

Rupp, André A.; Leucht, Michael; Hartung, Ralph
"Die Kompetenzbrille aufsetzen". Verfahren zur multiplen Klassifikation von Lernenden für Kompetenzdiagnostik in Unterricht und Testung
Unterrichtswissenschaft 34 (2006) 3, S. 195-219



Quellenangabe/ Reference:
Rupp, André A.; Leucht, Michael; Hartung, Ralph: "Die Kompetenzbrille aufsetzen". Verfahren zur multiplen Klassifikation von Lernenden für Kompetenzdiagnostik in Unterricht und Testung - In: Unterrichtswissenschaft 34 (2006) 3, S. 195-219 - URN: urn:nbn:de:0111-opus-55156 - DOI: 10.25656/01:5515

<https://nbn-resolving.org/urn:nbn:de:0111-opus-55156>
<https://doi.org/10.25656/01:5515>

in Kooperation mit / in cooperation with:

BELTZ JUVENTA

<http://www.juventa.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.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.
This document is solely intended for your personal, non-commercial use. Use of this document does not include any transfer of property rights and it is conditional to the following limitations: All of the copies of this documents must retain all copyright information and other information regarding legal protection. You are not allowed to alter this document in any way, to copy it for public or commercial purposes, to exhibit the document in public, to perform, distribute or otherwise use the document in public.

By using this particular document, you accept the above-stated conditions of use.

Kontakt / Contact:

peDOCS
DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation
Informationszentrum (IZ) Bildung
E-Mail: pedocs@dipf.de
Internet: www.pedocs.de

Digitalisiert

Unterrichtswissenschaft

Zeitschrift für Lernforschung

34. Jahrgang / 2006 / Heft 3

Thema:

Kompetenzmodelle

Verantwortliche Herausgeber:

Olaf Köller, Manfred Prenzel

Erratum 194

André A. Rupp, Michael Leucht, Ralph Hartung
„Die Kompetenzbrille aufsetzen“ – Verfahren zur
multiplen Klassifikation von Lernenden für
Kompetenzdiagnostik in Unterricht und Testung..... 195

Timo Ehmke, Dominik Leiß, Werner Blum, Manfred Prenzel
Entwicklung von Testverfahren für die Bildungsstandards
Mathematik. Rahmenkonzeption, Aufgabenentwicklung,
Feld- und Haupttest 220

Olaf Köller, Ulrich Trautwein, Kai S. Cortina, Jürgen Baumert:
Rezeptive Kompetenzen in Englisch am Ende der gymnasialen
Oberstufe. Verankerung deutscher Abiturienten am
Gemeinsamen Europäischen Referenzrahmen für die Sprachen..... 239

Allgemeiner Teil

Jörg Zumbach, Johann Pixner
Textverständlichkeit und kognitive Belastung beim Lernen
mit Text und Hypertext 256

Andreas Hartinger
Interesse durch Öffnung des Unterrichts – wodurch? 272

„Die Kompetenzbrille aufsetzen“

Verfahren zur multiplen Klassifikation von Lernenden für
Kompetenzdiagnostik in Unterricht und Testung

Competence Diagnostic in Mathematics – Methodes for Multiple
Classification of Learners and Competence Diagnostic in Schools

Der vorliegende Beitrag bietet eine Einführung in die Grundlagen testtheoretischer Ansätze zur Kompetenzdiagnostik auf der Basis statistischer Modelle zur multiplen Klassifikation von Lernenden. Solche Modelle erlauben es, durch die Spezifizierung der zur Lösung einzelner Aufgaben relevanten Kompetenzen statistisch abgesicherte Kompetenzprofile für Lernende zu erstellen und somit detaillierte diagnostische Rückmeldungen zur Bestimmung von individuellem Förderbedarf zu entwickeln. In einem ersten Abschnitt werden zunächst Kernmerkmale der mit diesen Modellen zu erfassenden Kompetenzen im Sinne der länderübergreifenden Bildungsstandards diskutiert. In einem zweiten Abschnitt wird verdeutlicht, was man aus der Systematik der Aufgabenentwicklung für Testzwecke für die Aufgabenentwicklung im Unterricht lernen kann. In einem dritten Abschnitt werden theoretische und praktische Vorzüge kognitiv-diagnostischer Modelle anhand des so genannten DINA Modells beispielhaft beschrieben. Ferner werden Ausblicke auf alternative Modelle gegeben sowie praktische Herausforderungen bei deren Verwendung diskutiert.

This article is an introduction to the conceptual and statistical foundations of multiple classification models. Through careful assessment design, these models can be used to develop complex performance profiles of learners that provide detailed diagnostic feedback that is useful for the creation of remedial programs tailored to the needs of individual learners. In the first part of the article, core characteristics of the theoretical construct of ‘competences’ are reviewed, which are operationalized by these models. This is done specifically in light of the central role the notion of competence has played in the development of the national educational standards in Germany. In the second section, a classroom task on mathematical problem-solving is used to illustrate what can be learned from the input re-

quirements of such models in large-scale assessment studies for the process of item development in general. In the third section, theoretical and practical advantages of such models are further illustrated with the DINA model. Finally, alternative models for cognitive diagnosis as well as challenges for their practical implementation are reviewed.

1. Einführung

Wie stelle ich fest, welche spezifischen Kompetenzen meine Schülerinnen und Schüler in meinem Fach erworben haben? Wie gebe ich ihnen detaillierte Rückmeldungen über die festgestellten Leistungsstände? Wie helfe ich ihnen, diejenigen Kompetenzen, über die sie bereits verfügen, zu festigen, und diejenigen Kompetenzen, über die sie noch nicht verfügen, zu erwerben?

Solche Fragen über Fähigkeiten von Schülerinnen und Schülern dienen der individuellen Förderung für einen nachhaltigen Wissens- und Kompetenzerwerb und gehören zum alltäglichen Geschäft von Lehrkräften. In Deutschland hat die Feststellung von komplexen Leistungsprofilen von Lernenden in den letzten Jahren eine besondere Priorität erfahren, was sich insbesondere auf das relativ schwache Abschneiden deutscher Schülerinnen und Schüler in internationalen Schulleistungsvergleichen wie TIMSS (z.B. Baumert, Bos, & Lehmann, 2000) oder PISA (z.B. Baumert et al., 2001; Prenzel et al., 2003) sowie den darin enthüllten sozialen und ethnischen Disparitäten begründet.

Mit der Verabschiedung *länderübergreifender Bildungsstandards* durch die *Kultusministerkonferenz der Länder der Bundesrepublik Deutschland (KMK)* im Herbst 2003 und 2004 liegen erstmals für alle 16 Länder verbindliche Leistungsanforderungen für Schülerinnen und Schüler in Form von zu erlernenden Kompetenzen vor. Die Bildungsstandards legen dabei a priori ein mittleres Anforderungsniveau im Hinblick auf diese fachbezogene Kompetenzen fest, welches Schüler am Ende der Primarstufe oder bei Erlangen des Hauptschul- bzw. des Mittleren Abschlusses erreicht haben sollen. In diesem Sinne handelt es sich bei den Bildungsstandards der KMK um *performanzbezogene Regelstandards* (KMK, 2004b; vgl. Klieme et al., 2003), wobei zu bedenken ist, dass erst empirische Daten über die Leistungen deutscher Schülerinnen und Schüler verlässliche Auskunft geben werden, wie Regelstandards zu definieren sind.

Um die Bildungsstandards der KMK in eine angemessene Aufgaben- und Lernkultur zu überführen, damit das Erreichen der Standards überprüft und individueller Förderbedarf identifiziert werden kann, muss entschieden werden, mit Hilfe welcher Methoden Aussagen über derzeitige Kompetenzprofile sowie über Kompetenzerwerb und Kompetenzentwicklung empirisch abgesichert werden können. Ein wichtiges Evaluationsinstrument für größer

angelegte Untersuchungen stellen dabei standardisierte Tests in nationalen und internationalen large-scale assessment-Studien wie PISA, TIMSS, TOSCA (z.B. Köller, Watermann, Trautwein, & Lüdtke, 2003) oder DESI (z.B. Klieme & Beck, im Druck) dar. Ergebnisse dieser Studien stellen umfassende Informationen für normbezogene Rückmeldungen auf Klassen- oder Schulebene bereit, vernachlässigen jedoch individuelle Rückmeldungen.

Methodisch gesehen begründet sich dies zum einen aus der komplexen Designstruktur dieser Untersuchungen, bei denen große Aufgabenpools anhand von vielen rotierten Testheften unterschiedlichen Schülergruppen vorgelegt werden. Dabei sind die Designs so ausgerichtet, dass multiple Schätzungen für Schülerleistungen resultieren, welche auf Klassen- oder Schulebene unverzerrt und ausreichend präzise sind. Zum anderen ist dies durch die verwendeten Skalierungsmodelle begründbar. Hier werden überwiegend statistische Modelle aus der probabilistischen Testtheorie (z.B. Embretson & Reise, 2000; Rost, 2004a) herangezogen, mit deren Hilfe Schülerinnen und Schüler auf eindimensionalen Fähigkeitskontinua verortet werden können (für Beispiele siehe z.B. Bos et al., 2003; Köller, Watermann, Trautwein, & Lüdtke, 2003; siehe auch die Artikel in Kohler & Schrader, 2004).

In diesem Beitrag sollen statistische Modelle vorgestellt werden, die diagnostische Schlussfolgerungen für Einzelschüler – und damit gezielte pädagogische Interventionen – für solche Untersuchungen erlauben. Ferner soll aufgezeigt werden, was aus der Systematik der Aufgabenentwicklung zu Testzwecken im Hinblick auf eine spätere Kalibrierung von Daten mit diesen Modellen für Aufgabenentwicklung im Unterricht gelernt werden kann. Um die theoretischen Vorzüge und praktischen Grenzen dieser Modelle detaillierter darstellen zu können, werden im folgenden Abschnitt diejenigen Merkmale von Kompetenzen diskutiert, die für deren kognitiv-diagnostische Modellierung besonders relevant sind. Im zweiten Abschnitt wird anhand einer Aufgabe zur Illustration der Bildungsstandards im Fach Mathematik gezeigt, was unter einer kompetenzorientierten Konzeptualisierung von Aufgaben für Unterricht und Testung zu verstehen ist. Im dritten Abschnitt werden die Möglichkeiten einer statistisch abgesicherten Kompetenzdiagnostik anhand des so genannten DINA Modells (z.B. Junker & Sijtsma, 2001; de la Torre & Douglas, 2004) veranschaulicht. Ferner wird eine kurze Übersicht über alternative Modelle gegeben. Der Artikel schließt mit Betrachtungen zum Potenzial kognitiv-diagnostischer Modelle im Hinblick auf ihre Verwendung bei der Rückmeldung von Schülerkompetenzen bezüglich der länderübergreifenden Bildungsstandards der KMK.

2. Zum Kompetenzbegriff

In der empirischen Bildungsforschung der letzten Jahre wird zur Beschreibung der Schülerleistungen vielfach der Begriff der *Kompetenz* verwendet. In

der Expertise zur Entwicklung nationaler Bildungsstandards (Klieme u.a., 2003), welche ein wichtiger Input bei der Entwicklung der länderübergreifenden Bildungsstandards war (siehe KMK, 2004b) werden mit Weinert (2001) Kompetenzen definiert als 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, um die Problemlösungen in variablen Situationen erfolgreich und verantwortungsvoll nutzen zu können.“ (S. 27f). Kompetenz wird so mit dem Potenzial gleichgesetzt, in bestimmten Situationen auf bestimmten Gebieten Probleme zu lösen (vgl. Prenzel, Drechsel, Carstensen & Ramm, 2004, S. 18; Kanning, 2003, S. 12). Offensichtlich ist dieser Kompetenzbegriff aus psychologischer Sicht sehr weit gefasst, da er nicht nur kognitionspsychologische (d.h. Fähigkeiten und Fertigkeiten), sondern auch lernpsychologische Elemente (d.h. motivationale, volitionale, und soziale Bereitschaften) beinhaltet. Er erlaubt somit die Subsummierung einer ganzen Reihe von Testbatterien und Fragebögen zur Leistungs-, Persönlichkeits- und Verhaltensdiagnostik (vgl. Kubinger, 2006; Rost, 2004a), um die von ihnen gemessenen Konstrukte in Beziehung zueinander zu stellen.

Solch eine umfassende definitorische Klammer schafft Freiheiten in der Gestaltung von normativen Vorgaben wie den Bildungsstandards, jedoch entstehen durch sie große Herausforderungen für eine Operationalisierung der Einzelkonstrukte und deren nomologischer Vernetzung (Embretson, 1983). Hierbei ist besonders die notwendige Präzisierung der Konstrukte hervorzuheben, wenn Messinstrumente zuverlässige Daten generieren sowie valide Aussagen über Schülerinnen und Schüler zulassen sollen. Diesbezüglich ergeben sich vier Merkmale, anhand derer Kompetenzen im hier verstandenen Sinn etwas näher beschrieben werden sollen.

Erstens ist auf semantischer und funktionaler Ebene die Abgrenzung dessen, was als Kompetenz beschrieben wird, primär eine Frage der Detailtiefe (Rupp, im Druck; Rupp & Mislevy, im Druck). In der Regel stellen Kompetenzen Aufschlüsselungen gröber definierter *Fähigkeiten* dar, welche allerdings ebenso als *Kompetenzbereiche* bezeichnet werden könnten. So könnte man zum Beispiel sagen, dass die *mathematische Kompetenz* „mit Symbolen und Formalismen umzugehen“ ohne Frage einen Teilaspekt des *mathematischen Kompetenzbereichs* oder der *mathematischen Fähigkeiten* von Schülerinnen und Schülern darstellt (vgl. Prenzel et al. 2004, S. 50; siehe auch KMK, 2005). Derartige Aufschlüsselungen nehmen üblicherweise Bezug auf fachdidaktische Vorstellungen zum Aufbau dieser Fähigkeiten (für Mathematik siehe u.a. Freudenthal, 1977, 1983; Niss, 2003). Somit ist die Interpretation von Variablen in statistischen Modellen zur Kompetenzdiagnostik, welche diese Kompetenzen abbilden sollen, immer nur durch den jeweiligen Anwendungskontext präzisierbar.

Zweitens werden die Beschreibungen von Kompetenzen und deren Beziehung zur Komplexität von Löseprozessen oft in so genannten *Kompetenzmodellen* zusammengefasst. In den Bildungsstandards zum Fach Mathematik in der Sekundarstufe I wird zum Beispiel ein konzeptioneller Raum aufgespannt, der sich aus einem Cluster von sechs *allgemeinen mathematischen Kompetenzbereichen* und einem Cluster von fünf *inhaltlichen mathematischen Kompetenzbereichen* zusammensetzt und mit Hilfe von drei *Komplexitätsstufen*, so genannten Anforderungsbereichen, ergänzt wird (KMK, 2003, 2004a). Ähnlich unterteilt das PISA-Rahmenkonzept (OECD, 2003) in Anlehnung an Niss (2003) insgesamt acht mathematische Kompetenzbereiche, die in drei Clustern mit unterschiedlichen Komplexitätsstufen zusammengefasst werden. Ein Desideratum kann nun sein, solche theoretisch verankerten Kompetenzmodelle empirisch abzubilden (z.B. Klieme, Baumert, Köller, & Bos, 2000, für TIMSS III). Dies kann explizit durch die Festlegung einer bestimmten Dimensionalität für ein statistisches Modell erreicht werden, sowie durch die a priori Zuordnung von multiplen Kompetenzen zu einzelnen Aufgaben, was unter anderem als Input für kognitiv-diagnostische Modelle dienen kann.

Drittens schließt die Konzeptualisierung von separaten Kompetenzen selbstverständlich nicht aus, dass zur erfolgreichen Bearbeitung von Aufgaben die *Kombination mehrerer Kompetenzen* erforderlich sein kann, was aus didaktischer Sicht insbesondere bei komplexen authentischen Aufgaben ausdrücklich gewünscht ist (vgl. etwa Blum et al., 2004, S.50). Es ist dabei wichtig zu unterscheiden, ob die Kompetenzen *nicht-kompensatorisch* oder *kompensatorisch* zueinander in Beziehung stehen. Mit anderen Worten, es ist wichtig theoretisch festzulegen, ob es eine sinnvolle Annahme ist, dass man ein Defizit in einer Kompetenz durch einen Überschuss in einer anderen Kompetenz beim Lösen einer Aufgabe ausgleichen kann (für Beispiele und Diskussionen siehe z.B. Bolt & Lall, 2003; Maris, 1995; Rupp, 2002). Dies spiegelt sich in der verbalen Beschreibung des Kompetenzmodells sowie in der Beschreibung der Aufgaben anhand der Kompetenzen wider und kognitiv-diagnostische Modelle greifen die Verwendung der Kompetenzen explizit durch ihre Strukturen auf¹.

Viertens ist es wichtig festzuhalten, dass Kompetenzen *nicht direkt beobachtbar* sind und damit ganz bewusst vom direkt erfassbaren Lösungsverhalten der Schülerinnen und Schüler unterschieden werden müssen. So wird unter anderem „aus den Antworten auf die Testfragen zu einem bestimmten Teilgebiet [...] die mathematische oder naturwissen-

¹ Dies geschieht durch die Verwendung eines *Produkts* über latente Variablen in der Formel für nicht-kompensatorische Modelle und einer *Summe* über latente Variablen in der Formel für kompensatorische Modelle.

schaftliche Kompetenz oder Lese- oder Problemlösekompetenz erschlossen“ (Prenzel, Drechsel, Carstensen & Ramm, 2004, S. 19). Aus statistischer Sicht spiegelt sich dies in der Verwendung von *latenten Variablen* (z.B. Skrandal & Rabe-Hesketh, 2004) in Modellen für Kompetenzdiagnostik wider, welche die einzelnen Kompetenzen abbilden.

Zusammenfassend kann man also sagen, dass die theoretischen Desiderata der Bildungsstandards bezüglich Aussagen über Kompetenzstrukturen eine natürliche Passung mit den in diesem Artikel vorgestellten kognitiv-diagnostischen Modellen haben, da sie (1) latente Kompetenzen durch latente Variablen abbilden, diese (2) kompensatorisch und nicht-kompensatorisch auf verschiedene Wege verknüpfen können, somit (3) komplexe Kompetenzstrukturen abbilden können und (4) dadurch Aussagen über einzelne Kompetenzen oder ganze Kompetenzbereiche, je nach gewünschter Detailtiefe der Rückmeldungen, ermöglichen. Im folgenden Abschnitt wird nun die Entwicklung von kompetenzbasierten Aufgaben im Hinblick auf diese Merkmale zum Zwecke einer Modellierung von Daten mit kognitiv-diagnostischen Modellen beschrieben.

3. *Komplexe Kompetenzstrukturen in Unterrichts- und Testaufgaben*

In diesem Abschnitt soll anhand einer Aufgabe zur Illustration der von der KMK verabschiedeten Bildungsstandards für das Fach Mathematik (KMK, 2003, 2004a) verdeutlicht werden, was eine kompetenzorientierte Entwicklung von Aufgaben für den Unterricht und für die Testung bedeutet. Insbesondere soll gezeigt werden, wie leichte Veränderungen ein und derselben Aufgabe sich in unterschiedlichen Anforderungsprofilen im Hinblick auf die von ihr aktivierten Kompetenzen ausdrücken. Hierbei soll die Repräsentation der unterschiedlichen Kompetenzstrukturen der Aufgabenvariationen in Tabellen- oder Matrixform verdeutlicht werden, da diese den Kerninput für die kognitiv-diagnostischen Modelle darstellen. Der Abschnitt schließt mit Schlussfolgerungen im Hinblick auf eine kompetenzorientierte und damit auf die Bildungsstandards der KMK bezogene Unterrichtsentwicklung.

Grundlage für diesen Abschnitt ist eine Aufgabe, die repräsentativ für Aufgaben aus einem Buch zur Implementierung der Bildungsstandards im Fach Mathematik (Blum, Drüke-Noe, Hartung, & Köller, im Druck) ist. Die Aufgabe bezieht sich auf die Satzgruppe des Pythagoras und ist in sechs unterschiedlichen Varianten in Tabelle A1 wiedergegeben. Zur Beschreibung dieser Aufgabe wird das für die Bildungsstandards der KMK entwickelte Kompetenzmodell herangezogen (KMK, 2003, 2004a). Dieses Modell unterscheidet zuerst sechs *allgemeine mathematische Kompetenzbereiche*, welche durch noch spezifischere Teilkompetenzen aufgeschlüsselt werden. Die allgemeinen mathematischen Kompetenzen oder Kompetenzbereiche

werden, wie im vorigen Abschnitt beschrieben, als Dispositionen von Schülerinnen und Schülern verstanden. Ferner werden sie innerhalb verschiedener inhaltlicher Kompetenzbereiche, so genannter *Leitideen*, aktiviert. Diese werden ebenfalls durch noch spezifischere *inhaltsbezogene Kompetenzen* präzisiert. Eine Leitidee vereinigt dabei Inhalte verschiedener mathematischer Sachgebiete und durchzieht das mathematische Curriculum spiralförmig. Die allgemeinen und inhaltsbezogenen Kompetenzbereiche sind in Tabelle A1 noch einmal aufgeführt.

Tab. A1: Überblick über die in den Bildungsstandards der KMK aufgeführten mathematischen Kompetenzen für die Sekundarstufe I

Allgemeine mathematische Kompetenzen	Inhaltsbezogene mathematische Kompetenzen / Leitideen
1. Mathematisches Problemlösen 2. Mathematisches Modellieren 3. Mathematische Darstellungen verwenden 4. Mathematisch kommunizieren 5. Mathematisch argumentieren 6. Mit symbolischen, formalen und technischen Elementen der Mathematik umgehen	1. Zahl 2. Messen 3. Raum und Form 4. Funktionaler Zusammenhang 5. Daten und Zufall

Bei der Klassifizierung einer konkreten Beispielaufgabe in den Bildungsstandards wird nun stets angegeben, welche allgemeinen mathematischen Kompetenzen oder Kompetenzbereiche zur Lösung der Aufgabe innerhalb welcher Leitideen angewendet werden müssen. Allerdings zeigt eine genaue Betrachtung der Aufgabenbeispiele auch, dass der diagnostische Schwerpunkt auf den allgemeinen mathematischen Kompetenzbereichen liegt, da nur deren Teilkompetenzen detailliert aufgelistet werden.

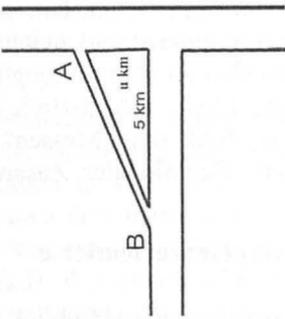
3.1 Kompetenzbereiche zur Lösung von Aufgabenvariationen

Die Annahmen darüber, welche Kompetenzbereiche bei welcher Variation der Beispielaufgabe in Tabelle A2 zu ihrer Lösung erforderlich sind, wurden *a priori* von Experten aus der Fachdidaktik, Mathematik und Psychologie erstellt und sind in Tabelle 1 zusammengefasst. Tabelle 1 wird in solch einem Zusammenhang auch als eine Aufgaben \times Kompetenz *Q-Matrix* (Tatsuoka, 1983) bezeichnet, wobei sie üblicherweise unterschiedliche Aufgaben – nicht Aufgabenvariationen – enthält.

Drei wichtige Merkmale von Aufgabenbeschreibungen anhand von Kompetenzen oder Kompetenzbereichen werden in Tabelle 1 verdeutlicht. Erstens ist erkennbar, dass die sechs Variationen des gleichen Aufgabenstamms die Aktivierung unterschiedlicher allgemeiner mathematischer Kompetenzbereiche erfordern. Während z.B. bei Variation I lediglich der Kompetenzbereich „mit symbolischen, formalen und technischen Elementen der Mathematik umgehen“ erforderlich ist, werden zur Lösung von Variation II be-

Tab. A2: Variationen einer Mathematikaufgabe zur Illustration nationaler Bildungsstandards:

Variation	Material	Aufgabe
I	---	Die Katheten eines rechtwinkligen Dreiecks sind 3 cm und 5 cm lang. Berechne die Länge der Hypotenuse.
II		Um von Punkt A nach Punkt B zu gelangen, benutzen viele Autofahrer lieber den direkten ‚Schleichweg‘ über die Nebenstraße, als über die sehr befahrenen Hauptstraßen zu fahren. In Prozent der Länge des Hauptstraßenweges, um wie viel kürzer ist der Schleichweg?
III		Um von Punkt A nach Punkt B zu gelangen, benutzen viele Autofahrer lieber den direkten ‚Schleichweg‘ über die Nebenstraße, als über die sehr befahrenen Hauptstraßen zu fahren. Wenn die Durchschnittsgeschwindigkeit auf den Hauptstraßen 50 km pro Stunde und auf der Nebenstraße 30 km pro Stunde beträgt, bringt der Schleichweg dann eine Zeitersparnis?
IV		Um von Punkt A nach Punkt B zu gelangen, benutzen viele Autofahrer lieber den direkten ‚Schleichweg‘ über die Nebenstraße, als über die sehr befahrenen Hauptstraßen zu fahren. Nimm an, dass die Durchschnittsgeschwindigkeit auf der Nebenstraße 30 km pro Stunde beträgt. Wie hoch darf die Durchschnittsgeschwindigkeit auf den Hauptstraßen höchstens sein, damit der Schleichweg eine Zeitersparnis bringt?

Variation	Material	Aufgabe
V		<p>Um von Punkt A nach Punkt B zu gelangen, benutzen viele Autofahrer lieber den direkten ‚Schleichweg‘ über die Nebenstraße, als über die zwei sehr befahrenen Hauptstraßen zu fahren. Nimm an, dass die Durchschnittsgeschwindigkeit auf der Nebenstraße 30 km pro Stunde beträgt. Wie hoch darf die Durchschnittsgeschwindigkeit auf den Hauptstraßen höchstens sein, damit der Schleichweg eine Zeitersparnis bringt?</p> <p>(a) Erstelle eine Tabelle für die Beziehung zwischen der Höchstgeschwindigkeit (v) und der Länge der einen Hauptstraße (u).</p> <p>(b) Erkläre – ohne Hilfe von Formeln – dass die Funktion in Abhängigkeit von u und v an einer Stelle ein Maximum hat.</p>
VI		<p>Um von Punkt A nach Punkt B zu gelangen, benutzen viele Autofahrer lieber den direkten ‚Schleichweg‘ über die <i>Canalettostraße</i> und die <i>Nicolaistraße</i> als über die mehr befahrenen Hauptstraßen <i>Stübelallee</i> und <i>Fetscherstraße</i> zu fahren. Unter welchen Bedingungen lohnt sich dieser Schleichweg? Begründe Deine Antwort.</p>

reits fünf allgemeine mathematische Kompetenzbereiche benötigt. Ähnliches gilt auch für die Leitideen, innerhalb derer diese aktiviert werden müssen.

Tab. 1: Aufgabenvariationen und zu deren Lösung erforderliche Kompetenzbereiche

Variation	Allgemeine mathematische Kompetenzbereiche						Inhaltsbezogene mathematische Kompetenzbereiche				
	AK	AK	AK	AK	AK	AK6	LI	LI	LI	LI	LI
	1	2	3	4	5		1	2	3	4	5
I	0	0	0	0	2	0	0	0	2	0	0
II	2	1	1	1	1	0	2	0	2	0	0
III	2	1	1	1	1	0	0	0	2	1	0
IV	2	1	1	1	1	0	0	0	2	2	0
V	2	2	2	2	2	2	0	0	2	1	0
VI	2	2	2	2	2	2	0	0	2	2	0

Anm. AK = Allgemeine mathematische Kompetenzbereiche, LI = Inhaltsbezogene mathematische Kompetenzbereiche bzw. Leitideen. AK 1 = mathematisch argumentieren, AK 2 = mathematisches Problemlösen, AK 3 = mathematisches Modellieren, AK 4 = mathematische Darstellungen verwenden, AK 5 = mit symbolischen, formalen und technischen Elementen der Mathematik umgehen, AK 6 = mathematisch kommunizieren. LI 1 = Zahl, LI 2 = Raum und Form, LI 3 = Messen, LI 4 = Funktionaler Zusammenhang, LI 5 = Daten und Zufall. 0 = Kompetenzbereich wird nicht benötigt, 1 = Kompetenzbereich wird benötigt, ist aber nicht essentiell, 2 = Kompetenzbereich wird benötigt und ist essentiell.

Zweitens zeigt Tabelle 1, dass nicht alle Kompetenzbereiche zur Lösung der Aufgabenvariationen in gleich starkem Ausmaß benötigt werden. Zur Lösung von Variation II etwa ist lediglich der allgemeine mathematische Kompetenzbereich „mathematisch argumentieren“ essentiell notwendig, während die allgemeinen mathematischen Kompetenzbereiche 2 bis 5 für die Aufgabenlösung nur unterstützend erforderlich sind.

Drittens veranschaulicht Tabelle 1, dass unterschiedliche Aufgabenvariationen die gleichen allgemeinen mathematischen Kompetenzbereiche innerhalb unterschiedlicher Leitideen aktivieren. So machen zum Beispiel die Variationen II und IV gleichermaßen die Aktivierung der allgemeinen mathematischen Kompetenzbereiche 1 bis 5 erforderlich, jedoch müssen diese für Variation II in den Leitideen „Zahl“ und „Messen“ und in Variation IV in den Leitideen „Messen“ und „Funktionaler Zusammenhang“ aktiviert werden.

3.2 Implikationen für eine kompetenzorientierte Unterrichtsentwicklung

Aus dem gerade skizzierten Beispiel kann im Hinblick auf eine Unterrichtsentwicklung, die auf die Diagnose und Förderung von einzelnen Kompetenzen und Kompetenzbereichen – und damit auf die Erfüllung der Bildungsstandards der KMK – ausgerichtet ist, Folgendes abgeleitet werden. Selbst durch kleine Veränderungen an sehr einfachen, möglicherweise derzeit bereits im Unterricht eingesetzten, Aufgaben lassen sich das Anforder-

rungsniveau der Aufgaben und die durch diese Aufgaben aktivierten Kompetenzen gezielt steuern.

Unterrichtsentwicklung im Hinblick auf die Bildungsstandards der KMK erfordert also vor allem einen sensiblen Umgang mit Aufgabenstellungen bezüglich der durch sie angesprochenen Kompetenzen. Solch eine Sensibilisierung ermöglicht es Lehrerinnen und Lehrern, *autonom kompetenzorientiert und standardbezogen* zu arbeiten. Voraussetzungen hierfür jedoch sind (a) eine Vertrautheit mit den Kompetenzmodellen, die der Entwicklung verschiedener Aufgaben zugrunde liegen sowie (b) ein Verständnis von theoretisch geleiteten Einschätzungen, welche Kompetenzen oder Kompetenzbereiche durch welche Art von Aufgabenstellung aktiviert werden. Einen Zugang zu diesen Informationen in Bezug auf das Fach Mathematik bilden Materialien wie das schon angesprochene Buch zur Illustration der Bildungsstandards (Blum et al., im Druck).

Wie dieses Beispiel illustriert hat, ist eine kompetenzorientierte Systematik in der Aufgabenentwicklung nicht nur für standardisierte Testverfahren nützlich, sondern kann auch bei der Entwicklung und der Auswertung von Unterrichtsaufgaben zur formativen Evaluation und Diagnose von Schülerinnen und Schülern nützlich sein. Im folgenden Abschnitt wird nun gezeigt, wie kognitiv-diagnostische Modelle konkret aussehen und welche Rückmeldungen sie erlauben.

4. Kognitiv-diagnostische Modelle für Kompetenzdiagnostik

In diesem Abschnitt wird zunächst ein kurzer Überblick über die derzeitige Praxis der Kompetenzdiagnostik in *large-scale assessment*-Studien gegeben. In einem zweiten Teil werden diesem Vorgehen die Grundgedanken der kognitiv-diagnostischen Modellierung gegenüber gestellt. In einem dritten Teil wird mit dem DINA Modell beispielhaft ein kognitiv-diagnostisches Modell vorgestellt. Schließlich werden ein Ausblick auf alternative Modelle gegeben und modellübergreifende Herausforderungen an Verfahren zur kognitiv-diagnostischen Modellierung verdeutlicht.

4.1 Traditionelle Methoden der Kompetenzdiagnostik

Generell kann man die Skalierung von Aufgaben und Lernenden in *large-scale assessment*-Studien in vier Kernschritte aufteilen. In einem ersten Schritt werden Aufgaben üblicherweise vor ihrer Darbietung von Experten hinsichtlich ihres Anforderungscharakters im Hinblick auf mehrere Kompetenzen eingeschätzt (z.B. Watermann & Klieme, 2002). Wie schon in früheren Abschnitten angesprochen, wurden bei PISA 2003 zur Klassifikation der Aufgaben im Fach Mathematik vier Leitideen sowie drei Kompetenzcluster benutzt, wobei die drei Kompetenzcluster acht generelle mathematische Kompetenzbereiche zusammenfassen und nach Komplexität geordnet sind. Die Leitideen und Kompetenzcluster waren bei PISA 2003 insofern für die Testentwicklung konstitutiv, als dass der Test so gestaltet wurde,

dass jede der vier Leitideen durch etwa gleiche Anzahlen von Aufgaben repräsentiert war. Dabei entfielen etwa die Hälfte aller Aufgaben auf das mittlere Kompetenzcluster und je ein Viertel der Aufgaben auf das untere und obere Kompetenzcluster.

In einem zweiten Schritt werden die Aufgaben dann einer ausreichend großen Stichprobe von Schülerinnen und Schülern dargeboten, so dass sowohl die Aufgaben als auch die Lernenden auf eindimensionalen Kontinua angeordnet werden können. Dazu werden üblicherweise Modelle der *probabilistischen Testtheorie* (z.B. Embretson & Reise, 2000; Rost, 2004a), allen voran das *Rasch Modell* (z.B. Fisher & Molenaar, 1995; Rasch, 1960) für dichotom kodierte Aufgaben sowie das *Partial-credit Modell* (z.B. Masters, 1982) für polytom kodierte Aufgaben herangezogen.²

In einem dritten Schritt werden die eindimensionalen Fähigkeitskontinua dann in Teilbereiche, so genannte *Kompetenzstufen* oder *Kompetenzniveaus*, zerlegt. Formal sind diese Kompetenzstufen so definiert, dass Schülerinnen und Schüler in der ihnen zugeordneten Kompetenzstufe so genannte „Schwellenitems“ einer Stufe mit einer bestimmten Wahrscheinlichkeit, Schwellenitems höherer Kompetenzstufen allerdings nur mit sehr viel geringerer Wahrscheinlichkeit lösen können. Diese Wahrscheinlichkeit lag zum Beispiel bei 62% in PISA 2003 (Prenzel, Drechsel, Carstensen & Ramm, 2004, S. 37) sowie 65% in TIMSS und DESI (Hartig, im Druck, S. 97).

In einem vierten Schritt erfolgt die inhaltliche Füllung der so gebildeten Kompetenzstufen durch die vor der Testung im ersten Schritt vorgenommenen Aufgabencharakterisierungen. Diejenigen Mengen von Aufgaben, deren Skalenwerte im jeweiligen Intervall einer Kompetenzstufe liegen, werden sorgfältig im Hinblick auf ihren Anforderungscharakter bezüglich der mathematischen Kompetenzen oder Kompetenzbereiche beschrieben. Das Ergebnis ist eine inhaltliche – jedoch statistisch gesehen separate – Beschreibung der Aufgaben jeder Kompetenzstufe und damit auch der Personen, die den Kompetenzstufen zugeordnet sind (Blum et al., 2004, S. 55f). Anders ausgedrückt werden auf diese Weise verschiedene Niveaus einer *quantitativen* Globalfähigkeit durch Aufgabenanforderungen anhand von *qualitativ* unterschiedlichen Teilkompetenzen *inhaltlich* charakterisiert.

Technisch gesehen ist die methodische Bandbreite zur Definition von Kompetenzstufen allerdings noch etwas größer. Die Festlegung der Kompetenzstufen kann, vereinfachend ausgedrückt, auf drei verschiedene Arten erfolgen, welche die oben beschriebenen Kernschritte unterschiedlich nutzen (Rost 2004b). So können die Kompetenzstufen durch (1) die Schwierig-

2 Das eindimensionale Rasch Modell ist in Tabelle 2 wiedergegeben. Seine Modellgleichung macht deutlich, dass die Verortung von Schülerinnen und Schüler auf einer kontinuierlichen Fähigkeitsdimension (θ) geschieht, auf der auch die Aufgaben durch ihre Schwierigkeitsparameter (β) verortet werden können.

keitsparameter der Aufgaben bzw. Antwortkategorien oder die Verteilung der Personenmesswerte, durch (2) die Charakterisierung der Testaufgaben durch Aufgabenmerkmale, sowie durch (3) unterschiedliche Lösungsprofile der Testaufgaben definiert werden. Hierbei werden dann verschiedene methodische Verfahren eingesetzt, welche von eindimensionalen Raschmodellen im ersten Fall über linear-logistische Testmodelle (z.B. Fisher, 1997; Hartig & Frey, 2005) im zweiten Fall bis hin zu Mischverteilungsmodellen (z.B. Rost, 1990) im dritten Fall reichen. Allerdings ist anzumerken, dass solch quantitativen Vorgehensweisen des *proficiency scaling* (z.B. Beaton & Allen, 1992) oftmals mit eher qualitativ orientierten Verfahren des *standard setting* (z.B. Zieky & Perie, 2006) empirisch zu einander in Bezug gesetzt werden müssen. Dies ist notwendig, um Kompetenzstufen zu definieren, die nicht nur statistisch sondern auch politisch für Entscheidungsträger vertretbar sind (Haertel & Loricé, 2004).

4.2 Grundgedanken kognitiv-diagnostischer Modellierung

Im Vergleich zu traditionellen Modellen der probabilistischen Testtheorie, inklusive der linear-logistischen Testmodelle und der Mischverteilungsmodelle, verwenden kognitiv-diagnostische Modelle die Einschätzung der Aufgaben bezüglich der von ihnen benötigten Kompetenzen oder Kompetenzbereiche *direkt zur multiplen Klassifikation der Lernenden* (für eine Diskussion siehe z.B. de la Torre & Douglas, 2004; Maris, 1999; Rupp, im Druck; Rupp & Mislevy, im Druck; Templin & Henson, 2005a; von Davier, 2005). Somit ist die Zielsetzung also keine *quantitative Verortung*, sondern eine *qualitative Klassifikation* anhand von mehreren Kompetenzen.

Statistisch gesehen unterscheiden sich kognitiv-diagnostische Modelle von traditionellen Modellen wie dem Rasch-Modell im Kern also dadurch, dass sie die erhobenen Schülerleistungen für jede Aufgabe auf mindestens eine latente Variable beziehen und dass diese latenten Variablen nicht-kontinuierlich sind. Anders ausgedrückt handelt es sich bei ihnen um multidimensionale Faktorenmodelle mit kategorialen latenten Variablen, die eine komplexe Faktorenstruktur abbilden können. Sie können demnach als *multiple Klassifikationsmodelle* (*multiple classification models*) oder auch als *restringierte latente Klassenmodelle* (*restricted latent class models*; Begriff nach Haertel, 1989) angesehen werden. Die Restriktion rührt daher, dass, anders als bei anderen latenten Klassenmodellen, die Anzahl der möglichen Klassen durch die Anzahl der betrachteten Kompetenzen sowie deren Abstufungen (z.B. ‚0-1‘ bzw. ‚vorhanden - nicht vorhanden‘ oder ‚0-1-2‘ bzw. ‚vorhanden - teilweise vorhanden - nicht vorhanden‘) vorbestimmt ist.

Der wichtigste Output dieser Modelle für Einzelschüler ist ein Kompetenzprofil, das den Lernenden aufzeigt, mit welcher Wahrscheinlichkeit sie jede der postulierten Kompetenzen oder Kompetenzbereiche schon gelernt haben. Als illustratives Beispiel zeigt Abbildung 1 ein solches Kompetenzprofil für die oben angesprochenen sechs allgemeinen mathematischen Kompe-

tenzbereiche. In der Abbildung ist für jeden Kompetenzbereich auf der dazugehörigen Skala die Wahrscheinlichkeit abgetragen, dass der Schüler diesen Kompetenzbereich schon zufrieden stellend erlernt hat.

In der Abbildung sieht man einen Schüler, der mit hoher Wahrscheinlichkeit die Kompetenzbereiche „mathematisches Problemlösen“ (AK 2), „mathematisch modellieren“ (AK 3) und „mathematische Darstellungen verwenden“ (AK 4) gemeistert hat, allerdings noch einen eindeutigen Lernbedarf in den Kompetenzbereichen „mathematisch argumentieren“ (AK 1), „mit symbolischen, formalen, und technischen Elementen der Mathematik umgehen“ (AK 5) und „mathematisch kommunizieren“ (AK 6) hat.

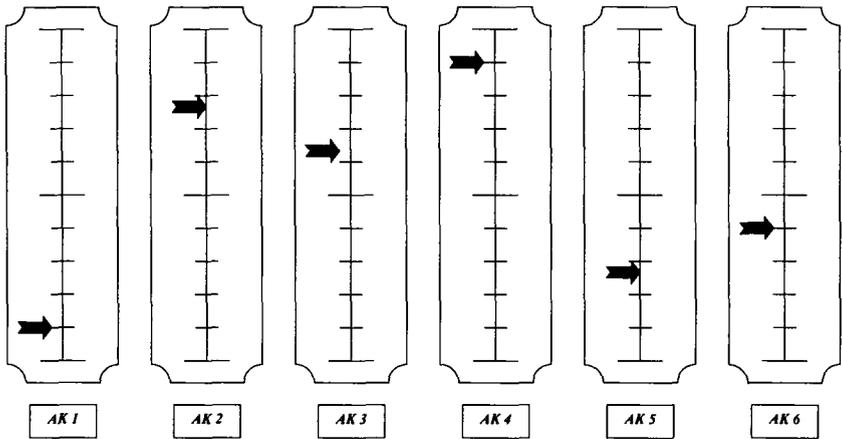


Abb. 1. Ein graphisches Kompetenzprofil eines Schülers in Bezug auf die sechs allgemeinen mathematischen Kompetenzbereiche.

Der wichtige Vorteil solcher Rückmeldungen ist die Identifizierung von konkretem individuellem Förderbedarf *auf der Basis statistischer Modelle*, da diese unerwartetes Antwortverhalten der Lernenden probabilistisch mit in die multiplen Klassifizierungen einbeziehen und so verlässliche Aussagen ermöglichen.

Auf der Basis solcher multiplen Klassifikationen von Schülerinnen und Schülern geben die Modelle ferner Auskunft über die relativen Schwierigkeiten einzelner Kompetenzen oder Kompetenzbereiche in Schülerpopulationen. Als illustratives Beispiel zeigt Abbildung 2 die Beziehung zwischen den globalen Summenscores eines Mathematiktests und den ersten beiden allgemeinen mathematischen Kompetenzbereichen „mathematisch argumentieren“ und „mathematisches Problemlösen“.

In der nachfolgenden Abbildung 2 sieht man deutlich, wie Schüler mit einer höheren Mathematikkompetenz (d.h. mit einem höheren Summenscore) auch die beiden anderen Kompetenzbereiche mit einer höheren Wahr-

scheinlichkeit erlernen. Dieses Verhältnis ist aber beim Kompetenzbereich „mathematisches Problemlösen“ noch stärker ausgeprägt als beim Kompetenzbereich „mathematisch argumentieren“. Mit anderen Worten, der Kompetenzbereich „mathematisch argumentieren“ ist für diese Schülerpopulation insgesamt schwieriger zu erlernen als der Kompetenzbereich „mathematisches Problemlösen“, weshalb er noch intensiver mit kompetenzorientierten Aufgaben geübt werden sollte. Anhand von ähnlichen Abbildungen kann man ebenfalls optimale Lernwege für Schüler mit unterschiedlichen Kompetenzprofilen darstellen, welche den bestmöglichen Leistungszuwachs in Bezug auf die gemessenen Kompetenzbereiche aufzeigen (z.B. Tatsuoka, 1995; Henson & Templin, 2004).

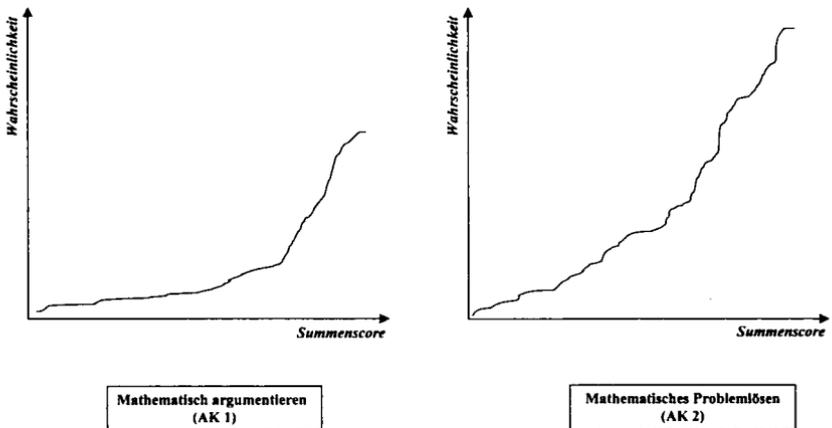


Abb. 2: Die Wahrscheinlichkeit (P) des Erlernens von zwei spezifischen Kompetenzen in Beziehung zum Summenschore einer Globalfähigkeit

Einige Modelle stellen zudem Informationen über die *kognitiven Trennschärfe* der Aufgabe bereit, da die theoretisch hergeleiteten Einträge in der *Q-matrix* empirisch korrigiert werden können. Dies zeigt an, wie gut einzelne Aufgaben Rückschlüsse über die postulierten Kompetenzen zulassen. Schließlich erlauben kognitiv-diagnostische Modelle mit Hilfe von Fitstatistiken die Durchführung von Hypothesentests über die postulierten kognitiven Prozesse, die bei der Aufgabenbearbeitung zugrunde liegen, da diese durch die im Modell verwendete *Q-matrix* operationalisiert werden. Wie kompetenzdiagnostische Modelle mathematisch aussehen, soll nun anhand des DINA Modells illustriert werden.

4.3 Das DINA-Modell

Das ‚deterministic inputs, noisy „and“ gate‘ (*DINA*) Modell (zurückgehend auf Macready & Dayton, 1977; bezeichnet als DINA von Junker & Sijtsma, 2001) ist ein *nichtkompensatorisches Modell*. Nichtkompensatorisch ist es in dem Sinn, dass alle zur Aufgabenlösung erforderlichen Kompetenzen

vorhanden sein müssen, um diese tatsächlich zu lösen und somit einen Score von ‚1‘ zu erhalten³. Das stochastische Element des DINA-Modells (*noisy „and“ gate*) besteht darin, dass auch das Vorhandensein aller zur Aufgabenlösung erforderlicher Kompetenzen nicht deterministisch, sondern nur probabilistisch, eine tatsächliche korrekte Lösung bedingt (vgl. de la Torre & Douglas, 2004).

Die Wahrscheinlichkeit der richtigen Lösung einer Aufgabe wird durch zwei Modellparameter bestimmt. Die so genannte *guessing*-Wahrscheinlichkeit (g_j) repräsentiert dabei die Wahrscheinlichkeit einer richtigen Aufgabenlösung durch Raten beim Fehlen einer oder mehrerer Kompetenzen. Die so genannte *slipping*-Wahrscheinlichkeit (s_j) hingegen repräsentiert das unerwartete falsche Lösen einer Aufgabe trotz des Verfügens über die zu ihrer Lösung erforderlichen Kompetenzen. Junker und Sijtsma (2001) sprechen von g_j und s_j als „Irrtumswahrscheinlichkeiten“ im Sinne falsch positiver (*guessing*) bzw. falsch negativer (*slipping*) Antwortmuster. Diese Irrtumswahrscheinlichkeiten beeinflussen dann das nicht direkt beobachtbare Antwortverhalten einer Person i im Hinblick auf die Aufgabe j ; letzteres wird bei ihnen symbolisch als ξ_{ij} bezeichnet.

Das Zusammenwirken von ξ_{ij} , s_j und g_j bei der Bestimmung der Lösungswahrscheinlichkeit einer Aufgabe j durch eine Person i wird in der Formel des DINA-Modells deutlich, die in Tabelle 2 dargestellt ist. An den Modellparametern g_j und s_j kann erkannt werden, dass für *jede* Aufgabe nur *ein* *slipping*- und *ein* *guessing*-Parameter geschätzt werden. Diese können dann als ein Maß zur Einschätzung der Modellgeltung dienen, da sehr hohe *slipping*- oder *guessing*-Wahrscheinlichkeiten auf die Fehlpassung einer Aufgabe innerhalb des Modells oder eine Fehlspezifikation des Modells insgesamt hinweisen.

Für jede Schülerin und jeden Schüler erhält man durch die Kalibrierung dieses Modells die Wahrscheinlichkeiten, dass sie jede einzelne Kompetenz erworben haben sowie eine Zuordnung zu demjenigen Kompetenzprofil, das global am wahrscheinlichsten für sie ist, was schon im vorigen Abschnitt illustriert wurde. Ferner erhält man für jede Aufgabe die geschätzten *slipping*- und *guessing*-Parameter sowie deren Standardfehler und eine Reihe von Fitstatistiken. Software zur Schätzung des DINA Modells liegt mit einem frei verfügbaren Code vor⁴, der das Modell – und einige andere – mit Hilfe des Softwareprogramms Mplus (Muthén & Muthén, 2006) unter Windows-XP schätzt.

3 Das Modell unterteilt die Personen, die eine bestimmte Aufgabe lösen, somit für jede Aufgabe immer in zwei Klassen: Eine Gruppe, die alle benötigten Kompetenzen besitzt und eine andere Gruppe, der mindestens eine benötigte Kompetenz fehlt.

4 Der interessierte Leser kontaktiere Jon Templin von der Kansas University in den U.S.A. (jtemplin@ku.edu).

Tab. 2: Ausgewählte statistische Modelle zur Kompetenzdiagnostik

Modell	Mathematische Formel	Komponenten	Wertebereich
Rasch	$P(X_{ij} = 1) = \frac{\exp(\theta_i - \beta_j)}{1 + \exp(\theta_i - \beta_j)}$	θ_i β_j	$-\infty < \theta_i < \infty$ $-\infty < \beta_j < \infty$
DINA	$P(X_{ij} = x_{ij}) = \left[s_j^{1-x_{ij}} (1-s_j)^{x_{ij}} \right] \left[g_j^{x_{ij}} (1-g_j)^{1-x_{ij}} \right]^{1-\xi_{ij}}$	$s_j = P(X_{ij} = 0 \xi_{ij} = 1)$ $g_j = P(X_{ij} = 1 \xi_{ij} = 0)$ $\xi_{ij} = \prod_{k=1}^K \alpha_{ik}^{q_{jk}}$	$0 \leq s_j \leq 1$ $0 \leq g_j \leq 1$ $\xi_{ij} \in \{0, 1\}$
NIDA	$P(X_{ij} = x_{ij}) = \left\{ \prod_{k=1}^K [(1-s_k)^{\alpha_k} g_k^{1-\alpha_k}]^{q_{jk}} \right\} \left\{ 1 - \prod_{k=1}^K [(1-s_k)^{\alpha_k} g_k^{1-\alpha_k}]^{q_{jk}} \right\}^{1-x_{ij}}$	$s_k = P(\zeta_{ijk} = 0 \alpha_{ik} = 1, q_{jk} = 1)$ $g_k = P(\zeta_{ijk} = 1 \alpha_{ik} = 0, q_{jk} = 1)$	$0 \leq s_k \leq 1$ $0 \leq g_k \leq 1$ $\zeta_{ijk} \in \{0, 1\}$

Anm.: x_{ij} = Score für Person i und Item j ; θ_i = Kontinuierlicher Fähigkeitsparameter für Person i ; α_{ik} = Diskreter Kompetenzparameter für Person i im Hinblick auf Kompetenz k ; β_j = Kontinuierlicher Schwierigkeitsparameter für Item j ; s_j bzw. s_k = Slipping-Wahrscheinlichkeit für ein Item j bzw. eine Kompetenz k ; g_j bzw. g_k = Ratewahrscheinlichkeit für ein Item j bzw. eine Kompetenz k ; ξ_{ij} = Diskreter Kompetenzparameter für Person i und Item j ; ζ_{ijk} = Diskreter Antwortparameter für Person i , Item j , und Kompetenz k ; q_{jk} = Diskreter Eintrag in der Q-matrix für Kompetenz k und Item j . NIDA = Noisy input, deterministic, and' gate; DINA = Deterministic input, noisy 'and' gate.

4.4 Überblick über alternative Modelle der Kompetenzdiagnostik

Neben dem DINA-Modell existieren noch eine Reihe alternativer Modelle zur Kompetenzdiagnostik, die in vielerlei Hinsicht von diesem Modell abgegrenzt werden können. Auf struktureller Ebene etwa unterscheidet sich das ‚noisy inputs, deterministic „and“ gate (NIDA)‘ Modell (z.B. Junker & Sijtsma, 2001) vom DINA Modell dadurch, dass slipping- und guessingparameter für einzelne Kompetenzen – und nicht Aufgaben – geschätzt werden; dieses Modell ist auch in Tabelle A3 im Anhang aufgeführt. Andere Modelle, wie zum Beispiel das ‚Reparametrized Unified / Fusion‘ Modell (z.B. Hartz, 2002; Templin & Henson, 2005b), beinhalten Parameter, die explizit die Vollständigkeit und die Komplexität der Q-Matrix modellieren. Parallel hierzu sind ähnliche Modelle auch in anderen Disziplinen wie der Biometrik und der mathematischen Psychologie vorgestellt worden (z.B. Maris, 1999; Ünlü, im Druck).

So sind in den letzten Jahren Modelle entwickelt oder weiterentwickelt worden, die polytome Itemantworten modellieren (z.B. Bolt & Fu, 2004; Templin, He, Roussos, & Stout, 2003), polytome Kompetenzklassen unterscheiden (z.B. Templin, Roussos & Stout, 2003; von Davier, 2005), hierarchische Kompetenzstrukturen abbilden (z.B. de la Torre & Douglas, 2004; Templin & Henson, 2005a) und für komplexe Simulationsaufgaben eingesetzt werden können (z.B. Yan, Mislevy, & Almond, 2003; Williamson, Almond, Mislevy, & Levy, 2006). Ferner gibt es noch kompetenzdiagnostische Modelle, die für Daten aus Langzeitstudien, zur Beschreibung von Progressionen durch verschiedene Entwicklungsstadien, und zur Modellierung von verschiedenen Strategien angemessen sind. Für eine detaillierte Beschreibung solcher Modelle zur kognitiven Kompetenzdiagnostik wird hier unter anderem auf Junker (1999), Rupp und Mislevy (im Druck) sowie Leucht und Rupp (in Vorbereitung) verwiesen.

4.5 Herausforderungen für eine kognitiv-diagnostische Modellierung

Unabhängig davon, welches statistische Modell zur Kompetenzdiagnostik herangezogen wird, muss auf konzeptioneller Ebene immer entschieden werden, auf welcher Detailtiefe Rückmeldungen an Schülerinnen und Schüler gewünscht sind. So macht es etwa einen Unterschied, ob bei der Modellierung ein eher grob gefasster Kompetenzbereich wie „mit symbolischen, formalen und technischen Elementen der Mathematik umgehen“ (KMK, 2004a, S. 7) oder eine eher eng umrissene Kompetenz wie „Brüche addieren und subtrahieren“ (z.B. Junker & Sijtsma, 2001; Yan, Mislevy & Almond, 2003) betrachtet werden soll. Mit der Frage der Detailtiefe sind auch statistische Abwägungen eng verbunden. So haben verschiedene Studien (z.B. Hartz, 2002; Templin & Henson, 2005a) gezeigt, dass man schon für Modelle mit sechs bis acht Kompetenzen mehrere hundert, besser aber tausend Probanden benötigt, um die Modellparameter zuverlässig schätzen zu können. Somit kann die Verwendung dieser Modelle in der Praxis auch als

ein Kompromiss verstanden werden, die statistischen Freiheitsgrade in einer praktischen Situation angemessen zu nutzen.

Als Beispiel sei hier eine fiktive Situation genannt, in der – bedingt durch die Anzahl der Aufgaben und der Schüler – nur sechs Kompetenzen mit einem bestimmten kognitiv-diagnostischen Modell zuverlässig gleichzeitig modelliert werden können. Dies hätte zur Konsequenz, dass entweder eine ohnehin schon fokussierte Kompetenz wie ‚Brüche subtrahieren‘ ziemlich präzise in ihre konstitutiven Teilkompetenzen aufgeschlüsselt werden könnte oder aber ein gröber definierter Kompetenzbereich wie ‚algebraische Fähigkeit‘ nur in ebenfalls gröber definierte Teilkompetenzen aufgeschlüsselt werden könnte. Mit anderen Worten, während im ersten Fall der Problemlöseprozess sehr detailliert abgebildet werden könnte und sehr detaillierte Rückmeldungen an Schülerinnen und Schüler gegeben werden könnten, wäre dies im zweiten Fall nur weitaus undifferenzierter möglich. Allerdings bedeutet dies natürlich nicht, dass ein bestimmtes Szenario einem anderen generell vorzuziehen wäre, da, wie gesagt, die Detailtiefe der Rückmeldung immer von der praktischen Zielsetzung der Untersuchung abhängt. Dies unterstreicht ein weiteres Mal die Notwendigkeit, vor der Untersuchung eindeutig zu präzisieren, welche Aussagen über welche Form der Kompetenzstruktur mit den Ergebnissen der jeweiligen Untersuchung ermöglicht werden sollen (siehe Mislevy, Steinberg, & Almond, 2003).

Letztendlich ist es auch unabdingbar, dass die Beschreibungen der Kompetenzen und die dadurch resultierenden Charakterisierungen von Aufgaben vertrauenswürdig sind, da sie den theoretischen Kerninput für die kognitiv-diagnostischen Modelle darstellen. Dies sollte durch einen empirisch abgesicherten Konsens von Experten geschehen, wobei ein fehlender Konsens für einige Kompetenzzuordnungen nicht unbedingt problematisch sein muss. Alternative Charakterisierungen von Aufgaben durch verschiedene Personengruppen wie Lehrer, Fachdidaktiker, und Wissenschaftler könnten zum Beispiel empirisch genutzt werden, um die daraus resultierenden Klassifikationen der Lernenden unter Verwendung der verschiedenen Q-Matrizen zu vergleichen. Wenn die Kompetenzbeschreibungen und die Klassifizierungen der Aufgaben allerdings nicht theoretisch verankert werden können, so können auch die Entscheidungen auf der Basis der kognitiv-diagnostischen Modelle nicht valide sein, was als *garbage-in-garbage-out* Phänomen bekannt ist.

Ein wichtige praktische Erwägung bei der Arbeit mit kognitiv-diagnostischen Modellen, ebenso wie mit traditionellen statistischen Modellen, ist die Entwicklung von Rückmeldesystemen, die Schülern, Lehrern und anderen Entscheidungsträgern die relevanten Informationen auf einer gewünschten Detailtiefe so darstellen, dass sie als nützlich wahrgenommen werden. Im Bundesland Nordrhein-Westfalen etwa werden Eltern im Kontext der Lernstandserhebungen Kompetenzprofile und Referenzwerte zur Interpretation

der Ergebnisse ihrer Kinder zur Verfügung gestellt (siehe www.learnline.nrw.de/angebote/learnstand9/). Dabei erfolgen die Rückmeldungen explizit auf der Ebene inhaltlich-fachlicher Anforderungen und der von diesen Anforderungen angesprochenen Kompetenzen oder Kompetenzbereichen (vgl. etwa Leutner, Wirth & Fleischer, 2004).

Auch bei der Normierung der länderübergreifenden Bildungsstandards wäre eine individuelle Rückmeldung über Kompetenzprofile als diagnostische Zusatzinformation zur Förderung von Lernenden wünschenswert, welche durch die in diesem Artikel vorgestellten Modelle statistisch abgesichert werden könnten. Auf diese Weise könnte ein Einsatz kognitiv-diagnostischer Modelle bei der Diagnose von Schülerleistungen und der Gestaltung von Rückmeldeformaten nachhaltig zu einer Qualitätsentwicklung deutscher Schulen im Hinblick auf die Erfüllung der Bildungsstandards der KMK beitragen.

5. Zusammenfassung und Schlussbetrachtungen

In diesem Beitrag wurden kognitiv-diagnostische Modelle für Kompetenzdiagnostik vorgestellt, die als restringierte latente Klassenmodelle oder multiple Klassifikationsmodelle aufgefasst werden können. Ihr Potenzial für eine auf qualitativ verschiedene Schülerkompetenzen bezogene pädagogisch-psychologische Diagnostik wurde verdeutlicht. Es wurde gezeigt, was aus der Systematik zur Aufgabenentwicklung in der Testung für eine kompetenzorientierte Aufgabenkultur im Unterricht gelernt werden kann. Schliesslich wurden theoretische Vorteile und praktische Grenzen kognitiv-diagnostischer Modelle diskutiert und anhand des DINA Modells verdeutlicht.

Dabei können nach Auffassung der Autoren Rückmeldungen über Kompetenzprofile, Kompetenzerwerb und Kompetenzentwicklung zur Entwicklung von Fördermaßnahmen von einer Verwendung der beschriebenen Modelle nur profitieren. Deshalb erscheint es sinnvoll, die praktische Einsetzbarkeit von kognitiv-diagnostischen Modellen weiter zu erforschen und im Rahmen der Normierung der länderübergreifenden Bildungsstandards der KMK zu erproben. Diese herausfordernde Aufgabe bedarf einer Kollaboration von Experten verschiedener Disziplinen wie der Fachdidaktik, der Psychologie und der Psychometrik und ist ein dezidiertes Ziel einer Forschungsgruppe am Institut zur Qualitätsentwicklung im Bildungswesen in Berlin.

Literatur

Baumert, J., Bos, W. & Lehmann, R. (Hrsg.). (2000). TIMSS III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit. Opladen: Leske & Budrich.

- Baumert, J., Klieme, E., Neubrand, M., Prenzel, M., Schiefele, U., Schneider, W., Stanat, P., Tillmann, K.J. & Weiß, M. (Hrsg.) (2001). PISA 2000 - Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich. Opladen: Leske & Budrich.
- Beaton, A.E. & Allen, N. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17, 191-204.
- Blum, W., Drüke-Noe, K., Hartung, R., & Köller, O. (Hrsg.). (im Druck). Bildungsstandards Mathematik konkret. Sekundarstufe I: Aufgabenbeispiele, Unterrichtsideen, und Fortbildungsmöglichkeiten. Berlin: Cornelsen / Scriptor Verlag.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfing, F. & Carstensen, C. (2004). Mathematische Kompetenz. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs. Münster: Waxmann.
- Bolt, D. & Fu, J. (2004, April). A polytomous extension of the fusion model and its Bayesian parameter estimation. Artikel präsentiert auf der Jahrestagung des National Council on Measurement in Education, San Diego, CA.
- Bolt, D., & Lall, V. F. (2003). Estimation of compensatory and noncompensatory multidimensional item response models using Markov Chain Monte Carlo. *Applied Psychological Measurement*, 27, 395-514.
- Bos, W., Lankes, E.-M., Prenzel, M., Schwippert, K., Valtin, R., & Walther, G. (2003). Erste Ergebnisse aus IGLU: Schülerleistungen am Ende der vierten Jahrgangsstufe im internationalen Vergleich. Verfügbar online auf www.erzwiss.uni-hamburg.de/IGLU/home.htm
- de la Torre, J. & Douglas, J.A. (2004). Higher-Order latent trait models for cognitive diagnosis. *Psychometrika*, 69, 333-353.
- Embretson, S. E. (1983). Construct validity : Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179-197.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. Mahwah, NJ: Erlbaum.
- Fisher, G. H. (1997). Unidimensional linear logistic Rasch models. In W. J. van der Linden & R. K. Hambleton (Hrsg.), *Handbook of modern item response theory* (pp. 225-244). New York: Springer.
- Fischer, G.H. & Molenaar, I.W. (1995). *Rasch-Models – Foundations, recent developments, and applications*. New York: Springer.
- Freudenthal, H. (1977). *Mathematik als pädagogische Aufgabe* (Bd. 1, Bd. 2). Stuttgart: Klett.
- Freudenthal, H. (1983). *Didactical phenomenology of mathematical structures*. Dordrecht: Reidel.
- Haertel, E.H. (1989). Using restricted latent class models to map the skill structure of achievement items. *Journal of Educational Measurement*, 26, 301-321.
- Haertel, E. H., & Lorie, W. A. (2004). Validating standards-based test score interpretations. *Measurement: Interdisciplinary Research and Perspectives*, 2, 61-103.

- Hartig, J. (im Druck). Skalierung und Kompetenzniveaus. In E. Klieme & B. Beck (Hrsg.). (im Druck). Sprachliche Kompetenzen - Konzepte und Messung (pp. 95-112). Weinheim: Beltz
- Hartig, J., & Frey, A. (2005, Juli). Application of different explanatory item response models for model based proficiency scaling. Vortrag auf dem 14. internationalen Treffen der Psychometric Society, Tilburg, NL.
- Hartz, S. M. (2002). A Bayesian framework for the Unified Model for assessing cognitive abilities: Blending theory with practicality. Unpublished doctoral dissertation, University of Illinois, Urbana-Champaign, Department of Statistics.
- Henson, R. A., & Templin, J. L. (2004). Creating a proficiency scale with models for cognitive diagnosis. Unpublished Technical Report for the ETS External Diagnostic Group.
- Junker, B. W. (1999). Some statistical models and computational methods that may be useful for cognitively-relevant assessment. Unpublished manuscript. Available online at <http://www.stat.cmu.edu/~brian/nrc/cfa>
- Junker, B.W. & Sijtsma, K. (2001). Cognitive assessment models with few assumptions, and connections with nonparametric item response theory. *Applied Psychological Measurement*, 25, 258-272.
- Kanning, U. P. (2003). Diagnostik sozialer Kompetenzen. *Kompendien Psychologische Diagnostik (Band 4)*. Wien: Hogrefe.
- Klieme, E., & Beck, B. (Hrsg.). (im Druck). Sprachliche Kompetenzen - Konzepte und Messung. Weinheim: Beltz
- 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., Baumert, J., Köller, O. & Bos, W. (2000). Mathematische und naturwissenschaftliche Grundbildung: Konzeptuelle Grundlagen und die Erfassung und Skalierung von Kompetenzen. In J. Baumert, W. Bos & R. Lehmann (Hrsg), TIMSS III. Dritte Internationale Mathematik- und Naturwissenschaftsstudie – Mathematische und naturwissenschaftliche Grundbildung am Ende der Pflichtschulzeit (S. 85-133). Opladen: Leske & Budrich.
- KMK. (2003). Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Mathematik für den mittleren Schulabschluss (Beschluss von 04. Dezember 2003). München: Wolters Kluwer Deutschland.
- KMK. (2004a). Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Mathematik für den Hauptschulabschluss (Beschluss von 15. Oktober 2004). München: Wolters Kluwer Deutschland.
- KMK. (2004b). Bildungsstandards der Kultusministerkonferenz: Erläuterungen zur Konzeption und Entwicklung. München: Wolters Kluwer Deutschland.
- KMK. (2005). Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss (Beschluss von 16. Dezember 2004). München: Wolters Kluwer Deutschland.
- Kohler, B. & Schrader, F.W. (Hrsg.) (2004). Ergebnisrückmeldung und Rezeption – Von der externen Evaluation zur Entwicklung von Schule und Unterricht. Sonderheft der Zeitschrift *Empirische Pädagogik*, 18.

- Köller, O., Watermann, R., Trautwein, U. & Lüdtke, O. (Hrsg.) (2003). TOS-CA – Eine Untersuchung an allgemein bildenden und beruflichen Gymnasien. Opladen: Leske & Budrich.
- Kubinger, K. D. (2006). Klinische Diagnostik. Wien: Hogrefe.
- Leucht, M., & Rupp, A. A. (in Vorbereitung). Kompetenzdiagnostik mit statistischen Modellen: Eine Synthese theoretischer Forschungsergebnisse und praktischer Anwendungen. Manuskript in Vorbereitung.
- Leutner, D., Wirth, J. & Fleischer, J. (2004). Zentrale Lernstandserhebungen in der Jahrgangsstufe 9 im Jahr 2004 in NRW. Erster Kurzbericht zur wissenschaftlichen Begleitung. Universität Duisburg-Essen.
- Macready, G. B., & Dayton, C. M. (1977). The use of probabilistic models in the assessment of mastery. *Journal of Educational Statistics*, 99-120.
- Maris, E. (1995). Psychometric latent response models. *Psychometrika*, 60, 523-547.
- Maris, E. (1999). Estimating multiple classification latent class models. *Psychometrika*, 64, 187-212.
- Masters, G.N. (1982). A Rasch Model for partial credit scoring. *Psychometrika*, 47, 149-174.
- Mislevy, R. J., Steinberg, L. S. & Almond, R.G. (2003). On the structure of educational assessments. *Measurement: Interdisciplinary Research and Perspectives*, 1, 3-67.
- Muthén, L. K., & Muthén, B. O. (2006). Mplus Version 4.1 [Computer software]. Los Angeles, CA: Muthén & Muthén.
- Nichols, P. D., Chipman, S. F., & Brennan, R. L. (Hrsg.). (1995). *Cognitively diagnostic assessment*. Hillsdale, NJ: Erlbaum.
- Niss, M. (2003). Mathematical competencies and the learning of mathematics: The Danish KOM Project. In A. Gagatsis & S. Papastavridis (Hrsg.), 3rd Mediterranean conference on mathematical education. Athens – Hellas 3-4-5 January 2003 (S. 115-124). Athens: The Hellenic Mathematical Society.
- OECD (2003). *The PISA 2003 assessment framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- Prenzel, M., Drechsel, B., Carstensen, C. & Ramm, G. (2004). PISA 2003 – Eine Einführung. In M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, H.-G. Rolff, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Prenzel, M., Baumert, J., Blum, W., Lehmann, R., Leutner, D., Neubrand, M., Pekrun, R., Rolff, H.-G., Rost, J. & Schiefele, U. (Hrsg.) (2004). *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*. Münster: Waxmann.
- Rasch, G. (1960). *Probabilistic models for some intelligence or attainment tests*. Copenhagen: Nielsen & Lydiche (2nd Edition Chicago University of Chicago Press, 1980).
- Rost, J. (1990). Rasch models in latent classes: An integration of two approaches to item analysis. *Applied Psychological Measurement*, 14, 271-282.
- Rost, J. (2004a). *Lehrbuch Testtheorie und Testkonstruktion*. Bern: Huber.

- Rost, J. (2004b). Psychometrische Modelle zur Überprüfung von Bildungsstandards anhand von Kompetenzmodellen. *Zeitschrift für Pädagogik*, 50, 662-679.
- Rupp, A. A. (2002). Feature selection for choosing and assembling measurement models: A building-block-based organization. *International Journal of Testing*, 2, 311-360.
- Rupp, A. A. (im Druck). A guide for describing and investigating the conceptual foundations and statistical properties of cognitive psychometric models. *International Journal of Testing*.
- Rupp, A. A., & Mislevy, R. J. (im Druck). Cognitive foundations of structured item response models. In J. Leighton & M. Gierl (Hrsg.), *Cognitive diagnostic assessment in education: Theory and practice*. Cambridge University Press.
- Skrondal, A., & Rabe-Hesketh, S. (2004). Generalized latent variable modeling: Multilevel, longitudinal, and structural equation models. Boca Raton, FL: Chapman & Hall/CRC.
- Tatsuoka, K. K. (1983). Rule-space: An approach for dealing with misconceptions based on item response theory. *Journal of Educational Measurement*, 20, 345-354.
- Tatsuoka, K. K. (1995). Architecture of knowledge structures and cognitive diagnosis: A statistical pattern recognition and classification approach. In P. D. Nichols, S. F. Chipman & R. L. Brennan (Eds.), *Cognitively diagnostic assessment* (pp. 327-360). Hillsdale, NJ: Erlbaum.
- Templin, J. L. & Henson, R. A. (2005a). Measurement of psychological disorders using cognitive diagnosis models. Manuscript submitted for publication.
- Templin, J. L. & Henson, R. A. (2005b). The random effects reparametrized unified model: A model for joint estimation of discrete skills and continuous ability. Manuscript submitted for publication.
- Templin, J. L., Roussos, L. & Stout, W. (2003). An extension of the current fusion model to treat polytomous attributes. Unpublished Technical Report for the ETS External Diagnostic Group.
- Templin, J. L., He, X., Roussos, L. & Stout, W. (2003). The pseudo-item method: A simple technique for analysis of polytomous data with the fusion model. Unpublished Technical Report for the ETS External Diagnostic Group.
- Ünlü, A. (im Druck). Estimation of careless error and lucky guess probabilities for dichotomous test items: A psychometric application of a biometric latent class model with random effects. *Journal of Mathematical Psychology*.
- von Davier, M. (2005). A general diagnostic model applied to language testing data (Research Report RR-05-16). Princeton, NJ: Educational Testing Service.
- Watermann, R., & Klieme, E. (2002). Reporting results of large-scale assessment in psychologically and educationally meaningful terms: Construct validation and proficiency scaling in TIMSS. *European Journal of Psychological Assessment*, 18, 190-203.
- Weinert, F. E. (2001). Vergleichende Leistungsmessung in Schulen – Eine umstrittene Selbstverständlichkeit. In F.E. Weinert (Hrsg.), *Leistungsmessungen in Schulen* (S. 17-31). Weinheim: Beltz.

- Williamson, D. M., Almond, R. G., Mislevy, R. J., & Levy, R. (2006). An application of Bayesian networks in automated scoring of computerized simulation tasks. In: D. M. Williamson, R. J. Mislevy, & I. I. Bejar (Hrsg.), *Automated scoring of complex tasks in computer-based testing* (pp. 201-258). Mahwah, NJ: Erlbaum.
- Yan, D., Mislevy, R. J. & Almond, R. G. (2003). *Design and analysis in a cognitive assessment* (Research Report No. RR-03-32). Princeton, NJ: Educational Testing Service.
- Zieky, M., & Perie, M. (2006). *A primer on setting cut scores on tests of educational achievement*. Princeton, NJ: Educational Testing Service.

Anschrift der Autoren:

Prof. Dr. André A. Rupp, Institut zur Qualitätsentwicklung im Bildungswesen
Humboldt Universität zu Berlin, Unter den Linden 6, Sitz: Jägerstrasse 10/11,
10099 Berlin, Tel.: 030/20935347, Fax: 030/20935336, Email:
Andre.Rupp@IQB.hu-berlin.de