Kompetenzmodellierung: Struktur, Konzepte und Forschungszugänge des DFG-Schwerpunktprogramms

**Jens Fleischer · Karoline Koeppen · Martina Kenk ·
Eckhard Klieme · Detlev Leutner**

Zusammenfassung: Die Messung und Modellierung von Kompetenzen stellt eine zentrale Voraussetzung für die Optimierung von Bildungsprozessen sowie für die Qualitätssicherung und Qualitätsentwicklung im Bildungswesen dar. Das 2007 eingerichtete DFG-Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ verfolgt das Ziel, die wissenschaftlichen Anstrengungen in diesem Bereich – über Fachdisziplinen hinweg – voranzutreiben und zu koordinieren. Der vorliegende Beitrag beschreibt die Struktur, Konzepte und Forschungszugänge und skizziert den aktuellen Forschungsstand des DFG-Schwerpunktprogramms.

Schlüsselwörter: Kompetenz · Kompetenzmodellierung · Kompetenzdiagnostik · DFG-Schwerpunktprogramm

Modeling of competencies: structure, concepts and research approaches of the DFG priority program

Abstract: The assessment and modeling of competencies plays a key role in optimizing educational processes and improving educational systems. The DFG priority program “Competence

© Springer Fachmedien Wiesbaden 2013

Dipl.-Psych. J. Fleischer (✉) · Prof. Dr. D. Leutner
Institut für Psychologie, Universität Duisburg-Essen,
45117 Essen, Deutschland
E-Mail: jens.fleischer@uni-due.de

Prof. Dr. D. Leutner
E-Mail: detlev.leutner@uni-due.de

Dr. K. Koeppen · Dipl.-Päd. M. Kenk · Prof. Dr. E. Klieme
Bildungsqualität und Evaluation, Deutsches Institut für Internationale Pädagogische Forschung,
Schloßstr. 29, 60486 Frankfurt am Main, Deutschland
E-Mail: koeppen@dipf.de

Dipl.-Päd. M. Kenk
E-Mail: kenk@dipf.de

Prof. Dr. E. Klieme
E-Mail: klieme@dipf.de

Prof. Dr. D. Leutner
E-Mail: detlev.leutner@uni-due.de

Models for Assessing Individual Learning Outcomes and Evaluating Educational Processes” which was founded in 2007, aims at promoting and coordinating the scientific efforts in this field across disciplines. The present article describes the structure, concepts and research approaches and sketches the current state of research of the priority program.

Keywords: Competence · Competence assessment · Competence modeling · DFG priority program

Die Vermittlung relevanter Wissensstände und Fertigkeiten in konkreten Anforderungsbereichen, die häufig unter dem Begriff „Kompetenz“ gefasst werden, ist ein zentrales Anliegen schulischer und beruflicher Bildung (vgl. Köller 2009). Denn Prosperität, soziale Kohäsion und Entwicklungschancen einer Gesellschaft hängen in großem Maße von den erworbenen Kompetenzen ihrer Mitglieder ab. Der Messung von Kompetenzen kommt damit eine Schlüsselfunktion für die Optimierung von Bildungsprozessen sowie die Qualitätssicherung und Qualitätsentwicklung des Bildungswesens zu (Klieme und Leutner 2006; Koeppen et al. 2008). Von der Bildungsforschung wird in diesem Zusammenhang erwartet, Modelle der Struktur, Stufung und Entwicklung von Kompetenzen bereitzustellen, die sowohl kognitionspsychologisch als auch fachlich fundiert sind und als Grundlage zur Entwicklung geeigneter Messinstrumente zur Erfassung von Kompetenzen dienen können.

Um diesem Anspruch gerecht werden zu können und die wissenschaftlichen Anstrengungen zur Bewältigung dieser Herausforderungen über Fachdisziplinen hinweg voranzutreiben und zu koordinieren, wurde 2007 das DFG-Schwerpunktprogramm (SPP) „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ eingerichtet (Klieme und Leutner 2006). Das Schwerpunktprogramm, welches 2012 in seine dritte und letzte zweijährige Förderphase gegangen ist, umfasst 30 Projekte, die sich über Fachdisziplinen hinweg mit verschiedenen Aspekten der Modellierung und Messung von Kompetenzen in den Domänen Mathematik, Naturwissenschaft, Sprache/Lesen, Lehrerkompetenzen und fächerübergreifende Kompetenzen befassen.

Zentraler Bezugspunkt für die Arbeiten des Schwerpunktprogramms ist der Kompetenzbegriff, der – obwohl keineswegs neu – erst seit einem Jahrzehnt zum Gegenstand intensiver Diskussionen in der Psychologie und der Bildungsforschung geworden ist (vgl. z. B. Csapó 2004; Hartig und Klieme 2006; Klieme et al. 2008). Die vielfältige Bedeutung des Kompetenzbegriffs hat Weinert (2001) systematisch aufbereitet. Er unterscheidet dabei sechs Verwendungen des Begriffs. So werden Kompetenzen unter anderem als auf spezifische Kontexte bezogene kognitive Leistungsdispositionen, aber beispielsweise auch als für die Bewältigung anspruchsvoller Aufgaben notwendige motivationale Orientierungen verstanden (Weinert 2001; vgl. auch Klieme 2004). Aufgrund theoretischer und pragmatischer Überlegungen empfiehlt Weinert (2001) eine Einschränkung auf die erstgenannte Definition. Dieser Empfehlung folgend werden im Rahmen des Schwerpunktprogramms Kompetenzen definiert als „kontextspezifische kognitive Leistungsdispositionen, die sich funktional auf Situationen und Anforderungen in bestimmten Domänen beziehen“ (Klieme und Leutner 2006, S. 879) und die durch Erfahrung und Lernen sowie äußere Interventionen und institutionalisierte Bildungsprozesse erworben und beeinflusst werden (vgl. Klieme und Leutner 2006; Klieme et al. 2008). Mit dieser Definition sind zwei Einschränkungen verbunden: Zum einen sind Kompetenzen bereichsspezifisch auf bestimmte

Kontexte beziehungsweise Anforderungssituationen bezogen, wodurch allgemeine kognitive Fähigkeiten ausgeschlossen werden. Kompetenzen werden damit, in einem größeren Maße als dies für allgemeine kognitive Fähigkeiten der Fall ist, als prinzipiell erlernbar und trainierbar verstanden. Zum anderen erfolgt eine Beschränkung auf kognitive Aspekte, wodurch motivationale und emotionale Faktoren zunächst ausgeschlossen werden. Diese Einschränkungen erfolgen im Rahmen des Schwerpunktprogramms vor allem auch aus pragmatischen Erwägungen und sollen nicht die prinzipielle Bedeutung allgemeiner kognitiver Fähigkeiten wie zum Beispiel der Intelligenz oder motivationaler Einstellungen und Orientierungen sowie emotionaler Faktoren für erfolgreiches Handeln in spezifischen Anforderungssituationen in Abrede stellen (vgl. Wilhelm und Nickolaus 2013 in diesem Heft). Die im Rahmen des Schwerpunktprogramms verwendete Definition des Kompetenzbegriffs steht damit in wesentlichen Punkten in Übereinstimmung mit der Verwendung des Kompetenzbegriffs im Kontext internationaler Schulleistungsstudien (z. B. PISA, TIMSS, IGLU/PIRLS) sowie der KMK-Bildungsstandards (vgl. Pant et al. 2012; siehe auch Pant 2013 in diesem Heft). So werden Kompetenzen beispielsweise in den PISA-Studien als „prinzipiell erlernbare, mehr oder minder bereichsspezifische Kenntnisse, Fertigkeiten und Strategien“ definiert (Baumert et al. 2001, S. 22).

Basierend auf dieser Arbeitsdefinition des Kompetenzbegriffs lassen sich die zentralen Fragestellungen des Schwerpunktprogramms insgesamt vier aufeinander aufbauenden Forschungsbereichen zuordnen (vgl. Klieme und Leutner 2006; Klieme et al. 2008; Koeppen et al. 2008). Die Abb. 1 stellt diese Bereiche in Form eines Zwiebelmodells dar: Im Kern befindet sich die Entwicklung theoretischer Kompetenzmodelle (Bereich 1). Aufbauend auf diesen theoretischen Modellen werden adäquate psychomet-

Abb. 1: Forschungsbereiche und Struktur des Schwerpunktprogramms

rische Modelle entwickelt (Bereich 2), die wiederum die Grundlage für die Konstruktion von konkreten Messinstrumenten zur Erfassung von Kompetenzen darstellen (Bereich 3). Abgerundet wird das Forschungsprogramm durch Arbeiten, die sich mit Fragen der Nutzung der gewonnenen diagnostischen Informationen befassen (Bereich 4).

Im Folgenden werden diese Forschungsbereiche im Überblick beschrieben sowie der aktuelle Forschungsstand des DFG-Schwerpunktprogramms skizziert.

1 Forschungsbereich: Theoretische Kompetenzmodelle

Zentrale Fragestellung: Wie lassen sich Kompetenzen, unter Berücksichtigung ihres Bezugs auf Anforderungen in spezifischen Situationen, angemessen kognitiv modellieren?

Bei der Formulierung von theoretischen Modellen zur Beschreibung und Erklärung von Kompetenzen lassen sich *Kompetenzstrukturmodelle* und *Kompetenzniveaumodelle* unterscheiden (Klieme und Hartig 2007):

- Kompetenzstrukturmodelle befassen sich mit der Frage, welche und wie viele verschiedene Teilkompetenzen (Dimensionen) einer bestimmten Kompetenz differenziert werden können. Diese Dimensionen repräsentieren inhaltlich unterscheidbare Aspekte, anhand derer sich Kompetenzunterschiede zwischen Personen beschreiben lassen. Die Binnenstruktur von Kompetenzen ist dabei vornehmlich durch die in einem bestimmten Bereich zu bewältigenden Anforderungen und die hierfür als notwendig erachteten kognitiven Prozesse bedingt (vgl. Hartig und Klieme 2006). So können die Dimensionen einer Kompetenz beispielsweise durch unterschiedliche inhaltliche beziehungsweise curriculare Einheiten (vgl. z. B. Winkelmann et al. 2012), unterschiedliche Arten von Aufgaben- und Problemstellungen (vgl. z. B. Leutner et al. 2012), unterschiedliche Wissensarten (vgl. z. B. Fleischer et al. 2010; Hardy et al. 2010), unterschiedliche Repräsentations- und Aufgabenformate (vgl. z. B. Bayrhuber et al. 2010; Schnotz et al. 2010) oder unterschiedliche Verarbeitungsprozesse (vgl. z. B. Hartig und Jude 2008) definiert sein. Die Beschreibung von interindividuellen Unterschieden auf diesen Dimensionen erfolgt dann in der Regel auf der Grundlage von psychometrischen Modellen mit kontinuierlichen latenten Variablen (vgl. Hartig und Frey 2013 in diesem Heft).
- Kompetenzniveaumodelle dienen einer qualitativen, kriteriumsorientierten Beschreibung der Anforderungen, die Personen mit unterschiedlich stark ausgeprägten (Teil-)Kompetenzen bewältigen können. Zur Definition von Kompetenzniveaus wird eine an sich kontinuierliche Kompetenzdimension in diskrete, ordinale Kategorien unterteilt. Für diese Kategorien erfolgt im Anschluss eine kriteriumsorientierte Beschreibung der erfassten Kompetenzen (vgl. Hartig und Klieme 2006; Wilson 2005). Die Definition von Kompetenzniveaus stellt letztlich immer eine Vereinfachung und Informationsreduktion dar mit dem Ziel einer besseren Beschreibung und Kommunizierbarkeit der erfassten Kompetenzen (vgl. Leuders und Sodian 2013 in diesem Heft). Zur Unterteilung der kontinuierlichen Kompetenzdimensionen in Niveaus gibt es unterschiedliche Vorgehensweisen (für einen Überblick siehe z. B. Cizek und Bunch 2007; Pant et al. 2010; vgl. auch Hartig 2007). Obwohl Kompetenzniveaus

zumeist nicht echten Entwicklungsstufen im Sinne qualitativer Entwicklungssprünge entsprechen, wird häufig synonym auch der Begriff „Kompetenzstufe“ als Übersetzung des englischen „proficiency level“ verwendet (vgl. Helmke und Hosenfeld 2004).

Die Frage der *zeitlichen Entwicklung von Kompetenzen* ist ein weiterer wichtiger Aspekt der Modellierung von Kompetenzen, da sich weder aus Kompetenzstruktur- noch aus Kompetenzniveaumodellen direkt *Entwicklungsmodelle* von Kompetenzen ableiten lassen (vgl. Robitzsch 2013 in diesem Heft). So gibt es bisher kaum hinreichend theoretisch fundierte und empirisch validierte Kompetenzentwicklungsmodelle (Artelt und Schneider 2011; Klieme et al. 2008). Dementsprechend bleibt bislang weitgehend unklar, ob die zeitliche Entwicklung von Kompetenzen eher als kontinuierlicher Prozess verstanden werden kann (vgl. z. B. Blum et al. 2004; PISA-Konsortium Deutschland 2006) oder ob Kompetenzentwicklung vielmehr ein diskontinuierlicher Prozess ist, bei dem qualitativ unterschiedliche Stufen durchlaufen und Konzepte und Kompetenzstrukturen reorganisiert werden (vgl. z. B. Vosniadou 2008; siehe auch Leuders und Sodian 2013 in diesem Heft).

Kompetenzmodelle im hier beschriebenen Sinne integrieren quantitative und qualitative Beschreibungen von inter- und (bei der Beschreibung von Kompetenzentwicklungen) auch intra-individuellen Unterschieden. Das typische Kompetenzmodell strukturiert eine Kompetenz nach Kompetenzdimensionen und es graduert jede dieser Dimensionen nach Niveaus. Kompetenzmodelle müssen darüber hinaus bezüglich ihres Abstraktionsgrades hinreichend spezifisch sein, um als Grundlage für die modellbasierte Entwicklung von Messinstrumenten zur Erfassung von Kompetenzen genutzt werden zu können (vgl. Neumann 2013 in diesem Heft).

Hinsichtlich der theoriegeleiteten Modellierung von Kompetenzen sind in den letzten Jahren große Fortschritte zu verzeichnen. So beschreibt Wilson (2005) einen allgemeinen Rahmen für eine modell- und theoriegeleitete Vorgehensweise bei der Definition von Kompetenzmodellen und der darauf aufbauenden Konstruktion von Testaufgaben. Beispiele für die theoriegeleitete Definition von Kompetenzmodellen finden sich darüber hinaus im Rahmen von Large-Scale-Assessments. So beschreibt Hartig (2007) ein entsprechendes Vorgehen zur Definition von Kompetenzniveaus im Rahmen der DESI-Studie. Da die zur Definition von Kompetenzniveaus durchgeführte kriteriumsorientierte Interpretation von Testwerten auf den empirischen Schwierigkeiten der eingesetzten Testaufgaben beruht, kommt der Vorhersage dieser Schwierigkeiten eine besondere Bedeutung für die modellbasierte Fundierung der Testwertinterpretation zu (vgl. z. B. Neubrand et al. 2002; Watermann und Klieme 2002). Darüber hinaus kann die Vorhersage von Aufgabenschwierigkeiten einen wichtigen Beitrag zur Konstruktvalidierung im Sinne der Konstruktrepräsentation sensu Embretson (1983) liefern sowie zur Optimierung der Entwicklung zukünftiger Testaufgaben beitragen (Hartig und Frey 2012; Leucht et al. 2012; Schaper et al. 2008).

Im Zusammenhang mit der theoriegeleiteten Modellierung von Kompetenzen im Rahmen von Large-Scale-Assessments sind ebenfalls als Beispiel Arbeiten zur Lesekompetenz im Rahmen der PISA-Studie 2000 (Artelt et al. 2001) zu nennen sowie zur mathematischen Kompetenz im Rahmen von PISA-2003 (Blum et al. 2004) oder zur naturwissenschaftlichen Kompetenz im Rahmen von PISA-2006 (Prenzel et al. 2007). In diesen Arbeiten wurden unter Bezug auf einschlägige Theorien in den jeweiligen fachli-

chen Domänen Kompetenzen beschrieben und modelliert. Allerdings erfolgte die Zuordnung von Testaufgaben zu Kompetenzniveaus in der Regel post-hoc und die Modelle besitzen häufig noch einen sehr hohen Abstraktionsgrad, sodass sie für eine modellgeleitete Itemkonstruktion nur bedingt geeignet sind. Auch die in jüngerer Zeit durchgeführten Studien zur Evaluation und Normierung der nationalen Bildungsstandards in Deutschland liefern Beispiele für ein stärker theorie- und modellgeleitetes Vorgehen bei der Erfassung von Kompetenzen (vgl. z. B. Harsch et al. 2010). So liegen aus diesen Arbeiten inzwischen unter anderem Kompetenzniveau Modelle sowohl für die globale Mathematikkompetenz als auch für die verschiedenen Inhaltsbereiche der Mathematik, die Lesekompetenz in Deutsch und für Lese- und Hörverstehen in der ersten Fremdsprache vor (Bremerich-Vos und Böhme 2009; Reiss und Winkelmann 2009; Tesch et al. 2010).

Ebenso können unter reger Beteiligung der Projekte des Schwerpunktprogramms für die Bereiche mathematisch-naturwissenschaftliche Kompetenzen, fächerübergreifende Kompetenzen, sprachlich-kulturelle Kompetenzen, berufliche Kompetenzen sowie Lehrerkompetenzen große Fortschritte bei der Definition von Kompetenzmodellen verzeichnet werden, die als Grundlage zur Entwicklung von Testinstrumenten genutzt werden können (vgl. z. B. Blomberg et al. 2011; Duit und Möller 2010; Fleischer et al. 2012; Hartig et al. 2008; Klieme et al. 2010a; Leuders und Biehler *im Druck*). Hierbei werden zunehmend auch Entwicklungsaspekte von Kompetenzen untersucht (vgl. z. B. Artelt et al. 2012; Eggert et al. 2010; Hülür et al. 2011; Koerber et al. 2011; Robitzsch et al. 2011; Viering et al. 2010).

Empirische Daten zur zeitlichen Entwicklung von Kompetenzen sowie zu tragfähigen Kompetenzentwicklungsmodellen sind darüber hinaus in den kommenden Jahren auch im Zusammenhang mit den Arbeiten des Nationalen Bildungspanel (NEPS; Blossfeld et al. 2011) sowie aus den Projekten des DFG-Schwerpunktprogramms „Education as a Lifelong Process. Analyzing Data of the National Educational Panel Study (NEPS)“ (<https://spp1646.neps-data.de/>) zu erwarten.

2 Forschungsbereich: Psychometrische Modelle

Zentrale Fragestellung: Wie lassen sich theoretische Kompetenzmodelle in psychometrischen Modellen abbilden, um die Kompetenzkonstrukte einer differenziellen Erfassung zugänglich zu machen?

Psychometrische Testmodelle stellen die Verbindung zwischen theoretischem Kompetenzmodell und dem konkreten Testverhalten (einzelne Itemantworten) einer Person bei der Bearbeitung eines Kompetenztests dar. Sie bieten ein Regelsystem, welches es erlaubt, aus Unterschieden im Testverhalten auf interindividuelle Unterschiede in der Ausprägung spezifischer Kompetenzen zu schließen (vgl. Embretson 1983; Wilson 2003). Eine zentrale Frage ist hierbei, auf welche Weise das Verhältnis zwischen situativen Anforderungen einerseits und Personenmerkmalen andererseits modelliert werden muss, damit individuelle Testwerte als Fähigkeit zur Bewältigung spezifischer Anforderungen im Sinne des Kompetenzbegriffs interpretiert werden können (Hartig 2008). Hierfür bieten sich insbesondere Modelle der Item-Response-Theorie (IRT, vgl. z. B. Rost 2004) an, da sie gewissermaßen die Rekonstruktion von Kompetenzen aus der beobachteten Perfor-

manz ermöglichen (vgl. Hartig und Frey 2013 in diesem Heft). Ein besonderer Vorteil ist, dass individuelle Messwerte von Personen (Personenparameter) mit den Schwierigkeiten von Aufgaben (Itemparameter) verglichen werden können (vgl. Embretson und Reise 2000), was eine kriteriumsorientierte Interpretation der Testergebnisse ermöglicht. Bei der Erfassung von kontextualisierten Konstrukten wie Kompetenzen ist dies besonders wichtig, da es den Bezug individueller Kompetenzausprägungen auf spezifische situative Anforderungen und somit die Beschreibung von Kompetenzniveaus ermöglicht (vgl. Hartig und Klieme 2006). Dies ist bei der Auswertung mit Methoden der klassischen Testtheorie (vgl. z. B. Rost 2004) nicht möglich. Einige jüngere Entwicklungen im Rahmen der IRT – auch unter Beteiligung von Projekten des Schwerpunktprogramms wie beispielsweise Multidimensionale IRT-Modelle, IRT-Modelle zum Einbezug von Aufgabenmerkmalen sowie kognitive Diagnosemodelle – erscheinen hierbei für die Modellierung von Kompetenzen besonders vielversprechend zu sein (vgl. Frey und Hartig 2009; Klieme et al. 2008).

Insbesondere bei Kompetenzen in komplexeren und/oder breiteren Anwendungsfeldern lässt sich konkretes Testverhalten unter Umständen besser auf mehrere latente Dimensionen zurückführen als auf eine einzelne. Zur Modellierung derartiger Strukturen bieten sich mehrdimensionale IRT-Modelle an (vgl. Reckase 2009), bei denen jedes Testitem entweder einer spezifischen Dimension zugeordnet (*between-item multidimensionality*) oder auch durch mehrere Dimensionen erklärt wird (*within-item multidimensionality*). Diese Modelle finden bereits vielfach Anwendung, häufig im Rahmen von Large-Scale-Assessments (vgl. z. B. Adams und Wu 2002; Voss et al. 2005; Winkelmann et al. 2012). Ebenso werden sie in den Projekten des Schwerpunktprogramms eingesetzt (vgl. z. B. Bayrhuber et al. 2010; Gschwendtner et al. 2010; Hardy et al. 2010; Leutner et al. 2012; Roick et al. 2010; Winther 2011) und dort auch gezielt weiterentwickelt (vgl. Frey und Seitz 2009, 2010; Hartig und Hühner 2008, 2009).

IRT-Modelle zum Einbezug von Aufgabenmerkmalen wie das Linear-Logistische Testmodell (Fischer 1997) oder das Multicomponent-Latent-Trait-Model (Whitley 1980) dienen dazu, das zur Lösung der Testitems nötige Verhalten in seine Bestandteile zu zerlegen, indem die Itemparameter auf Anforderungsmerkmale dieser Items zurückgeführt werden. Sofern neben kognitiven auch inhaltsbezogene Itemmerkmale mit einbezogen werden, können diese Modelle auch dem kontextualisierten Charakter von Kompetenzen gerecht werden und damit einen wichtigen Beitrag zur Konstruktvalidierung von Kompetenzen liefern (vgl. Hartig und Frey 2012). Eine weitere interessante Entwicklung in jüngerer Zeit stellen sogenannte Explanatory Item-Response-Modelle (Wilson et al. 2008) dar, bei denen Zufallseffekte auf Itemebene integriert werden, wodurch die Modelle in ihrer Anwendung auf ein breiteres Spektrum an Kompetenzen flexibler werden.

Im Gegensatz zu traditionellen Modellen der IRT (vgl. z. B. Embretson und Reise 2000) gehen kognitive Diagnosemodelle davon aus, dass die Leistungen von Personen in einem Test nicht auf eine kontinuierliche latente Variable, sondern auf mehrere kategoriale latente Variablen (Kompetenzdimensionen) zurückgeführt werden können (vgl. Rupp et al. 2010). Es handelt sich bei diesen Modellen somit um multiple Klassifikationsmodelle, bei denen die Klassifikation der eingesetzten Testaufgaben zu den verschiedenen Dimensionen theoretisch begründet a priori stattfindet (vgl. Kunina-Habenicht et al. 2010; von Davier et al. 2008). Durch die Klassifikation von Personen hinsichtlich einer

oder mehrerer Kompetenzdimensionen können die Testergebnisse in Form von Kompetenzprofilen zurückgemeldet werden, wodurch sehr anschauliche Rückmeldungen spezifischer Stärken und Schwächen von Personen auf den postulierten Kompetenzen bzw. Kompetenzdimensionen möglich werden (vgl. z.B. Rupp et al. 2006). Ein weiterer Aspekt dieser Modelle, der sie für die Modellierung von Kompetenzen interessant erscheinen lässt, ist die Möglichkeit, nicht-kompensatorische Modelle zu schätzen. Diese Modelle gehen davon aus, dass verschiedene Teilkompetenzen zur Lösung einer bestimmten Aufgabe notwendig sind und Mängel in einer Teilkompetenz nicht durch Stärken in einer anderen Teilkompetenz ausgeglichen werden können. Sie entsprechen damit den Annahmen vieler psychologischer und fachdidaktischer Kompetenzmodelle und ermöglichen so eine explizite Prüfung dieser Annahmen, was mit den herkömmlichen statistischen Methoden nicht möglich ist (vgl. Hartig 2008). Auch wenn die Entwicklung dieser Modelle noch relativ am Anfang steht und sie unter anderem aufgrund ihrer Komplexität und der erforderlichen großen Stichproben nicht für alle Anwendungsbereiche geeignet sind (vgl. Leuders und Sodian 2013 in diesem Heft; Hartig und Frey 2013 in diesem Heft), erscheinen die bisherigen Ergebnisse – auch aus Projekten des Schwerpunktprogramms – vielversprechend (vgl. z. B. Kunina-Habenicht et al. 2009, 2010).

3 Forschungsbereich: Messkonzepte und Messverfahren

Zentrale Fragestellung: Wie lassen sich Kompetenzmodelle und darauf basierende psychometrische Modelle in konkrete empirische Messverfahren übertragen?

Zur Erfassung von Kompetenzen in unterschiedlichsten pädagogischen Kontexten (z. B. Individualdiagnostik, Evaluation von Programmen oder Institutionen, Large-Scale-Assessments, Grundlagenforschung etc.) werden in der Regel standardisierte Testinstrumente eingesetzt. Um der Komplexität verschiedener Kompetenzkonstrukte gerecht zu werden, erscheint es notwendig, bestehende Messkonzepte und Instrumente neuen Anforderungen anzupassen und zu erweitern. Messinstrumente zur Erfassung von Kompetenzen sollten allgemein einen klaren theoretischen Bezug aufweisen und Rückschlüsse über das Ausmaß des Beherrschens einer bestimmten Kompetenz in einem realen Anwendungsbezug ermöglichen (vgl. Klieme et al. 2008). Angesichts der Komplexität von Kompetenzstrukturen und der Notwendigkeit, die unterschiedlichen Fähigkeiten und Prozesse, die zur Bewältigung realer Alltagssituationen erforderlich sind, dezidiert erfassen zu können, ist es zunehmend wichtig geworden, dass Assessmentprozeduren auf kognitiven Kompetenzmodellen beruhen. Beispiele für eine auf theoretischen Kompetenzmodellen basierende empirische Erfassung von Kompetenzen liefern die Arbeiten des Berkeley Evaluation & Assessment Research (BEAR) Center, das sich auf die modellbasierte Kompetenzerfassung im naturwissenschaftlichen Unterricht konzentriert (vgl. z. B. Wilson 2005). Gleiches gilt für die unter Federführung des Instituts zur Qualitätsentwicklung im Bildungswesen (IQB) durchgeführten Arbeiten zur Evaluation und Normierung der nationalen Bildungsstandards in Deutschland (vgl. z. B. Granzer et al. 2009; Rupp et al. 2008).

Ebenso lassen sich aus den Projekten des Schwerpunktprogramms zahlreiche Beispiele für theorie- und modellbasierte Operationalisierungen von Kompetenzkonstrukten anfüh-

ren. So werden beispielsweise in den Projekten zur Erfassung von Lehrerkompetenzen neue und innovative Instrumente zur Erfassung der Beratungskompetenz von Lehrkräften (vgl. Bruder et al. 2010; Klug et al. 2012), zur Erfassung professioneller Unterrichtswahrnehmung über eine Kombination von Videosequenzen mit anderen standardisierten Verfahren (vgl. Jahn et al. 2011; Seidel et al. 2010) sowie zur experimentellen Untersuchung der diagnostischen Kompetenz von Lehrkräften bei Übergangsentscheidungen (Gräsel et al. 2010) entwickelt. Ebenso werden im Bereich der fächerübergreifenden Kompetenzen neue Instrumente zur Erfassung analytischer und dynamischer Problemlösekompetenz (Fleischer et al. 2010; Greiff und Funke 2010), zur Erfassung der Bewertungskompetenz in Problem- und Entscheidungssituationen (Gausmann et al. 2010) sowie zur Erfassung von Aspekten der Selbstregulationskompetenz beim Lernen aus Sachtexten (Schütte et al. 2012) und zur Erfassung von domänenspezifischem metakognitivem Wissen (Lingel et al. 2010) entwickelt. Hierbei kommen zum Teil auch Methoden der technologiebasierten Kompetenzerfassung (vgl. Hartig und Klieme 2007; siehe auch Frey und Hartig 2013 in diesem Heft) zum Einsatz. Diese Verfahren bieten unter anderem die Möglichkeit zur Simulation komplexer und dynamischer Systeme, welche mittels traditioneller Messmethoden kaum oder nur schwer zugänglich sind (vgl. Greiff und Funke 2009, 2010), sowie die Möglichkeit zur computergestützten adaptiven Testung (vgl. Frey und Seitz 2009), zur automatisierten Itemgenerierung (vgl. Holling et al. 2009; Zeuch et al. 2010) und zum dynamischen Testen (vgl. Dörfler et al. 2009, 2010).

4 Forschungsbereich: Nutzung von Informationen aus Diagnostik und Assessment

Zentrale Fragestellung: Welche Arten von Informationen aus Kompetenzmessungen können von Akteuren/innen im Bildungswesen auf welche Weise genutzt werden?

Kompetenzdiagnostik als Spezialfall der pädagogischen Diagnostik dient in erster Linie der Vorbereitung pädagogischer Entscheidungen auf individueller Ebene (vgl. Klieme und Leutner 2006). Demgegenüber dienen Assessment-Programme in Bildungsinstitutionen in der Regel der Vorbereitung von Entscheidungen auf Schul- und Unterrichtsebene (wie bei den landesweiten Vergleichsarbeiten in Deutschland; vgl. z. B. Leutner et al. 2007; Hosenfeld et al. 2008) oder der Vorbereitung politischer Entscheidungen auf Schulsystemebene (wie bei den nationalen und internationalen Large-Scale-Assessments; vgl. z. B. Granzer et al. 2009; Klieme et al. 2010c; Stanat et al. 2012). Hierbei ist offensichtlich, dass sich aus diesem Spannungsverhältnis unterschiedlicher Nutzungsperspektiven mitunter höchst unterschiedliche Anforderungen an die verwendeten Testinstrumente, das Testdesign sowie die Art der bereitgestellten Ergebnisse und Informationen als Resultat der Testung ergeben (vgl. Klieme et al. 2008; Leutner et al. 2007; Pellegrino et al. 2001). Leiss und Rakoczy (2013 in diesem Heft) fassen den derzeitigen Forschungsstand zur optimalen Gestaltung von Rückmeldungen aus Kompetenzmessungen zusammen. Es zeichnet sich ab, dass kompetenzbezogene Rückmeldungen mit Bezug auf Kompetenzniveaumodelle bei Schülerinnen und Schülern leistungs- und motivationsförderlicher sind als Rückmeldungen in Form von Schulnoten (vgl. Harks et al. im Druck). Dennoch gibt es bisher zu wenig systematische empirische Forschung zu diesem Themenkomplex, sodass sich nach wie vor eine Fülle offener Forschungsfragen stellen (vgl. Böge-

holz und Eggert 2013 in diesem Heft; McElvany und Rjosk 2013 in diesem Heft; Pant 2013 in diesem Heft sowie Leiss und Rakoczy 2013 in diesem Heft). Hierzu gehören unter anderem: Welche Art von diagnostischen Informationen sind für welche Empfänger (einzelne Schüler/-innen, Lehrkräfte, Schulen, Bildungspolitiker/-innen) hilfreich? Wie müssen diese Informationen aufbereitet werden, um von den Empfängern/innen sinnvoll genutzt werden zu können? Unter welchen Bedingungen können Informationen aus Assessment-Programmen auf Systemebene gegebenenfalls auch auf individueller Ebene genutzt werden? Welche Effekte (ggf. negative) können diagnostische Informationen aus standardisierten Kompetenztests auf die Gestaltung von schulischem Unterricht und die diagnostische Praxis von Lehrkräften haben (Stichwort: *washback effects*; vgl. z. B. Cizek 2001; siehe auch McElvany und Rjosk 2013 in diesem Heft)?

Im Rahmen des DFG-Schwerpunktprogramms werden beispielsweise die Effekte unterschiedlicher Arten von Ergebnismeldungen sowohl bei formativen als auch summativen Leistungsbeurteilungen auf die Motivation, Attribution sowie zukünftige Leistungen von Schülerinnen und Schülern untersucht (vgl. Besser et al. 2010; Harks et al. im Druck; Klieme et al. 2010b).

5 Fazit

Die oben dargestellten Forschungsbereiche des DFG-Schwerpunktprogramms bauen aufeinander auf. Dementsprechend befassten sich zahlreiche Einzelprojekte des Schwerpunktprogramms in verschiedenen Förderphasen schwerpunktmäßig mit unterschiedlichen Bereichen und Fragestellungen. So begannen die meisten Projekte in der ersten Phase zunächst mit der Entwicklung theoretischer Kompetenzmodelle. In den anschließenden Förderphasen wurden dann – unter Verwendung geeigneter psychometrischer Modelle – Testinstrumente zur Operationalisierung der untersuchten Kompetenzen entwickelt und eingesetzt und in einigen Fällen auch Studien zur Nutzung der gewonnenen Informationen durch verschiedene Akteure/innen im Bildungswesen durchgeführt.

Die Forschungsbereiche des Schwerpunktprogramms und die zugehörigen grundlegenden Fragestellungen beschreiben die zentralen Anforderungen, die sich bei der Modellierung von Kompetenzen stellen. Der oben skizzierte Forschungsstand verdeutlicht, wie im Rahmen des Schwerpunktprogramms mit diesen Anforderungen umgegangen wurde und zeigt, welche Fortschritte hinsichtlich der Kompetenzmodellierung im Bildungsbe-
reich in den letzten Jahren gemacht wurden. Gleichzeitig wird jedoch ebenfalls deutlich, welch anspruchsvolles Unterfangen die theoretisch fundierte und methodisch adäquate Erfassung von Kompetenzen nach wie vor darstellt.

Danksagung: Diese Veröffentlichung wurde ermöglicht durch Sachbeihilfen der Deutschen Forschungsgemeinschaft (Kennz.: KL 1057/9-2 und LE 645/11-2) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).

Literatur

- Adams, R., & Wu, M. (Hrsg.). (2002). *PISA 2000 technical report*. Paris: OECD.
- Artelt, C., & Schneider, W. (2011). Herausforderungen und Möglichkeiten der Diagnose und Modellierung von Kompetenzen und ihrer Entwicklung. Editorial. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 167–172.
- Artelt, C., Stanat, P., Schneider, W., & Schiefele, U. (2001). Lesekompetenz: Testkonzeption und Ergebnisse. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 69–137). Opladen: Leske + Budrich.
- Artelt, C., Neuenhaus, N., Lingel, K., & Schneider, W. (2012). Entwicklung und wechselseitige Effekte von metakognitiven und bereichsspezifischen Wissenskomponenten in der Sekundarstufe. *Psychologische Rundschau*, 63, 18–25.
- Baumert, J., Stanat, P., & Demmrich, A. (2001). PISA 2000: Untersuchungsgegenstand, theoretische Grundlagen und Durchführung der Studie. In Deutsches PISA-Konsortium (Hrsg.), *PISA 2000. Basiskompetenzen von Schülerinnen und Schülern im internationalen Vergleich* (S. 15–68). Opladen: Leske + Budrich.
- Bayrhuber, M., Leuders, T., Bruder, R., & Wirtz, M. (2010). Repräsentationswechsel beim Umgang mit Funktionen – Identifikation von Kompetenzprofilen auf der Basis eines Kompetenzstrukturmodells. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 28–39). Weinheim: Beltz.
- Besser, M., Leiss, D., Harks, B., Rakoczy, K., Klieme, E., & Blum, W. (2010). Kompetenzorientiertes Feedback im Mathematikunterricht: Entwicklung und empirische Erprobung prozessbezogener, aufgabenbasierter Rückmeldesituationen. *Empirische Pädagogik*, 24, 404–432.
- Blomberg, G., Seidel, T., & Prenzel, M. (2011). Neue Entwicklungen in der Erfassung pädagogisch-psychologischer Kompetenzen von Lehrpersonen [Themenschwerpunkt]. *Unterrichtswissenschaft*, 39, 98–172.
- Blossfeld, H.-P., Roßbach, H.-G., & Maurice, J. von. (2011). Education as a lifelong process. The German National Educational Panel Study (NEPS) (14. Sonderheft der Zeitschrift für Erziehungswissenschaft). Wiesbaden: VS Verlag für sozialwissenschaften.
- Blum, W., Neubrand, M., Ehmke, T., Senkbeil, M., Jordan, A., Ulfig, F., & Carstensen, C. H. (2004). Mathematische Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 47–92). Münster: Waxmann.
- Bögeholz, S., & Eggert, S. (2013). Welche Rolle spielt Kompetenzdiagnostik im Rahmen von Lehr-Lernprozessen? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0384-2). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Bremerich-Vos, A., & Böhme, K. (2009). Lesekompetenzdiagnostik – die Entwicklung eines Kompetenzstufenmodells für den Bereich Lesen. In D. Granzner, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik* (S. 219–249). Weinheim: Beltz.
- Bruder, S., Klug, J., Hertel, S., & Schmitz, B. (2010). Modellierung der Beratungskompetenz von Lehrkräften. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 274–285). Weinheim: Beltz.
- Cizek, G. J. (2001). More unintended consequences of high-stakes testing. *Educational Measurement, Issues, and Practice*, 20, 19–28.

- Cizek, G. J., & Bunch, M. B. (2007). *Standard-setting: A guide to establishing and evaluating performance standards on tests*. Thousand Oaks: Sage.
- Csapó, B. (2004). Knowledge and competencies. In J. Letschert (Hrsg.), *The integrated person. How curriculum development relates to new competencies* (S. 35–49). Enschede: CIDREE/SLO.
- Dörfler, D., Golke, S., & Artelt, C. (2009). Dynamic assessment and its potential for the assessment of reading competence. *Studies in Educational Evaluation*, 35, 77–82.
- Dörfler, D., Golke, S., & Artelt, C. (2010). Dynamisches Testen der Lesekompetenz. Theoretische Grundlagen, Konzeption und Testentwicklung. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 154–164). Weinheim: Beltz.
- Duit, R., & Möller, K. (2010). Themenschwerpunkt „Kompetenzdiagnose“ [Themenschwerpunkt]. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 263–327.
- Eggert, S., Bögeholz, S., Watermann, R., & Hasselhorn, M. (2010). Förderung von Bewertungskompetenz im Biologieunterricht durch zusätzliche Strukturierungshilfen beim Kooperativen Lernen – Ein Beispiel für Veränderungsmessung. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 299–314.
- Embretson, S. E. (1983). Construct validity: Construct representation versus nomothetic span. *Psychological Bulletin*, 93, 179–197.
- Embretson, S. E., & Reise, S. (2000). *Item response theory for psychologists*. Mahwah: Erlbaum.
- Fischer, G. H. (1997). Unidimensional linear logistic rasch models. In W. J. van der Linden & R. K. Hambleton (Hrsg.), *Handbook of modern item response theory* (S. 225–243). New York: Springer.
- Fleischer, J., Wirth, J., Rumann, S., & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz – Analyse von Aufgabenprofilen. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 239–248). Weinheim: Beltz.
- Fleischer, J., Leutner, D., & Klieme, E. (2012). Modellierung von Kompetenzen im Bereich der Bildung: Eine psychologische Perspektive (Themenheft der Psychologische Rundschau, Vol. 63(1)). Göttingen: Hogrefe Verlag.
- Frey, A., & Hartig, J. (2009). Assessment of competencies [Special section]. *Studies in Educational Evaluation*, 35, 55–101.
- Frey, A., & Hartig, J. (2013). Wann sollten computerbasierte Verfahren zur Messung von Kompetenzen anstelle von papier- und bleistift-basierten Verfahren eingesetzt werden? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi: 10.1007/s11618-013-0385-1). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Frey, A., & Seitz, N.-N. (2009). Multidimensional adaptive testing in educational and psychological measurement: Current state and future challenges. *Studies in Educational Evaluation*, 35, 89–94.
- Frey, A., & Seitz, N.-N. (2010). Multidimensionale adaptive Kompetenzdiagnostik: Ergebnisse zur Messeffizienz. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 40–51). Weinheim: Beltz.
- Gausmann, W., Eggert, S., Hasselhorn, M., Watermann, R., & Bögeholz, S. (2010). Wie verarbeiten Schüler/-innen Sachinformationen in Problem- und Entscheidungssituationen Nachhaltiger Entwicklung? In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 204–215). Weinheim: Beltz.

- Granzer, D., Köller, O., Bremerich-Vos, A., van den Heuvel-Panhuizen, M., Reiss, K., & Walther, G. (2009). *Bildungsstandards Deutsch und Mathematik*. Weinheim: Beltz.
- Gräsel, C., Krolak-Schwerdt, S., Nölle, I., & Hörstermann, T. (2010). Diagnostische Kompetenz von Grundschullehrkräften bei der Erstellung der Übergangsempfehlung. Eine Analyse aus der Perspektive der sozialen Urteilsbildung. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 286–295). Weinheim: Beltz.
- Greiff, S., & Funke, J. (2009). Measuring complex problem solving – The microDYN approach. In F. Scheuermann & J. Björnsson (Hrsg.), *The transition to computer-based assessment. New approaches to skills assessment and implications for large-scale testing* (S. 157–163). Luxembourg: Office for Official Publications of the European Communities.
- Greiff, S., & Funke, J. (2010). Systematische Erforschung komplexer Problemlösefähigkeit anhand minimal komplexer Systeme. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 216–227). Weinheim: Beltz.
- Gschwendtner, T., Geißel, B., & Nickolaus, R. (2010). Modellierung beruflicher Fachkompetenz in der gewerblich-technischen Grundbildung. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 258–269). Weinheim: Beltz.
- Hardy, I., Kleickmann, T., Koerber, S., Mayer, D., Möller, K., Pollmeier, J., Schwiippert, K., & Sodian, B. (2010). Die Modellierung naturwissenschaftlicher Kompetenz im Grundschulalter. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 115–125). Weinheim: Beltz.
- Harks, B., Rakoczy, K., Klieme, E., Hattie, J., & Besser, M. (im Druck). Indirekte und moderierte Effekte von Rückmeldung auf Leistung und Motivation. In H. Ditton & A. Müller (Hrsg.), *Rückmeldungen: Theoretische Grundlagen, empirische Befunde, praktische Anwendungsfelder*. München: Waxmann.
- Harsch, C., Pant, H. A., & Köller, O. (2010). *Calibrating standards-based assessment tasks for English as a first foreign language. Standard-setting procedures in Germany*. Münster: Waxmann.
- Hartig, J. (2007). Skalierung und Definition von Kompetenzniveaus. In B. Beck & E. Klieme (Hrsg.), *Sprachliche Kompetenzen. Konzepte und Messung – DESI-Studie* (S. 72–82). Weinheim: Beltz.
- Hartig, J. (2008). Psychometric models for the assessment of competencies. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 69–90). Göttingen: Hogrefe.
- Hartig, J., & Frey, A. (2012). Konstruktvalidierung und Skalenbeschreibung in der Kompetenzdiagnostik durch die Vorhersage von Aufgabenschwierigkeiten. *Psychologische Rundschau*, 63, 43–49.
- Hartig, J., & Frey, A. (2013). Sind Modelle der Item-Response-Theorie (IRT) das „Mittel der Wahl“ für die Modellierung von Kompetenzen? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0386-0). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Hartig, J., & Höhler, J. (2008). Representation of competencies in multidimensional IRT models with within-Item and between-Item multidimensionality. *Zeitschrift für Psychologie/Journal of Psychology*, 216, 89–101.
- Hartig, J., & Höhler, J. (2009). Multidimensional IRT models for the assessment of competencies. *Studies in Educational Evaluation*, 35, 57–63.

- Hartig, J., & Jude, N. (2008). Strukturen sprachlicher Kompetenzen. In E. Klieme, W. Eichler, A. Helmke, R. Lehmann, G. Nold, H.-G. Rolff, K. Schröder, G. Thomé & H. Willenberg (Hrsg.), *Unterricht und Kompetenzerwerb in Deutsch und Englisch. Ergebnisse der DESI-Studie* (S. 191–201). Weinheim: Beltz.
- Hartig, J., & Klieme, E. (2006). Kompetenz und Kompetenzdiagnostik. In K. Schweizer (Hrsg.), *Leistung und Leistungsdiagnostik* (S. 127–143). Heidelberg: Springer.
- Hartig, J., & Klieme, E. (2007). *Möglichkeiten und Voraussetzungen technologiebasierter Kompetenzdiagnostik*. Berlin: BMBF.
- Hartig, J., Klieme, E., & Leutner, D. (2008). *Assessment of competencies in educational contexts*. Göttingen: Hogrefe.
- Helmke, A., & Hosenfeld, I. (2004). Vergleichsarbeiten – Standards – Kompetenzstufen: Begriffliche Klärungen und Perspektiven. In M. Wosnitza, R. S. Jäger & A. Frey (Hrsg.), *Lernprozesse, Lernumgebung und Lerndiagnostik. Wissenschaftliche Beiträge zum Lernen im 21. Jahrhundert* (S. 56–75). Landau: Empirische Pädagogik.
- Holling, H., Bertling, J. P., & Zeuch, N. (2009). Automatic item generation of probability word problems. *Studies in Educational Evaluation*, 35, 71–76.
- Hosenfeld, I., Koch, U., Groß Ophoff, J., & Scherthan, F. (2008). Projekt VERA: Ergebnisorientierte Unterrichtsentwicklung durch internetgestützte externe Evaluation? In A. Breiter, A. Lange & E. Stauke (Hrsg.), *Schulinformationssysteme und datengestützte Entscheidungsprozesse* (S. 49–59). Bern: Peter Lang.
- Hülür, G., Wilhelm, O., & Robitzsch, A. (2011). Multivariate Veränderungsmodelle für Schulnoten und Schülerleistungen in Deutsch und Mathematik. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 173–185.
- Jahn, G., Prenzel, M., Stürmer, K., & Seidel, T. (2011). Varianten einer computergestützten Erhebung von Lehrerkompetenzen: Untersuchungen zu Anwendungen des Tools Observer. *Unterrichtswissenschaft*, 39, 136–153.
- Klieme, E. (2004). Was sind Kompetenzen und wie lassen sie sich messen? *Pädagogik*, 56, 10–13.
- Klieme, E., & Hartig, J. (2007). Kompetenzkonzepte in den Sozialwissenschaften und im erziehungswissenschaftlichen Diskurs. In I. Gogolin, M. Prenzel & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (8. Sonderheft der Zeitschrift für Erziehungswissenschaft, S. 11–29). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Klieme, E., & Leutner, D. (2006). Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen. Beschreibung eines neu eingerichteten Schwerpunktprogramms der DFG. *Zeitschrift für Pädagogik*, 52, 876–903.
- Klieme, E., Hartig, J., & Rauch, D. (2008). The Concept of competence in educational contexts. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 3–22). Göttingen: Hogrefe.
- Klieme, E., Leutner, D., & Kenk, M. (2010a). Kompetenzmodellierung – Zwischenbilanz des DFG-Schwerpunktprogramms und Perspektiven des Forschungsansatzes (56. Beiheft der Zeitschrift für Pädagogik). Weinheim: Beltz.
- Klieme, E., Bürgermeister, A., Harks, B., Blum, W., Leiß, D., & Rakoczy, K. (2010b). Leistungsbeurteilung und Kompetenzmodellierung im Mathematikunterricht.
- Klieme, E., Artelt, C., Hartig, J., Jude, N., Köller, O., Prenzel, M., Schneider, W., & Stanat, P. (2010c). *PISA 2009. Bilanz nach einem Jahrzehnt*. Münster: Waxmann.
- Klug, J., Bruder, S., Keller, S., & Schmitz, B. (2012). Hängen Diagnostische Kompetenz und Beratungskompetenz von Lehrkräften zusammen? Eine korrelative Untersuchung. *Psychologische Rundschau*, 63, 3–10.
- Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current issues in competence modelling and assessment. *Zeitschrift für Psychologie/Journal of Psychology*, 216, 61–73.

- Koerber, S., Sodian, B., Kropf, N., Mayer, D., & Schwippert, K. (2011). Die Entwicklung des wissenschaftlichen Denkens im Grundschulalter: Theorieverständnis, Experimentierstrategien, Dateninterpretation. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 16–21.
- Köller, O. (2009). Bildungsstandards. In R. Tippelt & B. Schmidt (Hrsg.), *Handbuch Bildungsforschung* (S. 529–548). Wiesbaden: VS Verlag für Sozialwissenschaften.
- 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. *Studies in Educational Evaluation*, 35, 64–70.
- Kunina-Habenicht, O., Wilhelm, O., Matthes, F., & Rupp, A. A. (2010). Kognitive Diagnosemodelle: theoretisches Potential und methodische Probleme. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 75–85). Weinheim: Beltz.
- Leiss, D., & Rakoczy, K. (2013). Wie können Ergebnisse der Kompetenzdiagnostik in Forschungsprojekten sinnvoll zurückgemeldet werden? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0389-x). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Leucht, M., Harsch, C., Pant, H. A., & Köller, O. (2012). Steuerung zukünftiger Aufgabenentwicklung durch Vorhersage der Schwierigkeiten eines Tests für die erste Fremdsprache Englisch durch Dutch Grid Merkmale. *Diagnostica*, 58, 31–44.
- Leuders, T., & Biehler, R. (im Druck). Kompetenzmodellierung für den Mathematikunterricht (Themenheft des Journals für Mathematik-Didaktik). Heidelberg: Springer.
- Leuders, T., & Sodian, B. (2013). Inwiefern sind Kompetenzmodelle dazu geeignet kognitive Prozesse von Lernenden zu beschreiben? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0381-5). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Leutner, D., Fleischer, J., Spoden, C., & Wirth, J. (2007). Landesweite Lernstandserhebungen zwischen Bildungsmonitoring und Individualdiagnostik. In I. Gogolin, M. Prenzel & H.-H. Krüger (Hrsg.), *Kompetenzdiagnostik* (8. Sonderheft der Zeitschrift für Erziehungswissenschaft, S. 149–167). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Leutner, D., Fleischer, J., Wirth, J., Greiff, S., & Funke, J. (2012). Analytische und dynamische Problemlösekompetenz im Lichte internationaler Schulleistungsvergleichsstudien. Untersuchungen zur Dimensionalität. *Psychologische Rundschau*, 63, 34–42.
- Lingel, K., Neuenhaus, N., Artelt, C., & Schneider, W. (2010). Metakognitives Wissen in der Sekundarstufe: Konstruktion und Evaluation domänenspezifischer Messverfahren. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 228–238). Weinheim: Beltz.
- McElvany, N., & Rjosk, C. (2013). Wann kann Kompetenzdiagnostik negative Auswirkungen haben? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0387-z). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Neubrand, M., Klieme, E., Lüdtke, O., & Neubrand, J. (2002). Kompetenzstufen und Schwierigkeitsmodelle für den PISA-Test zur mathematischen Grundbildung. *Unterrichtswissenschaft*, 30, 100–119.

- Neumann, K. (2013). Mit welchem Auflösungsgrad können Kompetenzen modelliert werden? In welcher Beziehung stehen Modelle zueinander, die Kompetenz in einer Domäne mit unterschiedlichem Auflösungsgrad beschreiben? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0382-4). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Pant, H. A. (2013). Wer hat einen Nutzen von Kompetenzmodellen? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0388-y). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Pant, H. A., Tiffin-Richards, S. P., & Köller, O. (2010). Standard-Setting für Kompetenztests im Large-Scale-Assessment. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 175–188). Weinheim: Beltz.
- Pant, H. A., Böhme, K., & Köller, O. (2012). Das Kompetenzkonzept der Bildungsstandards und die Entwicklung von Kompetenzstufenmodellen. In P. Stanat, H. A. Pant, K. Böhme & D. Richter (Hrsg.), *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik* (S. 49–55). Münster: Waxmann.
- Pellegrino, J. W., Chudowsky, N., & Glaser, R. (2001). *Knowing what students know. The science and design of educational assessment*. Washington, DC: National Academic Press.
- PISA-Konsortium Deutschland (2006). *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres*. Münster: Waxmann.
- Prenzel, M., Schöps, K., Rönnebeck, S., Senkbeil, M., Walter, O., Carstensen, C. H., & Hammann, M. (2007). Naturwissenschaftliche Kompetenz im internationalen Vergleich. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2006 – Die Ergebnisse der dritten internationalen Vergleichsstudie* (S. 63–105). Münster: Waxmann.
- Reckase, M. D. (2009). *Multidimensional item response theory*. Dordrecht: Springer.
- Reiss, K., & Winkelmann, H. (2009). Kompetenzstufenmodelle für das Fach Mathematik im Primarbereich. In D. Granzer, O. Köller, A. Bremerich-Vos, M. van den Heuvel-Panhuizen, K. Reiss & G. Walther (Hrsg.), *Bildungsstandards Deutsch und Mathematik* (S. 120–141). Weinheim: Beltz.
- Robitzsch, A. (2013). Wie robust sind Struktur- und Niveaumodelle? Wie zeitlich stabil und über Situationen hinweg konstant sind Kompetenzen? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0383-3). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Robitzsch, A., Dörfler, T., Pfost, M., & Artelt, C. (2011). Die Bedeutung der Itemauswahl und der Modellwahl für die längsschnittliche Erfassung von Kompetenzen. Lesekompetenzentwicklung in der Primarstufe. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie*, 43, 213–227.
- Roick, T., Stanat, P., Dickhäuser, O., Frederking, V., Meier, C., & Steinhauer, L. (2010). Strukturelle und kriteriale Validität der literarästhetischen Urteilskompetenz. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 165–174). Weinheim: Beltz.
- Rost, J. (2004). *Lehrbuch Testtheorie – Testkonstruktion* (2. Aufl.). Bern: Huber.
- Rupp, A. A., Leucht, M., & Hartung, R. (2006). „Die Kompetenzbrille aufsetzen“. Verfahren zur multiplen Klassifikation von Lernenden für Kompetenzdiagnostik in Unterricht und Testung. *Unterrichtswissenschaft*, 34, 195–219.

- Rupp, A. A., Vock, M., Harsch, C., & Köller, O. (2008). *Developing standards-based assessment tasks for English as a first foreign language – Context, processes, and outcomes in Germany*. Münster: Waxmann.
- Rupp, A. A., Templin, J., & Henson, R. A. (2010). *Diagnostic measurement. Theory, methods, and applications*. New York: Guilford.
- Schaper, N., Ulbricht, T., & Hochholdinger, S. (2008). Zusammenhang von Anforderungsmerkmalen und Schwierigkeitsparametern der MT21-Items. In S. Blömeke, G. Kaiser & R. Lehmann (Hrsg.), *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare. Erste Ergebnisse zur Wirksamkeit der Lehrerbildung* (S. 453–480). Münster: Waxmann.
- Schnotz, W., Horz, H., McElvany, N., Schroeder, S., Ullrich, M., Baumert, J., Hachfeld, A., & Richter, T. (2010). Das BITE-Projekt: Integrative Verarbeitung von Texten und Bildern in der Sekundarstufe I. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 143–153). Weinheim: Beltz.
- Schütte, M., Wirth, J., & Leutner, D. (2012). Lernstrategische Teilkompetenzen für das selbstregulierte Lernen aus Sachtexten. *Psychologische Rundschau*, 63, 26–33.
- Seidel, T., Blomberg, G., & Stürmer, K. (2010). „OBSERVER“ – Validierung eines videobasierten Instruments zur Erfassung der professionellen Wahrnehmung von Unterricht. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 296–306). Weinheim: Beltz.
- Stanat, P., Pant, H. A., Böhme, K., & Richter, D. (Hrsg.). (2012). *Kompetenzen von Schülerinnen und Schülern am Ende der vierten Jahrgangsstufe in den Fächern Deutsch und Mathematik*. Münster: Waxmann.
- Tesch, B., Porsch, R., Leupold, E., Grotjahn, R., Kleppin, K., Frenzel, J., Harsch, C., Leucht, M., Pant, H. A., Rupp, A. A., Tiffin-Richards, S. P., & Köller, O. (2010). Kompetenzstufenmodelle für Lese- und Hörverstehen in der ersten Fremdsprache. In O. Köller, M. Knigge & B. Tesch (Hrsg.), *Sprachliche Kompetenzen im Ländervergleich* (S. 50–63). Münster: Waxmann.
- Viering, T., Fischer, H. E., & Neumann, K. (2010). Die Entwicklung physikalischer Kompetenz in der Sekundarstufe I. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 92–103). Weinheim: Beltz.
- von Davier, M., DiBello, L., & Yamamoto, K. (2008). Reporting Test outcomes using models for cognitive diagnosis. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S. 151–174). Göttingen: Hogrefe.
- Vosniadou, S. (2008). *Handbook of research on conceptual change*. Mahwah: Erlbaum.
- Voss, A., Carstensen, C. H., & Bos, W. (2005). Textgattungen und Verstehensaspekte: Analysen von Leseverständnis aus den Daten der IGLU-Studie. In W. Bos, E.-M. Lankes, M. Prenzel, K. Schwippert, R. Valtin & G. Walther (Hrsg.), *IGLU. Vertiefende Analysen zu Leseverständnis, Rahmenbedingungen und Zusatzstudien* (S. 1–79). Münster: Waxmann.
- 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). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Hrsg.), *Defining and selecting key competencies* (S. 45–65). Seattle: Hogrefe.
- Whitely, S. E. (1980). Multicomponent latent trait models for ability tests. *Psychometrika*, 45, 479–494.

- Wilhelm, O., & Nickolaus, R. (2013). Was grenzt das Kompetenzkonzept von etablierten Kategorien wie Fähigkeit, Fertigkeit oder Intelligenz ab? In D. Leutner, E. Klieme, J. Fleischer & H. Kuper (Hrsg.), *Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen: aktuelle Diskurse im DFG-Schwerpunktprogramm* (18. Sonderheft der Zeitschrift für Erziehungswissenschaft, doi:10.1007/s11618-013-0380-6). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Wilson, M. (2003). On choosing a model for measuring. *Methods of Psychological Research Online*, 8, 1–22.
- Wilson, M. (2005). *Constructing measures. An item response modelling approach*. Mahwah: Erlbaum.
- Wilson, M., de Boeck, P., & Carstensen, C. (2008). Explanatory item response models: A brief introduction. In J. Hartig, E. Klieme & D. Leutner (Hrsg.), *Assessment of competencies in educational contexts* (S.91–120). Göttingen: Hogrefe.
- Winkelmann, H., Robitzsch, A., Stanat, P., & Köller, O. (2012). Mathematische Kompetenzen in der Grundschule. Struktur, Validierung und Zusammenspiel mit allgemeinen kognitiven Fähigkeiten. *Diagnostica*, 58, 15–30.
- Winther, E. (2011). Das ist doch nicht fair! – Mehrdimensionalität und Testfairness in kaufmännischen Assessments. *Zeitschrift für Berufs- und Wirtschaftspädagogik*, 107, 218–238.
- Zeuch, N., Geerlings, H., Holling, H., van der Linden, W. J., & Bertling, J. P. (2010). Regelgeleitete Konstruktion von statistischen Textaufgaben. Anwendung von linear logistischen Testmodellen und Aufgabencloning. In E. Klieme, D. Leutner & M. Kenk (Hrsg.), *Kompetenzmodellierung Eine aktuelle Zwischenbilanz des DFG-Schwerpunktprogramms* (56. Beiheft der Zeitschrift für Pädagogik, S. 52–63). Weinheim: Beltz.