

Rost, Jürgen

## **Definition von Kompetenzniveaus mit Hilfe von Mischverteilungsmodellen**

*Klieme, Eckhard [Hrsg.]; Beck, Bärbel [Hrsg.]: Sprachliche Kompetenzen. Konzepte und Messung. DESI-Studie (Deutsch Englisch Schülerleistungen International). Weinheim u.a. : Beltz 2007, S. 100-106*

urn:nbn:de:0111-opus-31440

in Kooperation mit:

# **BELTZ**

<http://www.beltz.de>

### **Nutzungsbedingungen**

pedocs gewährt 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 dem Gebrauch von pedocs und der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

#### **Kontakt:**

**peDOCS**

Deutsches Institut für Internationale Pädagogische Forschung (DIPF)

Informationszentrum (IZ) Bildung

Schloßstr. 29, D-60486 Frankfurt am Main

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

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

Bärbel Beck / Eckhard Klieme (Hrsg.)

# **Sprachliche Kompetenzen**

Konzepte und Messung

DESI-Studie

(Deutsch Englisch Schülerleistungen International)

Beltz Verlag · Weinheim und Basel

Dr. *Bärbel Beck* ist Diplompsychologin und Projektkoordinatorin am Deutschen Institut für Internationale Pädagogische Forschung (DIPF) in Frankfurt a.M.

Prof. Dr. *Eckhard Klieme* ist Direktor des Deutschen Instituts für Internationale Pädagogische Forschung (DIPF) in Frankfurt a.M.

Diese Studie wurde im Auftrag der Kultusministerkonferenz erstellt. Für die Richtigkeit des Ergebnisses der Studie trägt das »Deutsches Institut für Internationale Pädagogische Forschung« allein die Verantwortung.

Das Werk und seine Teile sind urheberrechtlich geschützt. Jede Nutzung in anderen als den gesetzlich zugelassenen Fällen bedarf der vorherigen schriftlichen Einwilligung des Verlages. Hinweis zu § 52a UrhG: Weder das Werk noch seine Teile dürfen ohne eine solche Einwilligung eingescannt und in ein Netzwerk eingestellt werden. Dies gilt auch für Intranets von Schulen und sonstigen Bildungseinrichtungen.

Lektorat: Peter E. Kalb

© 2007 Beltz Verlag · Weinheim und Basel

[www.beltz.de](http://www.beltz.de)

Herstellung: Klaus Kaltenberg

Satz: Deutsches Institut für Internationale Pädagogische Forschung

Druck: Druckhaus »Thomas Müntzer«, Bad Langensalza

Printed in Germany

ISBN 978-3-407-25398-9

# Inhaltsverzeichnis

<i>Bärbel Beck / Eckhard Klieme</i> Einleitung.....	1
--	---

## Übergreifende Konzeptualisierung sprachlicher Kompetenzen

<i>Nina Jude / Eckhard Klieme</i> Sprachliche Kompetenz aus Sicht der pädagogisch-psychologischen Diagnostik.....	9
---	---

<i>Günter Nold / Heiner Willenberg</i> Lesefähigkeit .....	23
---	----

<i>Claudia Harsch / Astrid Neumann / Rainer Lehmann / Konrad Schröder</i> Schreibfähigkeit.....	42
--	----

<i>Wolfgang Eichler / Günter Nold</i> Sprachbewusstheit .....	63
--	----

## Messung sprachlicher Kompetenzen

<i>Johannes Hartig</i> Skalierung und Definition von Kompetenzniveaus .....	83
--	----

<i>Jürgen Rost</i> Definition von Kompetenzniveaus mit Hilfe von Mischverteilungsmodellen .....	100
---	-----

## Kompetenzmodelle und Kompetenzniveaus im Bereich des Deutschen

<i>Heiner Willenberg</i> Lesen.....	107
--	-----

<i>Heiner Willenberg / Steffen Gailberger / Michael Krelle</i> Argumentation .....	118
---	-----

<i>Heiner Willenberg</i> Wortschatz.....	130
---	-----

<i>Günther Thomé / Jens Gomolka</i> Rechtschreiben.....	140
--	-----

<i>Wolfgang Eichler</i> Sprachbewusstheit .....	147
--	-----

*Albert Bremerich-Vos / Rüdiger Grotjahn*  
Lesekompetenz und Sprachbewusstheit:  
Anmerkungen zu zwei aktuellen Debatten ..... 158

**Kompetenzmodelle und Kompetenzniveaus im Bereich des Englischen**

*Günter Nold / Henning Rossa*  
Hörverstehen ..... 178

*Günter Nold / Henning Rossa*  
Leseverstehen ..... 197

*Claudia Harsch / Konrad Schröder*  
Textrekonstruktion: C-Test ..... 212

*Günter Nold / Henning Rossa*  
Sprachbewusstheit ..... 226

*Günter Nold / John H. A. L. De Jong*  
Sprechen ..... 245

*Hermann-Günter Hesse / Kerstin Göbel*  
Interkulturelle Kompetenz ..... 256

*Günther Schneider*  
Auf dem Weg zu Skalen für die rezeptiven  
Kompetenzen im Bereich des Englischen ..... 273

**Ausblick**

*Konrad Schröder*  
Kompetenz, Bildungsstandards und Lehrerbildung  
aus fachdidaktischer Sicht ..... 290

*Günter Nold*  
DESI im Kontext des Gemeinsamen  
Europäischen Referenzrahmens für Sprachen ..... 299

*Sauli Takala*  
Relating Examinations to the Common European Framework ..... 306

*Hermann Lange*  
Abschließendes Statement ..... 314

Die Autorinnen und Autoren ..... 318

Jürgen Rost

## Definition von Kompetenzniveaus mit Hilfe von Mischverteilungsmodellen

Das Konzept von Kompetenzniveaus ist aus der Intention heraus entstanden, im Rahmen der Erhebung von Bildungsstandards die Ergebnisse von Leistungstests kriteriumsorientiert (statt normorientiert) zu erfassen und zu interpretieren. Normorientierte Aussagen sind bei der Erfassung von Bildungsstandards sinnvoll, wenn es z.B. darum gehen soll, Bundesländer oder Schularten miteinander zu vergleichen. Primäre Aufgabe der Erhebung von Bildungsstandards ist es jedoch, die Leistung des Schülers an einem gesetzten qualitativen Kriterium der Leistungsfähigkeit zu messen. Das Konzept von Kompetenzniveaus stellt einen Versuch dar, quantitative Unterschiede in der Leistungsfähigkeit zugleich auch als qualitative Unterschiede hinsichtlich der kognitiven Ausstattung der Schüler zu erfassen.

Psychometrisch betrachtet stellt diese Konzeption die Anforderung, Leistungsunterschiede bei Schülern gleichzeitig quantitativ und qualitativ abzubilden. Die Item-Response-Theorie bietet hinreichend viele und elaborierte Modelle quantitative Leistungsunterschiede mit Tests zu erfassen. Klassifizierende Testmodelle auf der Basis der Latent-Class-Analyse (Lazarsfeld/Henry 1968; Rost 2004) bieten ebenso vielfältige Möglichkeiten, qualitative Unterschiede in den Schülerleistungen in Form von Klassen oder Typen von Schülern abzubilden. Die Schüler eines Typs zeichnen sich durch ein gemeinsames Profil der Aufgabenlösungswahrscheinlichkeiten aus.

Sollen qualitative und quantitative Unterschiede in den Schülerleistungen *gleichzeitig* erfasst werden – wie es das Konzept von Kompetenzniveaus vorsieht – so müssen entweder qualitative Elemente in den quantitativen psychometrischen Ansatz übertragen werden oder quantitative Elemente in einen qualitativen (klassifizierenden) Ansatz. Der vorangehende Beitrag (Hartig in diesem Band) stellt eine Methode dar, mit Hilfe von Aufgabenmerkmalen Segmente auf der quantitativen Leistungsdimension zu definieren. Diese zeichnen sich durch die kognitiven Leistungen der Schüler aus, die sich auf dem entsprechenden Niveau befinden. Der Ansatz, über Aufgabenmerkmale Abschnitte auf der latenten Dimension zu definieren, kann als ein Versuch gewertet werden, qualitative Interpretationskategorien in das eigentlich quantitative Messmodell von Rasch zu integrieren.

Der vorliegende Beitrag hingegen stellt einen Weg dar, quantitative Leistungsunterschiede in einen genuinen klassifikatorischen Ansatz zu implementieren. Hiermit ist der Ansatz gemeint, Klassen von Schülern oder Schülertypen zu identifizieren, die sich jeweils dadurch auszeichnen, dass die Leistungen der Schüler dieser Klasse (mittels des Rasch-Modells) eindimensional skalierbar sind. Die primäre Variable ist also die klassifizierende latente Variable, die unter der Bedingung kons-

truiert wird, dass innerhalb jeder Klasse eine quantitative Variable messbar ist. Das Modell, das eine solche Klassifikation vornimmt, ist das Mixed-Rasch-Modell oder Misch-Verteilungs-Rasch-Modell (Rost 1990, 1991).

Die Definition von Kompetenzniveaus mit Hilfe des Mixed-Rasch-Modells wird im Folgenden anhand eines Beispiels vorgestellt. Dieses Beispiel stammt aus den Daten des DESI-Projekts, genauer gesagt aus dem Multiple-Choice-Fragebogen im Anschluss an eine Hörverstehensaufgabe. In dieser Aufgabe erhielten die Schüler einen englischsprachigen Text aus einer Tonaufzeichnung vorgespielt und sollten im Anschluss daran acht Fragen zum Inhalt dieses Textes beantworten. Es gab jeweils drei Antwortalternativen, von denen eine richtig war, die mit 1 kodiert wurde, während den falschen Antworten eine 0 zugewiesen wurde.

Der Wortlaut des gehörten Textes sowie die genaue Formulierung der dazu gestellten Fragen kann aus Gründen der Geheimhaltung an dieser Stelle nicht wiedergegeben werden. Analoge Beispiele findet der Leser im Kapitel Hörverstehen in diesem Band.

Im Rahmen eines rein quantitativen Vorgehens würde man zunächst die Frage prüfen, inwieweit die acht Items dieser Hörverstehensaufgabe eindimensional sind, d.h. inwieweit sie eine eindimensionale Personenvariable messen. Eine entsprechende Hauptkomponentenanalyse weist den Test zwar als eindimensional aus (nur ein Eigenwert ist größer als eins) jedoch zeigt das Ladungsmuster der Zweikomponenten-Lösung, wie in Abbildung 1 dargestellt, dass die Items vier und sechs auf einer anderen Dimension liegen als die übrigen sechs Items.

Während für Item vier kein Grund ersichtlich ist, warum es sich von den übrigen sechs Items unterscheidet, stellt Item sechs die einzige Frage zum Text dar, in der es um die Gefühle einer Person aus dem gesprochenen Text geht. Zu der Frage „Wie fühlte sich der Mann?“ hatten die Schüler die richtige Lösung aus folgenden drei Alternativen auszuwählen: Worried, Sad und Disappointed. Während die anderen Items nach bloßen Fakten fragten, wie z.B. die Frage danach, in welchem Restaurant die beiden essen gehen wollten, muss bei Item sechs nicht nur das richtige Gefühl erschlossen, sondern auch noch das korrekte englische Wort dafür gewählt werden. Es wird sich im Folgenden zeigen, inwieweit diese beiden Aufgaben auch bei der primär klassifizierenden Methode des Mixed-Rasch-Modells auffällig sind.

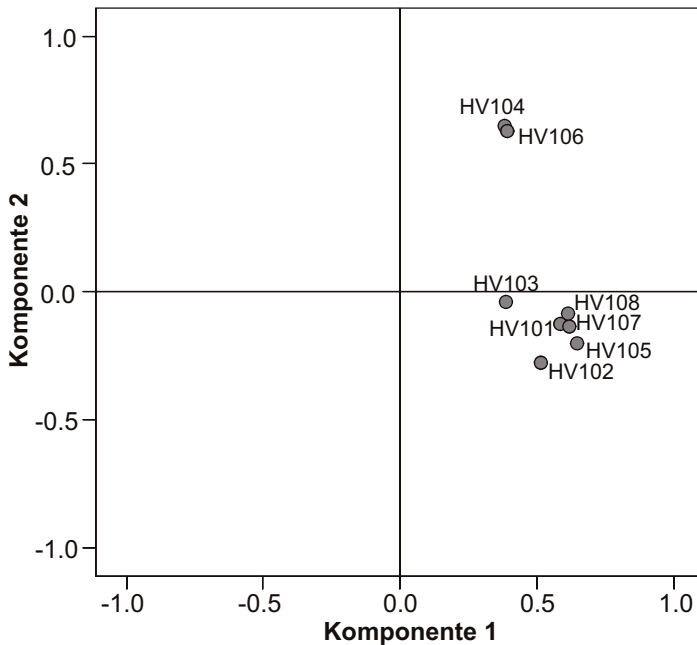


Abbildung 1: Ladungsmuster der acht Items auf den ersten beiden Hauptkomponenten.

Berechnet man zunächst die Zwei-Klassen-Lösung des Mixed-Rasch-Modells, so zeigen sich zwei Klassen, die auf einem deutlich unterschiedlichen Leistungsniveau liegen. Abbildung 2 gibt die beiden Verteilungen der Testpunktwerte wieder.

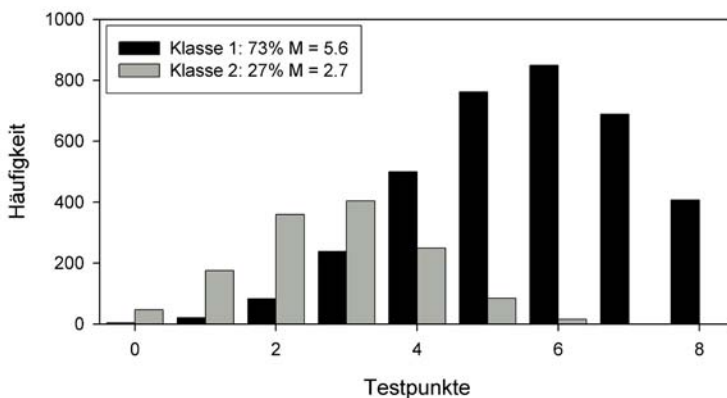


Abbildung 2: Die Verteilungen der Punktwerte in den beiden Klassen der Zwei-Klassen-Lösung des Mixed-Rasch-Modells.

Während die erste, größere Klasse im Durchschnitt sechs bis sieben Items richtig beantwortet hat, sind es bei der zweiten Klasse nur drei bis vier. Dennoch überlappen



sich beide Verteilungen, was bedeutet, dass die Zwei-Klassen-Lösung keine Teilung der Stichprobe an einem Cut-Off-Score beinhaltet, sondern eine leistungsstarke und eine leistungsschwache Klasse widerspiegelt, innerhalb derer eine beträchtliche quantitative Variation liegt. Die folgende Abbildung (Abb. 3) zeigt die Lösungswahrscheinlichkeiten in den beiden Klassen, wobei das obere Profil der ersten, größeren Klasse zuzuordnen ist und das untere der zweiten, kleineren Klasse.

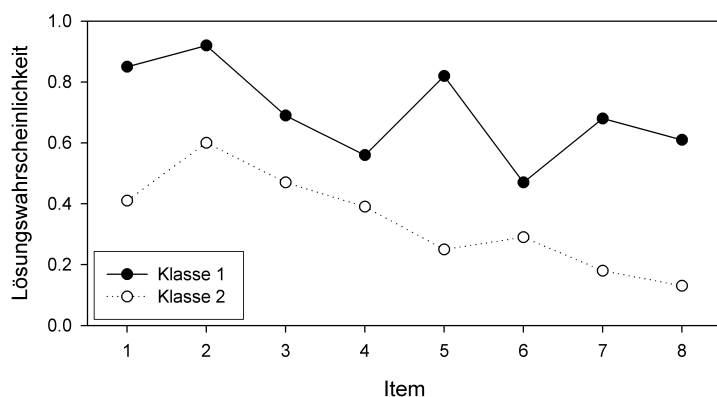


Abbildung 3: Die Itemprofile der Antwortwahrscheinlichkeiten in den beiden latenten Klassen der Zwei-Klassen-Lösung des Mixed Rasch-Modells.

Wie nach der zweidimensionalen Lösung der Hauptkomponentenanalyse (s.o.) erwartet, erweisen sich das vierte und das sechste Item als sehr trennschwach, d.h. die Profile haben an diesen Stellen einen sehr geringen Abstand. Das trifft allerdings auch auf das dritte Item zu, das in der zweidimensionalen Struktur nicht auffällig war.

Statistische Gütekriterien für die verschiedenen Klassenanzahlen des Mixed-Rasch-Modells zeigen an, dass die Vier-Klassen-Lösung sich noch besser eignet als die Zwei-Klassen-Lösung. Im Folgenden werden die eben berichteten Informationen übertragen auf die Vier-Klassen-Lösung wiedergegeben. Dass es sich bei der Vier-Klassen-Lösung tatsächlich um eine Aufspaltung der beiden Klassen in der Zwei-Klassen-Lösung handelt, belegt die folgende Kreuztabelle (vgl. Tabelle 1).

Tabelle 1: Die prozentualen Häufigkeiten von Schülern, die den Klassen eins und zwei der Zwei-Klassen-Lösung und den Klassen eins bis vier der Vier-Klassen-Lösung angehören.

		Zwei-Klassen-Lösung	
		Klasse 1	Klasse 2
		%	
Vier-Klassen-Lösung	Klasse 1	46	0
	Klasse 2	21	1
	Klasse 3	0	25
	Klasse 4	4	3

Lediglich die kleine Klasse vier mit sieben Prozent der befragten Schüler rekrutiert sich aus beiden Leistungsgruppen der Zwei-Klassen-Lösung. Die Score-Verteilungen

der beiden ersten Klassen (siehe Abb. 4) liegen nahezu übereinander, d.h. in beiden Klassen wurden etwa sechs der acht Items richtig beantwortet.

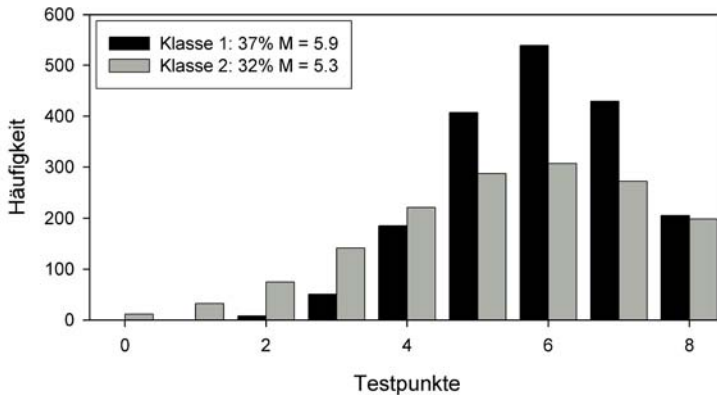


Abbildung 4: Die Verteilungen der Punktwerte in den beiden ersten Klassen der Vier-Klassen-Lösung.

Betrachtet man die in Abbildung 5 dargestellten Lösungsprofile in den beiden ersten Klassen, so zeigen sich weitgehend überlappende Profilverläufe. Einzige Ausnahme scheint das Item drei zu sein, welches in der kleineren der beiden leistungsstarken Klassen sehr viel schwieriger zu lösen ist als in der größeren.

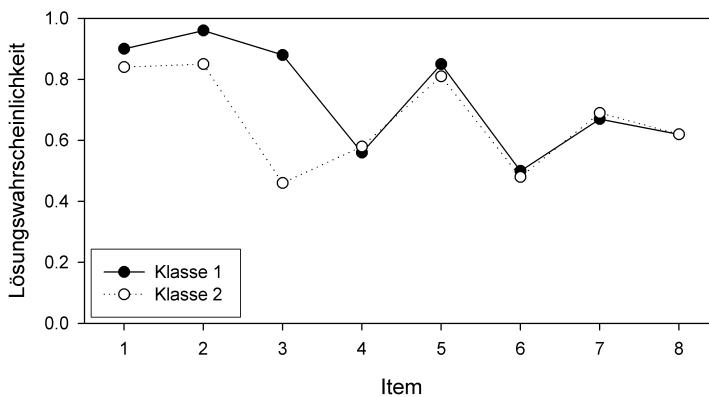


Abbildung 5: die Itemprofile der beiden ersten Klassen der Vier-Klassen-Lösung.

Es handelt sich bei Item drei um die Frage nach dem Hauptgrund für die Japanreise des Mannes aus der Erzählung. Item vier und sechs verhalten sich in Bezug auf die Ähnlichkeit der Profile beider Klassen unauffällig.

Die beiden verbleibenden Klassen sind erwartungsgemäß leistungsschwächer, wobei Klasse drei mit einem Erwartungswert von zwei bis drei gelösten Aufgaben noch geringere Leistungen aufweist als die vierte Klasse mit einem Erwartungswert von vier gelösten Aufgaben (siehe Abb. 6).

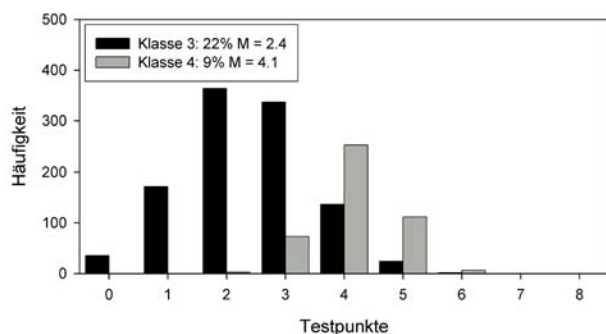


Abbildung 6: Die Verteilungen der Punktwerte in der dritten und vierten Klasse der Vier-Klassen-Lösung.

Die in Abbildung 7 dargestellten Lösungsprofile dieser beiden Klassen zeigen nun endlich ein Ergebnis, das Item sechs betrifft: Es ist das einzige der acht Items, das in beiden Klassen gleich schwierig zu beantworten ist. Anders herum ausgedrückt sind die Schüler der Klasse vier, die anscheinend auf dem Weg zu einem höheren Leistungsniveau sind, nicht in der Lage, die Frage sechs besser zu beantworten als dies in der leistungsschwächsten Gruppe geschieht.

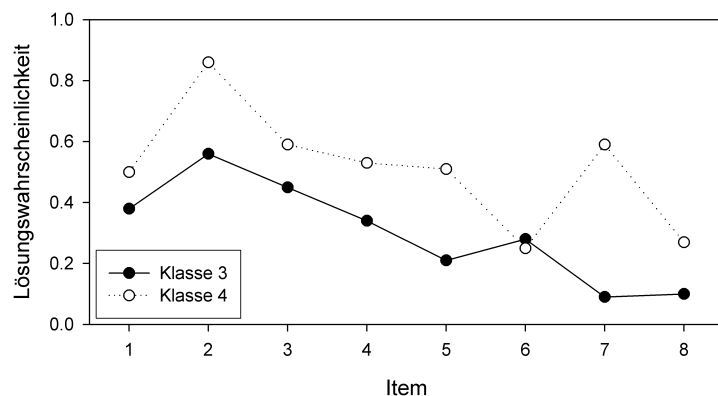


Abbildung 7: Die Itemprofile der dritten und vierten Klasse der Vier-Klassen-Lösung.

Man könnte dieses Ergebnis dahingehend interpretieren, dass die korrekte Wiedergabe von Gefühlen aus dem erzählten Text für Schüler auf einem mittleren Leistungsniveau schwieriger ist und in diesem Sinne eine höhere oder andersartige Kompetenz erfordert als die Beantwortung der übrigen Items. Daraus ließe sich schließen, dass im Englischunterricht die Kompetenz, Gefühlszustände zu erkennen, zu erinnern und zu benennen eine Hürde im unteren Leistungsbereich darstellen kann, die besondere Aufmerksamkeit erfordert.

Dieses Beispiel ist inhaltlich vielleicht nicht so aussagekräftig, es zeigt jedoch, was mit der gleichzeitigen Klassifikation und Quantifizierung von Leistungsdaten ge-

meint ist. Die vier Klassen wurden nicht a priori definiert, sondern vom Algorithmus anhand des Kriteriums gefunden, dass die Schüler innerhalb jeder Klasse mittels des Rasch-Modells skalierbar sein sollen. Die vom Modell identifizierten vier Klassen stellen keine Segmente auf dem latenten Kontinuum dar, wie es im vorangehenden Kapitel (Hartig in diesem Band) für die Definition von Kompetenzniveaus anhand von Aufgabenmerkmalen dargestellt wurde. Die Leistungswerte insbesondere der Klassen eins und zwei überlappen sich stark, so dass das *Antwortprofil*, also das *Lösungsmuster* der Schüler ausschlaggebend für die Klassenzugehörigkeit der Schüler ist. Generell erlauben die Lösungsprofile in einer Klasse eine qualitative Diagnose der Stärken und Schwächen der Schüler. Demgegenüber spiegeln die Verteilungen der Punktwerte die Breite des Leistungsspektrums wieder, innerhalb dessen Schüler mit einem derartigen Antwortprofil zu finden sind.

Die Methode des Mixed-Rasch-Modells zur Identifizierung von Kompetenzausprägungen und Kompetenzprofilen unterscheidet sich in wesentlichen Merkmalen vom Ansatz der Definition von Kompetenzniveaus anhand von Segmenten auf der Leistungsskala. Inwieweit es überschneidungsfreie Segmente mit einem unterschiedlichen qualitativen Antwortprofil gibt, wird empirisch untersucht und nicht a priori postuliert. Auch bedarf es nicht unbedingt starker Hypothesen über die qualitativen Kompetenzunterschiede auf unterschiedlichen Leistungsniveaus. Stattdessen liefert das Modell die Profile der latenten Klassen, wie sie anhand der Daten geschätzt worden sind.

Das primäre Ziel einer Analyse mit Hilfe des Mixed-Rasch-Modells besteht darin, eine gute Klassifikation zu finden, während die Definition von Kompetenzniveaus anhand von Aufgabenmerkmalen von der Eindimensionalität der Testleistung ausgeht und an dieser Annahme auch nicht rüttelt. Während das eine Verfahren nach Skalierbarkeit innerhalb der identifizierten Klassen sucht, strebt das andere Verfahren eine Klassifikation auf der gemessenen Variable an. Insofern sind beide methodischen Zugänge hinsichtlich der Priorität von qualitativer und quantitativer Analyse komplementär und es bleibt letztlich eine Frage der Testkonstruktion, des Testdesigns und der empirischen Ergebnisse, welches Modell für ein gegebenes Kompetenzkonstrukt die adäquate Kompetenzmessung liefert.

## Literatur

Lazarsfeld, P.F./Henry, N.W. (1968): Latent structure analysis. Boston: Houghton Mifflin Co.

Rost, J. (1990): Rasch Models in Latent Classes: An Integration of two approaches to item analysis. In: Applied Psychological Measurement 14, S. 271-282.

Rost, J. (1991): A Logistic Mixture Distribution Model for Polychotomous Item Responses. In: The British Journal for Mathematical and Statistical Psychology 44, S. 75-92.

Rost, J. (2004): Lehrbuch Testtheorie – Testkonstruktion. Bern: Huber.