

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

Klieme, Eckhard [Hrsg.]; Leutner, Detlev [Hrsg.]; Kenk, Martina [Hrsg.]: Kompetenzmodellierung. Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes. Weinheim ; Basel : Beltz 2010, S. 75-85. - (Zeitschrift für Pädagogik, Beiheft; 56)

urn:nbn:de:0111-opus-33635

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

Internet: 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. § 5 4(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 Co²CA
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

Olga Kunina-Habenicht/Oliver Wilhelm/Franziska Matthes/André A. Rupp

Kognitive Diagnosemodelle: Theoretisches Potential und methodische Probleme

Projekt Kognitive Diagnosemodelle¹

1. Kognitive Diagnosemodelle: Eine Einführung

Im Rahmen dieses Forschungsprojekts werden das theoretische Potential und die Schwierigkeiten neuartiger psychometrischer Modelle, sog. kognitiver Diagnosemodelle (CDMs) untersucht. Mit der Verwendung der CDMs sind im Wesentlichen drei Hoffnungen verknüpft. Zum einen sollen sie eine mehrdimensionale Abbildung der Kompetenzen ermöglichen, die auf einer theoretisch begründeten und vorab festgelegten Klassifikation der Aufgaben zu entsprechenden Fähigkeitsdimensionen basiert. Dies wird am Beispiel einiger Mathematikaufgaben erläutert.

Damit ein Schüler² eine bestimmte Aufgabe in einem Mathematiktest erfolgreich bearbeiten kann, muss er mehrere Teilprozesse korrekt ausführen. Das Lösen einer Sachaufgabe erfordert bspw. die Fertigkeit zum Leseverständnis und Modellieren. Modellieren bezeichnet die mentale Konstruktion der Problemsituation (des sog. realen Modells bzw. des Situationsmodells) und deren anschließende Übertragung in ein mathematisches Modell mit einer oder mehreren Operationen (vgl. Blum u.a. 2006). Die aufgestellte mathematische Rechnung muss schließlich durch die korrekte Anwendung von Rechenalgorithmen gelöst werden. Dabei können verschiedene Merkmale der Aufgabe eine Rolle spielen (z.B. kann in der Rechnung ein Zehner-Übertrag gefordert sein). Alle hier genannten Teilprozesse der Aufgabebearbeitung lassen sich als mathematische *Teilfertigkeiten* auffassen.

Die für die Lösung der Aufgabe erforderlichen Teilkompetenzen werden in einer *Q*-Matrix repräsentiert (s. Tab. 1). Ist für eine Aufgabe nur ein Eintrag in der *Q*-Matrix verzeichnet, so fordert die entsprechende Aufgabe lediglich die Beherrschung eines der postulierten Teilprozesse. Aufgaben, die die Ausführung von mehr als nur einem Teilprozess verlangen, weisen multiple Einträge in der *Q*-Matrix – sog. *Mehrfachladungen* – auf. Eine *Q*-Matrix beschreibt somit ein Klassifikationssystem von Teilfertigkeiten einer übergeordneten Kompetenz.

Die zweite Erwartung an CDMs ist, dass sie eine statistisch begründete Klassifikation der Schüler bzgl. einer oder mehreren Kompetenzen in sog. Kompetenzprofilen er-

1 Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennz.: RU 1424/3-1) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).

2 Im Folgenden wird für Personenbezeichnungen beider Geschlechter nur die männliche Form verwendet.

lauben. Um dies zu ermöglichen, wird dem Schüler für jede postulierte Teilfertigkeit eine Ausprägung auf einer kategorialen Variable zugewiesen. Das bedeutet konkret, dass die Schüler für jede Teilfertigkeit in zwei Fähigkeitsgruppen klassifiziert werden (Schüler A beherrscht die Fähigkeit (sog. „masters“) oder Schüler A beherrscht die Fähigkeit nicht (sog. „non-masters“). Eine solche statistische Klassifikation stellt selbstverständlich eine Vereinfachung dar. Mit der Formulierung „Schüler beherrscht die Fähigkeit in einer Grundrechenart“ meinen wir, dass dieser Schüler in einem bestimmten Zahlenraum sicher Rechnungen dieser Grundrechenart (ggf. mit Übertrag und verschiedenen Positionen der unbekanntenen Operatoren) im Kopf ausrechnen kann. Er kann darüber hinaus diese Grundrechenart flexibel zur Lösung von Sachaufgaben anwenden und erkennen, wann eine Anwendung der inversen Operation erforderlich ist.

Schüler, die durch die CDMs als „non-master“ eingestuft werden, können zwar einzelne Rechenaufgaben ausrechnen, haben jedoch Schwierigkeiten, die korrekte mathematische Operation zur Lösung von Sachaufgaben einzusetzen. Sie haben darüber hinaus Probleme zu erkennen, in welchen Fällen die Anwendung der inversen Operation erforderlich ist.

Über alle Teilfertigkeiten hinweg ergibt sich so ein Kompetenzprofil des Schülers, das an Schüler, Lehrkräfte und Eltern zurückgemeldet werden kann. Es wird somit möglich, auf spezifische Schwierigkeiten des Schülers einzugehen und weiteres Lernen optimiert zu fördern.

Schüler mit identischen Kompetenzprofilen werden in gleiche Kompetenzklassen zusammengefasst. Aus theoretischer Sicht ist denkbar, dass bestimmten Klassen mehr Schüler zugeordnet werden als anderen Klassen. Bspw. könnte es eine Kompetenzklasse geben, in der Kinder zwar schon sicher die Addition und Subtraktion im Zahlenraum bis 1000 im Kopf, aber noch nicht sicher die Multiplikation und Division im Kopf im Zahlenraum größer als 100 bzw. im Bereich des großen Ein-Mal-Eins beherrschen. Unplausibel wäre dagegen eine Kompetenzklasse mit dem Antwortmuster, bei dem Kinder korrekt Subtraktionsaufgaben lösen, aber Schwierigkeiten bei den Additionsaufgaben haben. Somit lassen sich einige der Kompetenzklassen zum Teil in eine geordnete Rangfolge bringen. So zeigen Schüler, die alle vier Grundrechenarten „beherrschen“, insgesamt bessere Testleistungen als Kinder, die nur die Strichrechnung erfolgreich bewältigen. Die Anzahl der Schüler in verschiedenen Kompetenzklassen gibt dann einen Hinweis darauf, in welcher Reihenfolge die Teilfertigkeiten in der Regel erworben werden.

Die dritte Hoffnung, die mit CDMs verbunden ist, betrifft die Modellierung von nicht-kompensatorischen Modellen, die mit etablierten statistischen Methoden nicht geschätzt werden können. In *kompensatorischen Modellen* wird angenommen, dass Mängel in einer Teilfertigkeit durch Kenntnisse einer anderen Fertigkeit ausgeglichen (bzw. kompensiert) werden können. Nicht-kompensatorische Modelle hingegen postulieren, dass alle Teilfertigkeiten erforderlich sind, um eine Aufgabe korrekt lösen zu können. Da viele didaktische und psychologische Modelle davon ausgehen, dass alle postulierten Teilfertigkeiten für die Lösung einer Aufgabe erforderlich sind, erlauben die nicht-kompensatorischen CDMs im Gegensatz zu etablierten statistischen Methoden explizit die Prüfung dieser Modelle.

2. Kognitive Diagnosemodelle: Methodische Aspekte

Der folgende Abschnitt soll ein grundlegendes Verständnis für CDMs und deren Einordnung neben anderen, etablierten statistischen Methoden vermitteln. Zusammenfassend lassen sich CDMs als konfirmatorische probabilistische Item-Response (IRT) Modelle mit kategorialen latenten Variablen, die Mehrfachladungsstrukturen erlauben, beschreiben. Im Folgenden werden die Begriffe „konfirmatorisch“ und „probabilistische IRT Modelle“ näher erläutert. Zunächst wird hierfür beschrieben, was unter einem latenten Faktor zu verstehen ist.

Wie die meisten Personeneigenschaften in der Bildungsforschung sind Kompetenzen in der Regel nicht direkt beobachtbar; man kann auf diese nur indirekt aufgrund von beobachteten Aufgabenlösungen schließen. In diesem Fall spricht man von einem *theoretischen Konstrukt*, das in einem statistischen Modell durch eine sog. *latente Variable* bzw. einen *latenten Faktor* repräsentiert wird.

Konfirmatorische Modelle prüfen, ob das theoretisch angenommene Strukturmuster der untersuchten Kompetenz die empirisch erhobenen Daten ausreichend gut beschreibt. *Probabilistische IRT-Modelle* stellen eine breite Klasse von statistischen Modellen dar, die probabilistische (d.h. wahrscheinlichkeitsbezogene) Zusammenhänge zwischen dem beobachteten Antwortverhalten und der interessierenden latenten Fähigkeit der Probanden modellieren. In spezifischen IRT-Modellen werden bei der Schätzung der latenten Fähigkeitsverteilungen Aufgabenschwierigkeit, Trennschärfe der Aufgaben oder die Ratewahrscheinlichkeit berücksichtigt. Im Unterschied zu probabilistischen IRT-Modellen werden in CDMs an Stelle von kontinuierlichen, *kategoriale* latente Variablen angenommen. Kontinuierliche Variablen beschreiben den Grad der Beherrschung einer Fertigkeit auf einer Skala mit fließenden Übergängen. Kategoriale Variablen hingegen sind meistens dichotom, sie haben also nur zwei Ausprägungen wie bspw. Beherrschung und Nichtbeherrschung einer Fertigkeit.

In der neueren methodischen Literatur, insbesondere im angelsächsischen Raum, wurden in den letzten Jahren verschiedene CDMs entwickelt (vgl. für eine Übersicht Rupp/Templin/Henson in Druck; Rupp/Templin 2008; diBello/Roussos/Stout 2007; Leighton/Gierl 2007). Zu diesen Modellen zählen u.a. die DINA und NIDA Modelle (vgl. z.B. de la Torre 2009; Junker/Sijtsma 2001), die DINO und NIDO Modelle (vgl. z.B. Rupp/Templin/Henson, in Druck) sowie die flexiblen Modellklassen des „general diagnostic model“ (vgl. z.B. von Davier 2005). Einen neuen integrativen Ansatz stellen log-lineare Modelle dar (vgl. Henson/Templin/Willse 2009). Bisher gibt es jedoch nur wenige erfolgreiche empirische Anwendungen dieser Modelle. In vielen Fällen handelt es sich um Reanalysen von Daten, die ursprünglich nur eine einzelne Dimension messen (sog. „Retrofitting“). Beim Retrofitting treten häufig sehr hohe Korrelationen zwischen den angenommenen latenten Teilfertigkeiten sowie Konvergenzprobleme auf. Diese Probleme können dafür sprechen, dass nicht alle postulierten Teilfertigkeiten statistisch und inhaltlich bedeutungsvoll voneinander separierbar sind.

3. Projektziele & Arbeitsprogramm

In dieser ersten Projektphase war das vorrangige Ziel, CDMs am Beispiel eines neu entwickelten Mathematiktests erfolgreich einzusetzen sowie diese Modelle mit anderen etablierten statistischen Auswertungsmethoden (IRT-Modellen und konfirmatorischen faktoranalytischen Modellen) zu vergleichen. Dieses Ziel war deswegen so entscheidend, weil der Nachweis des inkrementellen Nutzens der CDMs gegenüber etablierten statistischen Modellen umstritten ist und es bisher zu wenige Erfolgsfälle aus Anwendungen gibt, die kein Retrofitten darstellen (vgl. Wilhelm/Robitzsch 2009).

Zunächst stand die Entwicklung eines Klassifikationssystems mathematischer Fertigkeiten im Vordergrund. Anschließend wurden Aufgaben entsprechend dieses Klassifikationssystems entworfen. Die entwickelten Aufgaben wurden in einer ersten Pilotierungsstudie im April 2008 an 464 Grundschulkindern der dritten und vierten Klassen erprobt. Es folgte die Anpassung der Aufgaben und eine Normierung von Items mit zufriedenstellenden Itemschwierigkeiten und -trennschärfen in einer zweiten Studie im Juli 2008 an 747 Dritt- und Viertklässlern. In einer dritten Studie im November und Dezember 2008 wurden schließlich deutschlandweit Daten von 2123 Kindern der vierten Klassen an 47 Schulen erhoben, um das neu entwickelte Messinstrument abschließend zu normieren und anhand der Aufgaben zur Erfassung der Bildungsstandards Mathematik in der Grundschule sowie anhand des etablierten Mathematiktests DEMAT 3+ zu validieren. Im Folgenden soll der Prozess der Aufgabenentwicklung beschrieben werden.

4. Aufgabenentwicklung

Um den denkbaren Nutzen der CDMs optimal zu realisieren und um oben erörterte Schwierigkeiten beim Retrofitten zu vermeiden, wurden gezielt neue Aufgaben konstruiert, die auf einer vorab theoretisch definierten Q -Matrix basieren. Die Definition der Q -Matrix resultierte aus einem neu entwickelten Klassifikationssystem arithmetischer Kompetenzen. Die Entwicklung des Klassifikationssystems orientierte sich an Lehrplänen verschiedener Bundesländer, Bildungsstandards der Kultusministerkonferenz sowie an fachdidaktischer Literatur.

Um die Anforderungen der Aufgaben so eindeutig wie möglich zu halten, wurde bewusst auf die Verwendung komplexer offener Mathematikaufgaben verzichtet. Ein Problem komplexer offener Aufgaben besteht u.a. darin, dass beim Lösen solcher komplexer Aufgaben allgemeine kognitive Fähigkeiten und Sprachverständnis eine wichtige Rolle spielen, so dass unklar ist, inwieweit sich Kompetenzen wie „Kommunizieren“ oder „Argumentieren“ von solchen allgemeinen Fähigkeiten trennen lassen. Stattdessen wurden möglichst grundlegende Aufgaben zu den im Unterricht besonders intensiv behandelten Inhalten eingesetzt.

Mit diesen Prämissen und Einschränkungen versehen, wurde der Versuch unternommen, optimale Bedingungen für die erfolgreiche Schätzung theoretisch schlüssiger CDMs zu schaffen. Hierfür wurde für einfache Rechenaufgaben und Sachaufgaben ein

	$X + 22 = 27$	$113 + X = 204$	$106 - 8 = X$	$10 \times 8 = X - 15$	$250 : 5 = X \times 5$	Sachaufgabe 4
Addition	1	1	0	0	0	1
Subtraktion	0	0	1	1	0	1
Multiplikation	0	0	0	1	1	0
Division	0	0	0	0	1	0
Modellieren	0	0	0	0	0	1
Zehner- Übertrag	0	0	1	1	0	0
Zahlenraum größer 100	0	1	1	0	1	0
Umkehr- operation	1	1	0	1	1	0
Unbekannte Anfangsmenge	1	0	0	-	-	0
Unbekannte Teilmenge	0	1	0	-	-	0
Unbekannte Endmenge	0	0	1	-	-	1

Tab. 1: Q-Matrix für kontextfreie und kontextgebundene Aufgaben

atomistisches Klassifikationssystem entwickelt, das sich vornehmlich an der mathematischen Struktur der Aufgaben orientiert. Grundsätzlich ist denkbar, das vorliegende Aufgabenmaterial fachdidaktisch inspirierter zu klassifizieren und dabei die zugrunde liegenden kognitiven Denkhandlungen stärker zu betonen, anstatt sehr nah an Oberflächenaufgabenmerkmalen zu bleiben. Ein Ergebnis einer solchen Klassifikation könnten konkurrierende Messmodelle sein.

Drei zentrale Anforderungen an den neu entwickelten Mathematiktest wurden vorab als Fragestellungen formuliert: Die erste Fragestellung war, inwiefern der neu entwickelte Test ein reliables und valides Maß für die allgemeine mathematische Kompetenz liefert. Die zweite Fragestellung war, ob es der neu entwickelte Test auf einer elementareren Ebene erlaubt, die Beherrschung der vier Grundrechenarten und des Modellierens getrennt voneinander zu schätzen, auch wenn diese Teilkompetenzen voraussichtlich stark korreliert sind.

	+	22	=	27
	+	45	=	80
25	+		=	73
	+	6	=	104
113	+		=	204

$$10 \cdot 8 = \underline{\quad} - 15$$

$$\underline{\quad} + 6 = 72 : 9$$

Abb. 1: Beispiel für kontextfreie einfache und komplexe Rechenaufgaben

Die dritte Fragestellung war, inwiefern mittels des Tests auf einer noch elementareren Ebene zwischen Fertigkeiten zur Bewältigung spezifischer Aufgabenanforderungen unterschieden werden kann. Die Auswahl der relevanten Aufgabenanforderungen beruht zum Teil auf den Arbeiten von Carpenter u.a. (1999). Die Aufgaben wurden hierbei u.a. danach unterschieden, ob Anfangs-, Teil- oder Endmengen unbekannt sind. Bei der Aufgabenkonstruktion wurde weiterhin die Anwesenheit oder Abwesenheit eines Übertrags („Zehner-Übertrag“), verschiedene Zahlenbereiche („Rechnen im Zahlenraum größer 100“) und das Ausführen einer Umkehroperation (genauer einer Umkehroperation zu der Operation, die in der Aufgabe visuell gefordert wird) berücksichtigt. Die daraus resultierende Q-Matrix ist in Tabelle 1 dargestellt.

Aus diesem Prozess der Aufgabenentwicklung resultierten zwei Typen von Aufgaben. Der erste Typ beschreibt einfache und komplexe, kontextfreie Rechenaufgaben (s. Abb. 1). Ein zweiter Typ von Aufgaben sind kontextgebundene Sachaufgaben (s. Abb. 2). In der dritten Studie im November und Dezember 2008 wurden darüber hinaus Aufgaben aus dem Inhaltsbereich „Messen und Größen“ eingesetzt, die Umrechnungen und Anwendung arithmetischer Operationen auf verschiedene Maß-, Geld- und Längeneinheiten erforderten.

Sachaufgabe 4 T_A55

Marlen hat 55 Lieder auf ihrem MP3-Player. Leon hat auf seinem MP3-Player 15 Lieder *weniger als* Marlen. Beim Durchgehen der Liederliste stellen die Kinder fest, dass sie 10 gleiche Lieder haben.

*Wie viele **unterschiedliche** Lieder haben Marlen und Leon zusammen?*

Marlen und Leon haben zusammen verschiedene Lieder.

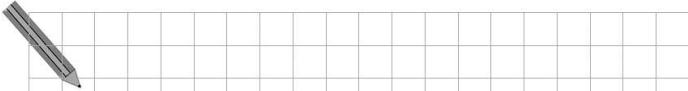




Abb. 2: Beispiel für eine kontextgebundene Sachaufgabe

5. Ergebnisse

Bei der Schätzung der CDMs ist die Anzahl und Beschaffenheit der theoretisch angenommenen relevanten Teilfertigkeiten von entscheidender Bedeutung. Um diese zu ermitteln, wurden im ersten Schritt mit Hilfe von konfirmatorischen Faktorenanalysen (CFA) (vgl. Kaplan 2000) drei denkbare Strukturmuster überprüft. Die geschätzten CFA-Modelle erlauben mittels verschiedener etablierter Modell-Fit-Indizes Aussagen über die Modellpassung und ermöglichen einen statistischen Vergleich der Modelle mittels des sog. Likelihood-Ratio-Tests.

Im ersten Modell wurde von einer allgemeinen arithmetischen Fertigkeit ausgegangen. Im zweiten Modell wurden drei verschiedene Fertigkeiten angenommen – Punktrechnung, Strichrechnung, Modellieren – wobei die Korrelationen zwischen allen drei Faktoren untereinander zugelassen wurden. Im dritten Modell wurde postuliert, dass die vier Grundrechenarten und das Modellieren fünf verschiedene, jedoch voneinander abhängige Konstrukte bzw. latente Faktoren darstellen.

Die Modellpassung für das erste Modell mit einer allgemeinen Arithmetikfähigkeit ist im Vergleich zu Modell 2 und Modell 3 deutlich schlechter. Aufgrund der hohen Korrelationen zwischen den latenten Faktoren ist es statistisch schwierig im Modell 3 zwischen Addition und Subtraktion bzw. Multiplikation und Division zu trennen. Die Datenlage in der Pilotierungsstudie spricht dafür, drei voneinander abhängige Konstrukte bzw. latente Faktoren zu unterscheiden: Punktrechnung, Strichrechnung und Modellieren (Modell 2). Darüber hinaus zeigen die Ergebnisse, dass Schüler in dritten und vierten Klassen große Schwierigkeiten mit Multiplikation und Division, insbesondere mit dem großen Ein-Mal-Eins haben.

Die zweite Frage, die in der Pilotierungsstudie beantwortet werden sollte, betraf den Vergleich der CDMs mit den korrespondierenden CFA-Modellen. Als ein Vertreter der CDMs wurde das *general diagnostic model* (GDM) gewählt, wobei hierbei das korrespondierende GDM zum CFA-Modell mit der besten Modellpassung geschätzt wurde.

Es zeigte sich, dass das GDM per se keine zusätzliche Information über die mehrdimensionalen CFA-Modelle hinaus bereitstellt. GDM erlaubt jedoch eine andere Aufbereitung der Information und erlaubt die direkte Schätzung der Kompetenzprofile der Schüler, die eine Auskunft darüber ermöglichen, welche Teilfertigkeiten vom Kind bereits beherrscht werden und welche noch nicht (vgl. Tab. 2). Diese Information kann potentiell für Rückmeldungen an Lehrkräfte und Eltern genutzt werden. Des Weiteren können die Häufigkeiten der beobachteten Kompetenzklassen (bspw. „Schüler beherrschen alle vier Rechenarten“ vs. „Schüler können addieren und subtrahieren, aber nicht ausreichend gut multiplizieren und dividieren“) Hinweise auf den Lernprozess des Erwerbs der arithmetischen Fertigkeiten liefern. Unsere Ergebnisse lassen vermuten, dass in der Regel zunächst Addition und Subtraktionskenntnisse erworben werden, die eine wichtige Voraussetzung für den Erwerb weiterer Teilfertigkeiten bilden. Aufbauend darauf erfolgt parallel der Erwerb der Modellierungskompetenz bzw. von Multiplikation und Division.

Eine ausführliche Beschreibung der hier berichteten Ergebnisse der Pilotierungsstudie sowie umfassende methodische Erläuterungen findet der interessierte Leser bei Kunina-Habenicht/Rupp/Wilhelm (2009).

In der vor kurzem abgeschlossenen Validierungsstudie wurde der neu entwickelte Mathematiktest zusammen mit den Bildungsstandardaufgaben in Mathematik für die Grundschule, DEMAT 3+ (vgl. Roick/Görlitz/Hasselhorn 2004) sowie dem kognitiven Fähigkeitstest (KFT; vgl. Heller/Perleth 1976) administriert. Erste Ergebnisse deuten auf mittelhohe Zusammenhänge zwischen dem neu entwickelten Mathematiktest und dem DEMAT 3+, den Bildungsstandardaufgaben und der Mathematiknote hin. Ausführliche Informationen hierzu sind zu finden bei Kunina/Wilhelm/Rupp (2009).

In den nächsten Analyseschritten streben wir die Anwendung des neuen integrativen Ansatzes der log-linearen Modelle auf die Daten aus der Validierungsstudie an. Anschließend wird die Validität der so berechneten Kompetenzprofile anhand der Zusammenhänge mit den eingesetzten externen Messinstrumenten und mit der Mathematiknote geprüft. Ein weiteres Ziel in diesem Projekt betrifft die Untersuchung geeigneter Modell-Fit-Maße, die Auskunft über die Modellpassung und damit die Prüfung der Adäquatheit der CDMs geben.

Teilfertigkeiten					
Addition	Subtraktion	Multiplikation	Division	Modellieren	3. Klasse
0	0	0	0	0	23,5 %
1	1	1	1	1	21,0 %
1	1	0	0	1	14,2 %
1	1	1	1	0	6,6 %
1	1	0	0	0	6,3 %
1	0	0	0	0	6,0 %
1	1	1	0	1	4,6 %
69,2 %	54,5 %	44,8 %	37,2 %	48,5 %	82,2 %

Anmerkungen: 0 – Teilfertigkeit wird beherrscht; 1 – Teilfertigkeit wird nicht beherrscht. In der letzten Zeile ist der Anteil der Schüler angegeben, die die entsprechende Teilfertigkeit erfolgreich erworben haben. Genauere Informationen und Hinweise zu den statistischen Berechnungen sind zu finden bei Kunina-Habenicht/Rupp/Wilhelm (2009).

Tab. 2: Prozentualer Anteil von Schülern der 3. Klassen in ausgewählten Kompetenzklassen

6. Fazit: Zukunft der kognitiven Diagnosemodelle

Im abschließenden Abschnitt wollen wir eine subjektive Einschätzung der Vorzüge und Probleme geben, die aus unserer Sicht derzeit mit den CDMs bestehen.

Eine Hoffnung, die mit den CDMs verbunden ist, betrifft die statistische und damit scheinbar objektive Klassifikation der Schüler in Kompetenzklassen. Diese Klassifikation ist jedoch nur dann bedeutungsvoll, wenn die Q -Matrix korrekt und reliabel definiert wurde. Da die Definition der Q -Matrix jedoch idealerweise über ein Expertentreffen erfolgt, bei denen die Experten Konsens über die Definition der Q -Matrix sowie Klassifikation der Aufgaben in die Q -Matrix finden sollen, ist die Subjektivität der Klassifikation der Schüler bei den CDMs im Prozess der Definition der Q -Matrix „versteckt“ (vgl. Gorin 2009).

Die Definition der Q -Matrix – oder ggf. konkurrierender Q -Matrizen – sollte nach Möglichkeit auf der Grundlage einer psychologisch oder fachdidaktisch fundierten Theorie basieren. In der Literatur wird häufig bemängelt, dass in vielen Forschungsbereichen die theoretische Fundierung der zugrunde liegenden Teilfertigkeiten und kognitiven Prozesse gar nicht oder kaum ausgereift sind (vgl. Maris/Bechger 2009; Wilhelm/Robitzsch 2009). Ob die Orientierung bei der Erzeugung einer Q -Matrix wie in dieser Arbeit vorrangig über Aufgabenattribute oder über vermeintlich ablaufende Denkprozesse erfolgt, ist unseres Erachtens zunächst zweitrangig. Entscheidend ist – auch im direkten Modellvergleich – Modelle mit guter Passung und domänengerechter nomothetischer Spanne zu etablieren. Ein weiterer kritischer Punkt betrifft die Validität der Kompetenzprofile, die CDMs liefern, die bisher kaum untersucht worden ist (vgl. Sinharay/Haberman 2009).

Aus statistischer Perspektive bestehen bei der Schätzung der CDMs noch viele ungeklärte Fragen. Ein wesentliches Problem betrifft die Tatsache, dass die Komplexität der Modelle und damit die Anzahl der zu schätzenden Parameter mit der Anzahl der angenommenen latenten Faktoren deutlich ansteigen. Dies führt häufig zu Schätzproblemen, die sich bspw. in mangelnder Reliabilität, großen Standardfehlern der Parameterschätzungen oder ausbleibender Konvergenz äußern können (vgl. Maris/Bechger 2009). Ferner ist bisher nur wenig über die Beurteilung der Güte der Modellpassung bei den CDMs – im Vergleich zu den CFAs – bekannt (vgl. Levy 2009).

In diesem Projekt untersuchen wir einige der genannten problematischen Aspekte und hoffen eine überzeugende Anwendung für empirische Daten zu präsentieren, die klare Vorteile der CDMs gegenüber bisherigen etablierten statistischen Modellen aufzeigt.

Literatur

- Blum, W./Drüke-Noe, C./Hartung, R./Köller, O. (Hrsg.) (2006): Bildungsstandards Mathematik: konkret. Berlin: Cornelsen.
- Carpenter, T.P./Fennema, E./Franke, M.L./Empson, S.B./Levi, L.W. (1999): Children's mathematics: Cognitively guided instruction. Portsmouth, NH: Heinemann.
- de la Torre, J. (2009): DINA model and parameter estimation: A didactic. In: Journal of Educational and Behavioral Statistics 34, S. 115–130.
- di Bello, L.V./Roussos, L.A./Stout, W. (2007): Review of cognitively diagnostic assessment and a summary of psychometric models. In: Rao, C.R./Sinharay, S. (Hrsg.): Handbook of Statistics, Vol. 26. Psychometrics. Amsterdam: Elsevier, S. 979–1030.
- Gorin, J.S. (2009): Diagnostic classification model: Are they necessary? Commentary on Rupp and Templin (2008). In: Measurement: Interdisciplinary Research and Perspectives 7, S. 30–33.
- Heller, K.A./Perleth, Ch. (1976): Kognitiver Fähigkeitstest für 4.–12. Klassen (KFT 4–12+). Göttingen: Hogrefe.
- Henson, R./Templin, J./Willse, J. (2009): Defining a Family of Cognitive Diagnosis Models Using Log-Linear Models with Latent Variables. In: Psychometrika 74, H. 2, S. 191–210.
- Junker, B.W./Sijtsma, K. (2001): Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. In: Applied Psychological Measurement 25, S. 258–272.

- Kaplan, D. (2000): Structural equation modelling: Foundations and extensions. Advanced Quantitative Techniques in the Social Sciences series. London: Sage.
- Kunina-Habenicht, O./Rupp, A.A./Wilhelm, O. (2009): A Practical Illustration of Multidimensional Diagnostic Skills Profiling: Comparing Results from Confirmatory Factor Analysis and Diagnostic Classification Models. In: Studies in Educational Evaluation 35 (2), S. 64–70.
- Kunina, O./Wilhelm, O./Rupp A.A. (2009): Validity of Proficiency Profiles in Arithmetic Ability based on Cognitive Diagnosis Models. Vortrag gehalten beim annual meeting of American Educational Research Association (AERA) in San Diego USA, 13.–17. April.
- Leighton, J./Gierl, M. (Hrsg.) (2007): Cognitive diagnostic assessment for education: Theory and practice. Cambridge: Cambridge University Press.
- Levy, R. (2009): Evidentiary reasoning in diagnostic classification models. In: Measurement: Interdisciplinary Research and Perspectives 7, S. 36–41.
- Maris, G./Bechger, T. (2009): Equivalent diagnostic classification models. In: Measurement: Interdisciplinary Research and Perspectives 7, S. 41–46.
- Roick, T./Görlitz, D./Hasselhorn, M. (2004): Deutscher Mathematiktest für dritte Klassen (DE-MAT 3+). Göttingen: Beltz Test.
- Rupp, A.A./Templin, J. (2008): Unique characteristics of diagnostic classification models: A comprehensive review of the current state-of-the-art. In: Measurement: Interdisciplinary Research and Perspectives 6, S. 219–262.
- Rupp, A.A./Templin, J./Henson, R. (2010): Diagnostic measurement: Theory, methods, and applications. New York: The Guilford Press.
- Sinharay, S./Haberman, S.J. (2009): How much can we reliably know about what examinees know? In: Measurement: Interdisciplinary Research and Perspectives 7, S. 46–49.
- von Davier, M. (2005): A general diagnostic model applied to language testing data (Research Report No. RR-05-16). Princeton, NJ: Educational Testing Service.
- Wilhelm, O./Robitzsch, A. (2009): Have cognitive diagnostic models delivered their good? Some substantial and methodological concerns. In: Measurement: Interdisciplinary Research and Perspectives 7, S. 53–57.

Anschrift der Autor/innen

Olga Kunina-Habenicht, Goethe-Universität Frankfurt, Institut für Psychologie, Arbeitsbereich Pädagogische Psychologie, Senckenberganlage 15, D-60325 Frankfurt a.M.
E-Mail: kunina@paed.psych.uni-frankfurt.de

Oliver Wilhelm, Universität Duisburg-Essen, Berliner Platz 6–8, CO 8/17, D-45127 Essen
E-Mail: oliver.wilhelm@uni-due.de

Franziska Matthes, Humboldt Universität zu Berlin, Institut zur Qualitätsentwicklung im Bildungswesen (IQB), Unter den Linden 6, D-10099 Berlin
E-Mail: franziska.matthes@student.hu-berlin.de

André A. Rupp, Department of Measurement, Statistics, and Evaluation (EDMS), University of Maryland, USA
E-Mail: ruppandr@umd.edu