionen*

Bei Modellen mit Mehrfachladungen können die Lösungswahrscheinlichkeiten einzelner Items von mehreren Fähigkeitsdimensionen abhängen. Diese Fähigkeitsdimensionen können unterschiedlich integriert werden. Eine grundsätzliche Frage ist, ob die latenten Variablen kompensatorisch oder nicht-kompensatorisch miteinander verknüpft sind (vgl. Hartig/Höhler 2008). Die meisten MIRT-Modelle mit Mehrfachladungen sind kompensatorisch, eine geringe Fähigkeitsausprägung in einer Dimension kann durch

eine hohe Ausprägung in einer zweiten Dimension ausgeglichen werden und umgekehrt. In einem nicht-kompensatorischen Modell wird für die erfolgreiche Bearbeitung einer komplexen Aufgabe eine hohe Fähigkeitsausprägung in allen Dimensionen benötigt. Der Unterschied zwischen beiden Verknüpfungen kann an zwei einfachen Modellen veranschaulicht werden. Die *Item-Response-Funktion* (IRF) für ein Item im zweidimensionalen Raschmodell mit einer Itemladung auf beiden Dimensionen und einer Itemschwierigkeit von Null ist:

$$\Pr(x_{vi}|\theta_1, \theta_2) = \text{logit}(\theta_1 + \theta_2) \quad (1)$$

mit

$$\text{logit}(y) \equiv \frac{e^y}{1 + e^y} \quad (2)$$

In diesem Modell hängt die Lösungswahrscheinlichkeit von der Summe beider Fähigkeitsdimensionen  $\theta_1$  und  $\theta_2$  ab. Da eine Fähigkeit die andere kompensieren kann, resultiert die gleiche Summe  $\theta_1 + \theta_2$  in der gleichen Wahrscheinlichkeit, ein Item richtig zu lösen.

Das *multicomponent latent trait model* (MLTM) von Embretson (1984) ist ein partiell nicht-kompensatorisches MIRT-Modell. Die IRF für ein Item in einer vereinfachten zweidimensionalen Version des MLTM, mit Komponentenschwierigkeiten und Rateparametern von null ist:

$$\Pr(x_{vi}|\theta_1, \theta_2) = \text{logit}(\theta_1) \cdot \text{logit}(\theta_2) \quad (3)$$

In dem nicht-kompensatorischen MLTM hängt die Lösungswahrscheinlichkeit von dem Produkt der Logits beider Fähigkeitsdimensionen  $\theta_1$  und  $\theta_2$  ab. Somit ist die Lösungswahrscheinlichkeit eines Items nur dann hoch, wenn alle dafür benötigten Fähigkeiten hoch ausgeprägt sind. In Abbildung 2 sind die IRFs für das kompensatorische Modell (vgl. Gleichung 1) und das nicht-kompensatorische Modell (vgl. Gleichung 2) graphisch dargestellt.

Durch die graphische Veranschaulichung wird deutlich, dass das nicht-kompensatorische Modell verglichen mit dem kompensatorischen Modell in allgemein geringer ausgeprägten Lösungswahrscheinlichkeiten resultiert. In dem nicht-kompensatorischen Modell sind die Anforderungen an die Fähigkeiten der getesteten Person höher. Die IRFs beider Modelle weichen besonders stark in Bereichen voneinander ab, in denen sich beide Dimensionen  $\theta_1$  und  $\theta_2$  unterscheiden: Wenn eine der beiden Dimensionen gering ausgeprägt ist, geht die Lösungswahrscheinlichkeit im nicht-kompensatorischen Modell gegen Null. Im kompensatorischen Modell kann die Lösungswahrscheinlichkeit dagegen substantiell von Null abweichen, wenn die Fähigkeit in einer Dimension ge-

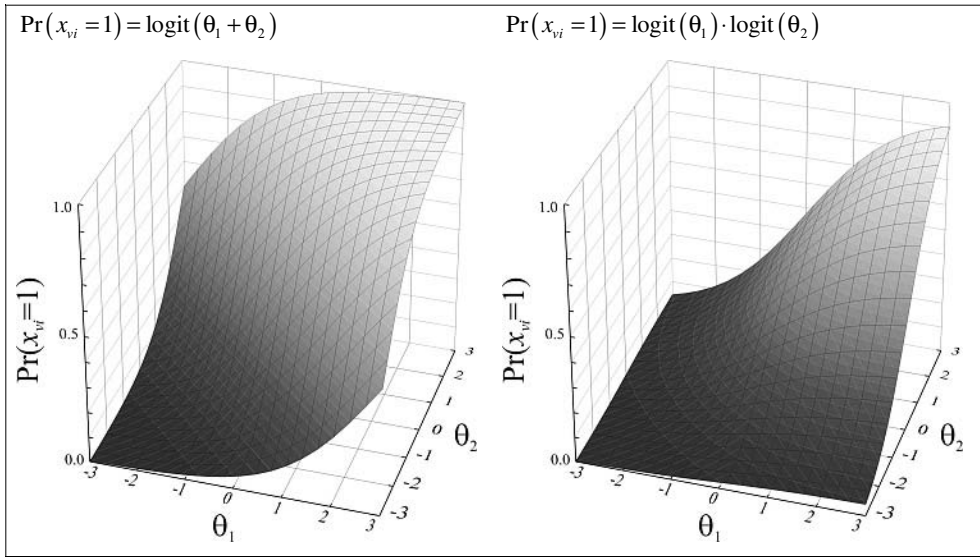


Abb. 2: IRFs für ein zweidimensionales kompensatorisches (links) und ein zweidimensionales nicht-kompensatorisches (rechts) Modell (Quelle: Hartig/Höhler eingereicht)

ring, aber in der anderen Dimension hoch ausgeprägt ist. Für die Bereiche, in denen  $\theta_1$  und  $\theta_2$  etwa gleich ausgeprägt sind, unterscheiden sich die beiden Modelle kaum.

Die Wahl einer kompensatorischen oder nicht-kompensatorischen Funktion für die Verknüpfung mehrerer latenter Variablen hat entscheidende Implikationen für die Bedeutung dieser Variablen. Eine kompensatorische Verknüpfung kann beispielsweise adäquat sein, wenn die latenten Variablen unterschiedliche Lösungsstrategien für die erfolgreiche Bearbeitung der Testaufgaben repräsentieren. Die Lösungswahrscheinlichkeit sollte hoch sein, wenn irgendeine Strategie ausreichend gut beherrscht wird und noch höher, wenn mehrere Lösungsstrategien zur Verfügung stehen.

Eine nicht-kompensatorische IRF kann dagegen angemessen sein, wenn die latenten Dimensionen Fähigkeiten repräsentieren, die unterschiedlichen (kognitiven) Operationen zugrunde liegen. Die grundlegende Annahme ist hier, dass für eine erfolgreiche Itembearbeitung alle dieser Operationen durchgeführt werden müssen (z.B. bei aufeinander aufbauenden Schritten eines Lösungsprozesses). Wenn eine dieser Fähigkeiten gering ausgeprägt ist, ist die erfolgreiche Bearbeitung eines entsprechenden Items stark erschwert, unabhängig von der Fähigkeitsausprägung in den anderen Dimensionen. Eine hohe Lösungswahrscheinlichkeit ergibt sich erst, wenn alle benötigten Fähigkeiten hoch ausgeprägt sind.

Die Anzahl der latenten Dimensionen hat unterschiedliche Implikationen für die Interpretation der Werte in  $\theta$  für die verschiedenen Modelle. In nicht-kompensatorischen Modellen mit Mehrfachladungen verändert sich die Skalierung der Fähigkeitschätzungen, wenn weitere Dimensionen in das Modell mit aufgenommen werden. In kompen-

satorischen Modellen ist die Skalierung der einzelnen Dimensionen dagegen unabhängig von zusätzlichen Dimensionen (vgl. Reckase 2007). Trotz der mathematischen und theoretischen Unterschiedlichkeit dieser beiden Modellierungsansätze gibt es Hinweise darauf, dass kompensatorische und nicht-kompensatorische Modelle mit der gleichen Anzahl latenter Dimensionen eine sehr ähnliche Anpassung an empirische Daten aufweisen. In Abbildung 2 wird deutlich, dass die IRFs gleich sind für Bereiche, in denen die Dichte einer bivariaten Verteilung bei positiv korrelierten Dimensionen am größten ist (vgl. ebd.; Spray u.a. 1990). Für positiv korrelierte Dimensionen sollten sich daher zwischen den beiden Modellen keine bedeutsamen Unterschiede bezogen auf die empirische Anpassungsgüte ergeben. Deutliche Unterschiede sind hingegen bei negativ korrelierten latenten Variablen zu erwarten.

#### 4. Schlussfolgerungen

Insbesondere für die Modellierung von Kompetenzen stellen MIRT-Modelle eine vielversprechende Methodologie dar. Es existiert bereits eine Bandbreite von Modellen, die für vielfältige Fragestellungen (nicht nur) innerhalb der empirischen Bildungsforschung jeweils passend sein können. Durch vertiefende Analysen mit komplexeren Modellen der IRT können zusätzliche Informationen über die Fähigkeiten der getesteten Individuen gewonnen werden. Aber auch theoretische Annahmen über spezifische (Teil-)Kompetenzen und darüber, wie diese miteinander interagieren, können mit den verschiedenen MIRT-Modellen getestet werden. Schon bei der theoretischen Entwicklung von Kompetenzmodellen sollte die psychometrische Umsetzung bedacht werden. So kann eine theoriegeleitete Itemkonstruktion bei gleichzeitiger Formulierung eines Messmodells gewinnbringend für eine möglichst differenzierte Interpretation und Kommunikation von Schülerkompetenzen eingesetzt werden. Von solchen Erkenntnissen können nicht nur die Schüler/innen, sondern auch Pädagogen/innen, Erzieher/innen und Entscheidungsträger/innen profitieren. Diese Aufgabe erfordert jedoch den möglichst frühen interdisziplinären Austausch zwischen Didaktik und Psychometrie. In der zweiten Förderphase des DFG-Projekts „Modellierung von Kompetenzen mit mehrdimensionalen IRT-Modellen“ ist die Neuentwicklung von Testaufgaben geplant, bei der die mögliche Dimensionalität der Aufgaben schon bei der Itemkonstruktion systematisch berücksichtigt werden soll.

#### Literatur

- Ackerman, T.A./Gierl, M.J./Walker, C.M. (2003): Using multidimensional item response theory to evaluate educational and psychological tests. In: *Educational Measurement: Issues and Practice* 22, S. 37–53.
- Adams, R./Wilson, M./Wang, W.-C. (1997): The multidimensional random coefficients multinomial logit model. In: *Applied Psychological Measurement* 21, S. 1–32.
- Beck, B./Klieme, E. (Hrsg.) (2007): *Sprachliche Kompetenzen. Konzepte und Messung*. Weinheim/Basel: Beltz.

- Blum, W./Neubrand, M./Ehmke, T./Senkbeil, M./Jordan, A./Ulfig, F./Carstensen, C.H. (2004): Mathematische Kompetenz. In: Prenzel, M./Baumert, J./Blum, W./Lehmann, R./Leutner, D./Neubrand, M./Pekrun, R./Rolff, H.-G./Rost, J./Schiefele, U. (Hrsg.): PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann, S. 47–92.
- Borsboom, D./Mellenbergh, G.J./van Heerden, J. (2003): The theoretical status of latent variables. In: *Psychological Review* 110, S. 203–219.
- Embretson, S.E. (1983): Construct validity: Construct representation versus nomothetic span. In: *Psychological Bulletin* 93, S. 179–197.
- Embretson, S.E. (1984): A general latent trait model for response processes. In: *Psychometrika* 49, S. 175–186.
- Gonzalez, E.J. (2003): Scaling the PIRLS reading assessment data. In: Martin, M.O./Mullis, I.V.S./Kennedy, A.M. (Hrsg.): PIRLS 2001 technical report. Chestnut Hill, MA: Boston College.
- Gonzalez, E.J./Galia, J./Li, I. (2004): Scaling methods and procedures for the TIMSS 2003 mathematics and science scales. In: Martin, M.O./Mullis, I.V.S./Chrostowski S.J. (Hrsg.): TIMSS 2003 Technical Report. International Association for the Evaluation of Educational Achievement.
- Hartig, J. (2007): Skalierung und Definition von Kompetenzniveaus. In: Beck, B./Klieme, E. (Hrsg.): Sprachliche Kompetenzen. Konzepte und Messung. DESI-Ergebnisse Band 1. Weinheim/Basel: Beltz, S. 83–99.
- Hartig, J. (2008): Psychometric models for the assessment of competencies. In: Hartig, J./Klieme, E./Leutner, D. (Hrsg.): Assessment of competencies in educational contexts: state of the art and future prospects. Göttingen: Hogrefe & Huber Publishers, S. 69–90.
- Hartig, J./Höhler, J. (eingereicht): Multidimensional IRT models for the assessment of competencies. In: *Studies in Educational Evaluation*.
- Hartig, J./Höhler, J. (2008): Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. In: *Zeitschrift für Psychologie/Journal of Psychology* 216, S. 89–101.
- Hartig, J./Jude, N./Wagner, W. (2008): Methodische Grundlagen der Messung sprachlicher Kompetenzen. In: Klieme, E./Eichler, W./Lehmann, R.H./Nold, G./Schröder, K./Thomé, G./Willenberg, H. (Hrsg.): Sprachliche Kompetenzen. Leistungsverteilungen und Bedingungsfaktoren. DESI-Ergebnisse Band 2. Weinheim: Beltz Pädagogik, S. 34–54.
- Hartig, J./Klieme, E. (2006): Kompetenz und Kompetenzdiagnostik. In: Schweizer, K. (Hrsg.): Leistung und Leistungsdiagnostik. Berlin: Springer, S. 127–143.
- Klieme, E./Eichler, W./Helmke, A./Lehmann, R.H./Nold, G./Rolff, H.-G. u.a. (2008): Unterricht und Kompetenzerwerb in Deutsch und Englisch: Ergebnisse der DESI-Studie. Weinheim: Beltz.
- McDonald, R.P. (1997): Normal-ogive multidimensional model. In: van der Linden, W.J./Hambleton, R.K. (Hrsg.): Handbook of modern item response theory. New York City, NY: Springer-Verlag, S. 257–269.
- Mislevy, R.J./Wilson, M./Ercikan, K./Chudowsky, N. (2002): Psychometric principles in student assessment. CSE technical report 583. Los Angeles, CA: National Center for Research on Evaluation, Standards, and Student Testing (CRESST).
- Muthén, B. (2002): Beyond SEM: general latent variable modeling. In: *Behaviormetrika* 29, S. 81–117.
- Rabe-Hesketh, S./Skrodingal, A./Pickles, A. (2004): Generalized multilevel structural equation modelling. In: *Psychometrika* 69, S. 167–190.
- Reckase, M.D. (2007): Multidimensional item response theory. In: Sinharay, S./Rao, C.R. (Hrsg.): Handbook of statistics, Vol. 26: Psychometrics. Amsterdam: Elsevier, S. 607–642.
- Reckase, M.D. (2009): Multidimensional item response theory. New York: Springer Verlag.



- Skrondal, A./Rabe-Hesketh, S. (2004): Generalized latent variable modeling. Multilevel, longitudinal and structural equation models. Boca Raton u.a.: Chapman & Hall.
- Spray, J.A./Davey, T.C./Reckase, M.D./Ackerman, T.A./Carlson, J.E. (1990): Comparison of two logistic multidimensional item response theory models. Research Report ONR90-8. ACT Inc., Iowa City, IA.
- Walker, C.M./Beretvas, S.N. (2003): Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. In: Journal of Educational Measurement 40, S. 255–275.
- Weinert, F.E. (2001): Concept of competence: a conceptual clarification. In: Rychen, D.S./Salganik, L.H. (Hrsg.): Defining and selecting key competencies. Seattle: Hogrefe & Huber Publishers, S. 45–65.
- Wilson, M. (2005): Constructing measures. An item response modelling approach. Mahwah: Lawrence Erlbaum Associates.

### **Anschrift des Autors/der Autorin**

Prof. Dr. Johannes Hartig, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Bildungsqualität und Evaluation, Schloßstr. 29, D-60486 Frankfurt a.M.  
E-Mail: hartig@dipf.de

Jana Höhler, Deutsches Institut für Internationale Pädagogische Forschung (DIPF), Bildungsqualität und Evaluation, Schloßstraße 29, D-60486 Frankfurt a.M.  
E-Mail: hoehler@dipf.de