

STEFAN RUMANN, JENS FLEISCHER, HOLGER STAWITZ, JOACHIM WIRTH UND DETLEV LEUTNER

## Vergleich von Profilen der Naturwissenschafts- und Problemlöse-Aufgaben der PISA 2003-Studie

Comparison of profiles of science and problem solving tasks in the PISA 2003-study

### Zusammenfassung

Ausgangspunkt zweier vorgestellter Studien sind Ergebnisse der PISA 2003-Studie. Neben den fachlichen Kompetenzen Mathematik, Lesen und Naturwissenschaften wurde in PISA-2003 zusätzlich als fächerübergreifende Kompetenz das analytische Problemlösen im internationalen Vergleich erhoben. Bei Schülerinnen und Schülern in Deutschland trat ein besonderes Ergebnis auf: Im Problemlösen zeigten sie signifikant überdurchschnittliche Ergebnisse, während sie in den Naturwissenschaften nur durchschnittliche Werte erzielten. Die vorliegenden Studien untersuchen, was die charakteristischen Merkmale der Naturwissenschafts- und Problemlöse-Aufgaben aus PISA sind und inwiefern diese Merkmale einen Einfluss auf die Schwierigkeit der Aufgaben haben.

Schlüsselwörter: PISA 2003, Naturwissenschaftliche Kompetenz, Problemlösekompetenz, Aufgabenanalyse

### Abstract

Two studies are presented which are based on results of the PISA study 2003. Beside the subject-specific competencies mathematics, reading and science, PISA also assessed cross-curricular problem solving. For the German students a conspicuous result was found: They performed significantly above the international average in problem solving, but showed average results in science. The present studies analyze the specific characteristics of science and problem solving tasks from PISA and examine if these characteristics have an influence on the difficulty of these tasks.

Keywords: PISA 2003, science competence, cross-curricular problem solving competence, task-analysis

## 1 Einleitung

Die Fähigkeit Probleme zu lösen wird als eine Schlüsselkompetenz bezeichnet (Funke, 2003), die sowohl in beruflichen Belangen wie auch in alltäglichen Situationen eine wichtige Rolle spielt (z.B. OECD, 2004). Problemlösen ist dementsprechend sowohl als fächerübergreifende als auch als fachspezifische Kompetenz Forschungsgegenstand verschiedener Fachdisziplinen wie beispielsweise der Psychologie oder der Naturwissenschaften (z.B. Bodner & Herron, 2002; Funke, 2003; Leutner, Fleischer, Wirth, Greiff & Funke, im Druck). Entsprechend wird Problemlösen auch in den Kerncurricu-

la verschiedener Schulfächer in Deutschland als zu erwerbende fachspezifische Kompetenz definiert (z.B. Melle, Parchmann & Sumfleth, 2004) und findet sich auch in den nationalen Bildungsstandards für die naturwissenschaftlichen Fächer Biologie, Chemie und Physik in unterschiedlichen Kompetenzbereichen (KMK, 2005a, 2005b, 2005c) wieder. In den Einheitlichen Prüfungsanforderungen für das Abitur im Fach Chemie wird Problemlösen in den Anforderungsbereich III eingeordnet und dort beschrieben als das „planmäßige und kreative Bearbeiten komplexer Problemstellungen mit dem Ziel, selbstständig [...] zu einer Lösung zu gelangen“ (KMK, 2004, S. 12). Zudem findet sich

Problemlösen auch in der Forschung zur naturwissenschaftlichen Erkenntnisgewinnung (scientific inquiry; vgl. Bybee, 2000; AAAS, 1993) und zum naturwissenschaftlichen Denken (scientific reasoning; vgl. Dunbar, 1995; Kuhn, Amsel & O'Loughlin, 1988).

Neben domänenspezifischen Kompetenzen hat auch Problemlösen Eingang in internationale Schulleistungsvergleichsstudien gefunden. So wurde in der PISA-Studie 2003 fächerübergreifendes Problemlösen in Form der analytischen Problemlösekompetenz erfasst (Leutner, Fleischer & Wirth, 2006; Leutner, Klieme, Meyer & Wirth, 2004; OECD, 2003).

Im Vergleich zwischen analytischer Problemlösekompetenz und naturwissenschaftlicher Kompetenz zeigt sich in PISA-2003 für Schülerinnen und Schüler in Deutschland ein interessantes Ergebnis: Auf der internationalen PISA-Skala ( $M=500$ ;  $SD=100$ ) übersteigt die im internationalen Vergleich überdurchschnittliche Problemlösekompetenz ( $M=513$ ) die im internationalen Vergleich nur durchschnittliche naturwissenschaftliche Kompetenz ( $M=502$ ) um 11 Punkte (Leutner et al., 2004; Rost, Walter, Carstensen, Senkbeil & Prenzel, 2004). Die Diskrepanz, die besonders im unteren Leistungsbereich ausgeprägt ist, ist vor dem Hintergrund der relativ hohen messfehlerbereinigten Korrelation zwischen beiden Kompetenzbereichen ( $r=.80$ ; OECD, 2005) umso erstaunlicher.

Dieser Niveauunterschied ist erklärungsbedürftig. Eine Hypothese ist, dass sich im fächerübergreifenden Problemlösetest kognitive Potenziale der Schülerinnen und Schüler offenbaren, die im Unterricht scheinbar nicht hinreichend zum Aufbau naturwissenschaftlicher Kompetenzen ausgeschöpft werden („Potenzialausschöpfungshypothese“; vgl. Leutner et al., 2004, OECD, 2004). Will man die beschriebene Diskrepanz zwischen naturwissenschaftlicher Kompetenz und fächerübergreifender Problemlösekompetenz erklären und in didaktische Maßnahmen zur Verbesserung der naturwissenschaftlichen Kompetenz umsetzen, ist es erforderlich, den Aufbau und die Struktur

beider Kompetenzbereiche weiter aufzuklären, als dies in der bisherigen Forschung geschehen ist. Hierzu werden in der vorliegenden Untersuchung strukturelle Gemeinsamkeiten und Unterschiede bezüglich der Konstruktion und der kognitiven Anforderungsprofile von Testaufgaben zur Erfassung naturwissenschaftlicher Kompetenz einerseits und fächerübergreifender Problemlösekompetenz andererseits näher analysiert werden (vgl. Fleischer, Wirth, Rumann, & Leutner, 2010).

## 2 Naturwissenschaftliche Kompetenz und fächerübergreifende Problemlösekompetenz in PISA-2003

Grundlegend für die Konstruktion der naturwissenschaftlichen Aufgaben in PISA-2003 ist das im internationalen Bewertungsrahmen definierte Verständnis von naturwissenschaftlicher Kompetenz im Sinne einer Scientific Literacy (OECD, 2003). Basierend auf Bybee's (1997) vierstufiger Taxonomie von Scientific Literacy wird die naturwissenschaftliche Grundbildung fünfzehnjähriger Jugendlicher als Fähigkeit konzeptualisiert, „naturwissenschaftliches Wissen anzuwenden, naturwissenschaftliche Fragen zu erkennen und aus Belegen Schlussfolgerungen zu ziehen, um Entscheidungen zu verstehen und zu treffen [...]“ (OECD, 2003, S. 133; vgl. Rost et al., 2004). Ausgehend von dieser allgemeinen Definition werden drei qualitative „Aspekte“ naturwissenschaftlicher Grundbildung unterschieden, die gleichrangig bei der Aufgabenkonstruktion berücksichtigt wurden und die damit wesentlich die inhaltliche Ausgestaltung der Aufgabenstämme beeinflussen. Die Aufgaben des internationalen Naturwissenschaftstest lassen sich entsprechend den drei Aspekten klassifizieren. Bei diesen Aspekten handelt es sich um:

*Verstehen naturwissenschaftlicher Konzepte.* Hierin spiegelt sich der fachinhaltliche Anspruch der Naturwissenschaftsaufgaben wider. So lassen sich alle Aufgaben einer naturwissenschaftlichen Bezugsdomäne,

Biologie, Chemie oder Physik, zuordnen. Da PISA konzeptuelles Wissen und nicht Faktenwissen („nominal literacy“) testet, werden die für die drei Naturwissenschaften relevanten basalen Ideen („big ideas“, bspw. „Struktur und Eigenschaft der Materie“) in der Aufgabenstellung berücksichtigt.

*Durchführen naturwissenschaftlicher Prozesse.* Hierunter versteht man die Fähigkeit, wissenschaftliche Untersuchungen methodisch und inhaltlich zu verstehen, hieraus evidenzbasierte Schlussfolgerungen zu ziehen und damit insgesamt naturwissenschaftliche Prozesse zu erklären.

*Anwendungsgebiete in relevanten Kontexten.* Die Fähigkeit, konzeptuelle oder prozedurale Kompetenzen anzuwenden, zeigt sich vor allem in alltagsnahen, relevanten Kontexten oder Situationen. Der internationale Bewertungsrahmen unterscheidet hierbei die Kontexte „Leben und Gesundheit“, „Erde und Umwelt“ und „Technologie“.

Wie bereits erwähnt, wurde neben den drei fachlichen Domänen Lesen, Mathematik und Naturwissenschaften in PISA-2003 zusätzlich, als fächerübergreifende Kompetenz, analytisches Problemlösen getestet. Das Lösen von Problemen lässt sich allgemein beschreiben als „zielorientiertes Denken und Handeln in Situationen, für deren Bewältigung keine routinierten Vorgehensweisen verfügbar sind“ (Klieme, Funke, Leutner, Reimann & Wirth, 2001, S. 185). In ähnlicher Weise wird fächerübergreifende Problemlösekompetenz in der PISA-Studie 2003 definiert als „Fähigkeit einer Person, kognitive Prozesse zu nutzen, um sich mit solchen realen, fachübergreifenden Problemstellungen auseinanderzusetzen und sie zu lösen, bei denen der Lösungsweg nicht unmittelbar erkennbar ist und die zur Lösung nutzbaren Wissensbereiche nicht einem einzelnen Fachgebiet [. . .] entstammen“ (OECD, 2003, S. 156; vgl. Leutner et al., 2004). Beim Lösen von Problemen handelt es sich somit nicht um den Umgang mit Routine-Situationen, für die eine Lösung aus dem Gedächtnis abgerufen werden kann (in Abgrenzung zu einer Aufgabe; Funke,

2003). Ebenfalls zentral ist der Prozesscharakter des Problemlösens. So beschreibt Polya (1945) die Bearbeitung eines Problems als eine Abfolge von vier Prozessschritten, die sich in ähnlicher Art und Weise auch im „Assessment Framework“ zu PISA-2003 wiederfinden: (1) understanding the problem, (2) devising a plan, (3) carrying out the plan und (4) looking back (vgl. OECD 2003). Fächerübergreifende analytische Problemlösekompetenz wurde in PISA-2003 durch drei Arten von Problemstellungen operationalisiert, die sich hinsichtlich der zu erreichenden Ziele, der beteiligten Prozesse und der potenziell schwierigkeitsbestimmenden Merkmale unterscheiden (vgl. Leutner et al., 2004):

*Entscheidungen treffen.* Die Lösung dieser Art von Problemstellungen erfordert es, die gegebene Problemsituation sowie alternative Vorgehensweise zu verstehen und eine Entscheidung vor dem Hintergrund verschiedener einschränkender Bedingungen zu treffen. Häufig müssen hierzu Informationen aus unterschiedlichen Quellen kombiniert werden.

*Systeme analysieren und entwerfen.* Zur Lösung dieser Art von Problemstellungen muss entweder die Logik und Funktionsweise eines komplexen Systems verstanden oder selbst ein funktionierendes System konstruiert werden, welches bestimmte Zielzustände erreichen soll. Hierfür müssen häufig Zusammenhänge und Wechselwirkungen zwischen Variablen identifiziert werden.

*Fehler suchen.* Die Lösung dieser Art von Problemstellungen erfordert das Verstehen zentraler Merkmale eines Systems oder eines Mechanismus, um eine Fehlfunktion diagnostizieren zu können. Hierfür müssen häufig verbale und graphische Informationen kombiniert werden (OECD, 2003, 2004).

Bei der Konstruktion der Aufgaben wurden zudem drei *Kontexte* unterschieden, je nachdem wie dicht die Aufgabe am persönlichen Umfeld des Aufgaben-Bearbeiters angesiedelt ist („privat-persönlich“, „Arbeit und Freizeit“ und „Gesellschaft“).

### 3 Testaufgaben und Aufgabenmerkmale

Nach Klauer (1987, S. 15) ist eine Aufgabe allgemein durch die Verknüpfung einer Stimulus-Komponente mit einer Response-Komponente charakterisiert: „Die Stimulus-Komponente besteht aus einem bestimmten Inhalt, der in einer bestimmten Art und Weise vorgelegt wird. Die Response-Komponente besteht aus der Handlung, die an der Stimulus-Komponente ausgeübt werden soll.“

Dabei kann eine Aufgabe in vielerlei Funktionen auftreten: Im schulischen Kontext dienen Aufgaben einerseits häufig dazu, den Unterricht zu strukturieren, andererseits aber auch Unterrichtsinhalte zur Rekonstruktion fachlicher Inhalte anzubieten (Fischer, 2001; Leutner, Fischer, Kauertz, Schabram & Fleischer, 2008). Darüber hinaus werden im naturwissenschaftlichen Unterricht Aufgaben eingesetzt, um die Fähigkeit zum Problemlösen selbst zu unterrichten. Im Zuge der Output-Orientierung und der Messung von Bildungserfolg, z. B. durch die Einführung der nationalen Bildungsstandards (z. B. für Chemie: KMK, 2005b), hat die Bedeutung der Konstruktion von Aufgaben zur Kompetenzmessung in der Forschung und das Interesse an Kompetenz-Testaufgaben in den Bildungseinrichtungen stark zugenommen (Kauertz, Fischer, Mayer, Sumfleth & Walpuski, 2010).

In vielen empirischen Studien wurde bereits nachgewiesen, dass bestimmte – auch formale – Aufgabenmerkmale einen Einfluss auf die Aufgabenbearbeitung haben (z. B. Marx, & Rinkens, 2008; Rupp, Vock, Harsch & Köller, 2008): So konnte zum Beispiel im Bereich der Antwortformate eine unterschiedliche kognitive Anforderung bei verschiedenen Formaten und damit eine unterschiedliche Schwierigkeit nachgewiesen werden (z. B. Martinez, 1999). Neben dem Aufgabenformat werden von Schnotz & Kürschner (2007) die Länge bzw. die Relevanz des Textes als ein schwierigkeitsbestimmendes Aufgabenmerkmal identifiziert. Unterschiedliche Schwierigkeitsstufen inner-

halb von Texten aufgrund unterschiedlicher sprachlicher Komplexität sind in der Linguistik weitreichend beschrieben (z. B. Bamberger & Vanecek, 1984). Auch hinsichtlich des Aufgabenkontexts lassen sich Effekte auf die Motivation und Leistung der Aufgabenbearbeiter in bestimmten Arbeitssituationen und Aufgabenstellungen nachweisen (z. B. Cordova & Lepper, 1996). Die Vielzahl von möglichen Einflüssen auf die Wirkung der Aufgabe bei Schülerinnen und Schülern verdeutlicht, dass die Konstruktion von Testaufgaben besonderer Aufmerksamkeit bedarf. Es gibt eine Reihe von Forschungsansätzen, die Vorschläge zur Konstruktion von Testaufgaben machen. So empfiehlt beispielsweise Hammann (2006) den Einsatz von Anwendungsaufgaben zur Kompetenzdiagnose in den Naturwissenschaften, wobei er zugleich ein Defizit in diesem Bereich feststellt. Er benennt vier „Säulen“, nach denen man sich bei der Konstruktion solcher Aufgaben richten kann: 1. Der Aufgabenkontext, 2. die Kompetenzen, 3. das Wissen und 4. die affektive Dimension der Aufgabe. Dieser Ansatz lässt aber die Frage nach der äußeren Gestalt, wie beispielsweise dem Antwortformat, außer Acht. Doch selbst innerhalb eines gegebenen Antwortformats sind Faktoren zu berücksichtigen, die unterschiedlich auf die Schwierigkeit von Aufgaben wirken können: Beispielsweise müssen bei der Konstruktion von Distraktoren in Multiple-Choice-Aufgaben verschiedene Aspekte berücksichtigt werden (Green, 1984). So zeigen Haladyna und Downing (1989), dass bei der Erstellung von Distraktoren verschiedene Vorgaben, wie beispielsweise gleiche Satzlängen der unterschiedlichen Antwortoptionen, beachtet werden sollten (Haladyna & Downing, 1989, vgl. Haladyna, 2004). Studien, die sich mit Aufgabenformaten beschäftigen, dokumentieren, dass es keineswegs trivial ist, eine Aufgabe für einen Kompetenztest zu formulieren. Es ist somit deutlich geworden, dass viele Merkmale, selbst die äußere Form einer Aufgabe, mit Hinblick auf Schwierigkeit und Motivation zu einer unterschiedlichen Wirkung führen können.

## 4 Fragestellung

Ziel der vorliegenden Studie war es, strukturelle Gemeinsamkeiten und Unterschiede bezüglich der Konstruktion und der kognitiven Anforderungen von Testaufgaben zur Erfassung naturwissenschaftlicher Kompetenz und fächerübergreifender Problemlösekompetenz zu analysieren. Dadurch lassen sich Hinweise auf mögliche Erklärungsansätze für die bei PISA-2003 gefundenen Niveauunterschiede zwischen diesen beiden Kompetenzen ableiten. Hierzu wurden zunächst die PISA-2003 Testaufgaben zur naturwissenschaftlichen Kompetenz und zum Problemlösen einer ausführlichen Merkmals- und Anforderungsanalyse unterzogen. In einem zweiten Schritt wurden die so gewonnenen Aufgabenprofile zur Konstruktion neuer Testaufgaben verwendet: Zum einen wurde das Naturwissenschafts-Profil auf die Problemlöseaufgaben übertragen und zum anderen das Problemlöse-Profil auf die Naturwissenschaftsaufgaben. Mit diesen abgewandelten Aufgaben wurden neue Daten erhoben, um zu prüfen inwiefern die Abwandlung der Aufgaben einen Einfluss auf die Schwierigkeit dieser Aufgaben hat. Im Einzelnen wurden dabei zwei Forschungsfragen behandelt:

1. Was sind die charakteristischen Merkmale der Naturwissenschafts- und Problemlöseaufgaben aus PISA-2003?
2. Inwiefern lässt sich durch eine gezielte Veränderung der charakteristischen Merkmale der Naturwissenschafts- und Problemlöseaufgaben die Schwierigkeit dieser Aufgaben verändern?

## 5 Studie I – Aufgaben- und Anforderungsanalyse

### 5.1 Methode

Bei der Analyse von Testaufgaben lässt sich zwischen niedrig-inferenten und hoch-infe-

renten Merkmalen unterscheiden. *Niedrig-inferente* Merkmale bezeichnen relativ leicht zu erfassenden Merkmale, wie beispielsweise die Sichtstruktur von Aufgaben (z. B. das Antwortformat). Merkmale, deren Erfassung ein höheres Maß an schlussfolgerndem Denken und Interpretation erfordern, wie beispielsweise das sprachliche Niveau einer Aufgabe, werden als *hoch-inferent* bezeichnet (vgl. Clausen, Reusser & Klieme, 2003). Zur Analyse der PISA-Aufgaben wurden zwei Instrumente zur Erfassung beider Arten von Aufgabenmerkmalen entwickelt. Zur Erfassung niedrig-inferenter Merkmale wurde ein Kategoriensystem in Anlehnung an die COACTIV-Studie (Jordan et al., 2006) entwickelt. Das Kategoriensystem enthält 12 detailliert beschriebene Beurteilungsdimensionen (im Folgenden als Facetten bezeichnet; vgl. Seidel, Prenzel, Duit & Lehrke, 2003), teilweise mit Unterfacetten, die folgende Bereiche abdecken: Aufgabenstellung, Aufgabeninformation, Antwortformat, Lösungsvorgaben, Anzahl von Operatoren (Denk- und Handlungsaufforderungen), Lösungs- und Strukturierungshilfen, Sprache, Fächertyp, Themengebiet (Schulfach), Anwendungsfeld, Aufgabenkontext, Aufgabentyp.

Zur Erfassung hoch-inferenter Merkmale wurde ein Ratinginstrument in Anlehnung an die Konzeption von Langer und Schulz von Thun (2007) entwickelt, bei dem die Ausprägung von Aufgabenmerkmalen auf 4-stufigen Ratingskalen (von „trifft voll und ganz zu“ bis „trifft überhaupt nicht zu“) eingeschätzt werden sollte. Die insgesamt 44 Items des Instruments beziehen sich auf fünf übergeordnete Facetten: Vorwissen, Aufgabengestaltung, Aufgabeninformationen, Lösungsweg, lösungsrelevante Bedingungen. Anhand dieser Instrumente und zugehöriger detaillierter Manuale wurden die Testaufgaben aus PISA-2003 zur naturwissenschaftlichen Kompetenz ( $n = 35$ ) und zum Problemlösen ( $n = 19$ ) durch drei geschulte Beurteiler unabhängig voneinander und ohne Kenntnis der jeweiligen Domäne (Naturwissenschaft vs. Problemlösen) nach den Beurteilungsfacetten klassifiziert.

## 5.2 Ergebnisse

Als Indikator für die Objektivität der Beurteilungsfacetten wurden für jedes Rater-Paar Cohens-Kappa-Koeffizienten als Maß der Beurteilerübereinstimmung berechnet. Um eine hinreichend hohe Beurteilerübereinstimmung zu gewährleisten, wurden im Falle der niedrig-inferenten Facetten des Kategoriensystems für die weiteren Analysen nur diejenigen Beurteilungsfacetten berücksichtigt, bei denen die Kappa-Koeffizienten aller Rater-Paare über .71 lagen (vgl. Wirtz & Caspar, 2002). Bei den hoch-inferenten Merkmalen, für die auf Grund des erhöhten Maßes an schlussfolgerndem Denken bei der Einschätzung per se eine geringere Beurteilerübereinstimmung zu erwarten war, wurden nur diejenigen Facetten berücksichtigt, bei denen mindestens ein Rater-Paar Kappa-Koeffizienten von über .61 und die anderen Rater-Paare Werte von über .41 erreichten (vgl. Landis & Koch, 1977). Lediglich bei zwei Beurteilungsfacetten („Fächertyp“ und „Aufgabentyp“) des Kategoriensystems konnte trotz Überarbeitung des Manuals keine hinreichende Beurteilerübereinstimmung erreicht werden. Für die übrigen 21 Items konnte eine für die weiteren Analysen hinreichend hohe Beurteilerübereinstimmung erreicht werden.

Zur Prüfung von Unterschieden zwischen den Domänen wurden, je nach Skalenniveau der Facetten, Chi-Quadrat-Tests sowie t-Tests für unabhängige Stichproben berechnet. Das Signifikanzniveau wurde auf 5% festgesetzt und eine sequentielle  $\alpha$ -Fehler-Adjustierung für multiple Tests nach Holm (1979) vorgenommen.

Im Folgenden werden zentrale Ergebnisse der Aufgabenklassifikation für die Facetten berichtet, für die nach den obigen Vorgaben eine hinreichend hohe Objektivität angenommen werden kann und für die sich auch nach  $\alpha$ -Fehler-Adjustierung noch statistisch signifikante Unterschiede zwischen

Naturwissenschafts- und Problemlöseitems zeigen ließen.

### Zentrale Ergebnisse der niedrig-inferenten Aufgabenanalyse

*Aufgabenstellung.* Bei den Naturwissenschaftsitems werden häufiger Fragen gestellt (63% der Items) während die Aufgabenstellung bei den Problemlöseitems häufiger (58% der Items) aus einer Aufforderung besteht ( $\chi^2_{(2)} = 4.95$ ;  $p < .05$ ).

*Aufgabeninformationen.* Bei den Problemlöseitems werden in 68% der Fälle kurze Einleitungstexte (maximal 10 Propositionen<sup>1</sup>) verwendet, während bei den Naturwissenschaftsitems in 69% der Fälle längere Einleitungstexte (mehr als 10 Propositionen) verwendet werden ( $t_{(52)} = 4.66$ ;  $p < .05$ ). Die Naturwissenschaftsitems werden darüber hinaus durch einen eher technisch gestalteten Einleitungssatz begonnen, der erläutert, wie mit dem Stimulus-Material umgegangen werden soll. Die Problemlöseitems werden hingegen hauptsächlich durch einen eher inhaltlicher Einleitungssatz begonnen, der die Aufgaben-Bearbeiter darauf vorbereitet, was in der Aufgabe inhaltlich zu tun ist ( $\chi^2_{(2)} = 23.11$ ;  $p < .01$ ). Bei 47% der Problemlöseitems werden mehr als vier Zahlen gegeben, während dies bei den Naturwissenschaftsitems nur bei 20% der Items vorkommt ( $t_{(52)} = 2.46$ ;  $p < .05$ ).

*Antwortformat.* Die Naturwissenschaftsitems weisen häufiger als die Problemlöseitems ein Single- oder Multiple-Choice Antwortformat auf (57% vs. 37% der Fälle). Bei den Problemlöseitems sind hingegen häufiger Graphiken oder Tabellen zu erstellen bzw. zu vervollständigen, als dies bei den Naturwissenschaftsitems der Fall ist (32% vs. 3% der Fälle) ( $\chi^2_{(7)} = 31.36$ ;  $p < .05$ ).

*Sprache.* Wie zu erwarten werden bei den Naturwissenschaftsitems häufiger und mehr Fachbegriffe verwendet, als dies bei den

<sup>1</sup> Unter Propositionen versteht man die „kleinsten Bedeutungseinheiten, die als selbstständige Behauptung stehen können“ (Anderson, 1989, S. 112).

Problemlöseitems der Fall ist ( $t_{(52)} = 13.31$ ;  $p < .05$ ). Bei 79% der Problemlöseitems werden keine Fachbegriffe verwendet, während bei 94% der Naturwissenschaftsitems zwei und mehr Fachbegriffe verwendet werden.

*Themengebiet (Schulfach).* Die in den Problemlöseitems angesprochenen Themengebiete werden zu 53% der Mathematik zugeordnet, während dies nur bei 3% der Naturwissenschaftsitems der Fall ist. Diese werden zu 78% einem der drei naturwissenschaftlichen Fächer zugeordnet, was lediglich bei 11% der Problemlöseitems der Fall ist ( $\chi^2_{(8)} = 43.92$ ;  $p < .05$ ).

*Aufgabenkontext.* Die Problemlöseitems werden mehrheitlich einem persönlichen Kontext zugeordnet (53%), während die Naturwissenschaftsitems mehrheitlich (53%) einem globalen Kontext zugeordnet werden ( $\chi^2_{(4)} = 27.52$ ;  $p < .05$ ).

formation als lösungsrelevant eingeschätzt, während bei den Naturwissenschaftsitems stärker auch nichtlösungsrelevante Informationen enthalten sind ( $t_{(49)} = -8.06$ ;  $p < .05$ ).

*Lösungsweg.* Zur Lösung der Problemlöseitems müssen häufiger mehrere verschiedene kognitive Prozesse durchlaufen werden, als dies bei den Naturwissenschaftsitems der Fall ist ( $t_{(43)} = -3.46$ ;  $p < .05$ ). Zur Lösung der Problemlöseitems ist darüber hinaus in höherem Ausmaß Handlungswissen notwendig, während das zur Lösung der Naturwissenschaftsitems notwendige Wissen eher dem deklarativen Sachwissen zugeordnet werden kann ( $t_{(39)} = 7.74$ ;  $p < .05$ ).

*Lösungsrelevante Bedingungen.* Die zur Lösung der Items zu beachtenden Bedingungen werden bei den Problemlöseitems stärker explizit hervorgehoben, als dies bei den Naturwissenschaftsitems der Fall ist ( $t_{(52)} = -5.23$ ;  $p < .05$ ).

## Zentrale Ergebnisse der hoch-inferenten Aufgabenanalyse

*Vorwissen.* In den Problemlöseitems wird hauptsächlich Alltagsvokabular verwendet, während die zentralen Begriffe in den Naturwissenschaften als eher nicht aus dem Alltag bekannt bewertet werden ( $t_{(52)} = -4.77$ ;  $p < .05$ ).

*Aufgabengestaltung.* Die Verständlichkeit der Sprache wird bei den Naturwissenschaftsaufgaben deutlich geringer eingeschätzt als dies bei den Problemlöseitems der Fall ist ( $t_{(52)} = -4.66$ ;  $p < .05$ ). Zudem wird die Präsentation der Aufgabeninformationen bei den Problemlöseitems als stärker motivierend eingeschätzt, als dies bei den Naturwissenschaftsaufgaben der Fall ist ( $t_{(51)} = -3.80$ ;  $p < .05$ ). Darüber hinaus werden die Problemlöseitems ebenfalls als stärker alltagsnah im Vergleich zu den Naturwissenschaftsaufgaben eingeschätzt ( $t_{(26)} = -5.23$ ;  $p < .05$ ).

*Aufgabeninformationen.* Wie schon bei den Ergebnissen des niedrig-inferenten Kategoriensystems wird bei den Problemlöseitems ein Großteil der gegebenen Aufgabenin-

## 6 Studie II – Übertragung der Aufgabenprofile und Testung der Aufgabenschwierigkeiten

### 6.1 Methode

Die in Studie I identifizierten charakteristischen Aufgabenprofile der Naturwissenschaftsitems und der Problemlöseitems wurden in Studie II dazu verwendet, die jeweiligen Items abzuwandeln und dadurch neue Testitems zu konstruieren. Hierbei wurde zum einen das Naturwissenschafts-Profil auf die Problemlöseitems und zum anderen das Problemlöse-Profil auf die Naturwissenschaftsitems übertragen. Die so entstandenen neuen Aufgabentypen werden im Folgenden als *abgewandelte* Aufgaben bezeichnet. Drei Problemlöseitems wurden aus den weiteren Analysen ausgeschlossen, da für sie eine Abwandlung nicht sinnvoll möglich erschien, sodass für Problemlösen  $n = 16$  Items für die Abwandlung zur Verfügung standen. Um eine bessere Vergleichbarkeit zwischen den Aufgabentypen zu gewährleisten, wurde für Naturwissenschaften eine nach Itemschwie-

rigkeit (siehe OECD, 2005) stratifizierte Stichprobe von  $n=16$  Items gezogen. Bei der Abwandlung der Naturwissenschaftsitems wurde versucht, so viele Merkmale der Problemlöseitems wie möglich, für die sich in Studie I signifikante Unterschiede gezeigt hatten, auf die Naturwissenschaftsitems zu übertragen. In analoger Weise wurde bei der Abwandlung der Problemlöseitems verfahren. Tabelle 1 stellt die Merkmale dar, die bei den Items abgewandelt wurden. Einige Merkmale ließen sich jedoch nicht wechselseitig auf Items des jeweils anderen Kompetenzbereichs übertragen. So konnte bei den abgewandelten Naturwissenschaftsitems beispielsweise nicht vollständig auf Fachbegriffe verzichtet werden. Bei einigen abgewandelten Problemlöseaufgaben mussten z. B. Skizzen erhalten bleiben, da sie lösungsrelevante Informationen enthielten, die nicht in einen Fließtext überführt werden konnten.

Die abgewandelten Aufgaben wurden anschließend von zwei geschulten Ratern anhand des Kategoriensystems und des Ratinginstruments aus Studie I klassifiziert und mit den Originalaufgaben verglichen. Nach Überarbeitung einzelner Items ließen sich für alle Merkmale signifikante Unterschiede zwischen abgewandelten und originalen Aufgaben in der intendierten Art und Weise zeigen.

*Instrumente.* Die insgesamt 64 Testitems (32 Problemlöse- und 32 Naturwissenschaftsitems, davon jeweils 16 originale und 16 abgewandelte Items) wurden in einem rotierten Multiple-Matrix-Design (vgl. OECD, 2005) auf 16 Testhefte (à 24 Items) verteilt. Die Testdauer betrug drei Schulstunden à 45 min verteilt über zwei Tage.

*Stichprobe.* Die Testinstrumente wurden 647 Schülerinnen und Schülern (46.4% weiblich; Altersdurchschnitt  $M=15.5$  Jahre) zu Beginn der Jahrgangsstufe 10 aus 48 Klassen (HS:

Tab. 1: Aufgabenmerkmale, die zur Abwandlung der Items verwendet wurden

Merkmals	Naturwissenschaftliche Aufgaben	Problemlöseaufgaben
Einleitungssatz:	eher technischer Art	eher auf Inhalt bezogen
Art der Aufgabeninformation:	eher Textinformation	eher Graphiken und Tabellen
Textlänge:	eher lange Einleitungstexte	eher kurze Einleitungssätze
Textstruktur:	Fließtext	deutlich gegliedert
Textrelevanz:	auch irrelevanter Text	nahezu ausschließlich lösungsrelevante Informationen
Sprachniveau:	komplexe Sätze mit vielen Fachbegriffen	Alltagssprache ohne Fachbegriffe
Art der Aufgabenstellung:	eher Fragen	eher Aufforderungen bzw. Handlungsanweisungen
Antwortformat:	eher Single- und Multiple-Choice	eher Vervollständigung und Erstellung von Tabellen und Graphiken
Kontext:	eher globaler Kontext	eher persönlicher Kontext

14 Klassen<sup>2</sup>, GS: 7 Klassen, GYM: 7 Klassen) in Nordrhein-Westfalen zur Bearbeitung vorgelegt.

## 6.2 Ergebnisse

Die originalen und die abgewandelten Testitems wurden nach dem dichotomen Raschmodell mit der Software ConQuest 2.0 (Wu, Adams, Wilson & Haldane, 2007) skaliert. Dabei wurde ein 2-dimensionales Modell (Naturwissenschaften vs. Problemlösen) spezifiziert, welches die originalen und die abgewandelten Naturwissenschaftsitems der ersten Dimension zuordnet und die originalen und die abgewandelten Problemlöseitems der zweiten Dimension zuordnet. Als hinreichend guter Itemfit wurden INFIT-Werte zwischen 0.8 und 1.2 angesehen (vgl. Bond & Fox, 2001). 12 Items wurden aufgrund zu geringer INFIT-Werte aus dem Modell entfernt. Die Reliabilitäten beider Dimensionen, erfasst über die EAP-Reliabilität (vgl. Rost, 2004), liegen bei  $r = .74$  (für Naturwissenschaften) und bei  $r = .76$  (für Problemlösen). Die Varianz als Indikator für die Trennschärfe der Items liegt für die Dimension Naturwissenschaften bei  $\sigma^2 = 1.18$  und für die Dimension Problemlösen bei  $\sigma^2 = 1.71$ . Die Korrelationen der beiden latenten Dimensionen des Modells beträgt  $r = .83$  und entspricht damit in etwa der Korrelation zwischen Naturwissenschaften und Problemlösen in PISA-2003 ( $r = .80$ ; Leutner et al., 2004).

Um zu prüfen, ob die Abwandlung der Naturwissenschafts- und Problemlöseaufgaben einen Einfluss auf die Schwierigkeit dieser Aufgaben hat, wurden jeweils getrennt für die beiden spezifizierten Dimensionen (Naturwissenschaften vs. Problemlösen) die Schwierigkeitsparameter der originalen Items mit denen der abgewandelten verg-

lichen. Für die abgewandelten Naturwissenschaftsitems ergibt sich eine mittlere Itemschwierigkeit von  $M = -0.374$  ( $SD = 0.720$ ), die geringer ist als die mittlere Itemschwierigkeit der originalen Naturwissenschaftsitems ( $M = 0.315$ ,  $SD = 0.986$ ). Dieser hypothesenkonforme Unterschied ist statistisch signifikant ( $t_{(25)} = 2.02$ ;  $p(\text{einseitig}) = .027$ ) und weist eine große Effektstärke auf ( $d = 0.80$ ). Für die abgewandelten Problemlöseitems ergibt sich eine mittlere Itemschwierigkeit von  $M = 0.851$  ( $SD = 1.334$ ), die höher ist als die mittlere Itemschwierigkeit der originalen Problemlöseitems ( $M = -0.312$ ,  $SD = 0.951$ ). Dieser hypothesenkonforme Unterschied ist ebenfalls signifikant und weist ebenfalls eine hohe Effektstärke auf ( $t_{(25)} = -2.48$ ;  $p(\text{einseitig}) = .010$ ;  $d = 1.01$ ).

## 7 Zusammenfassende Diskussion

Ausgehend von den unterschiedlichen Leistungsergebnissen der Schülerinnen und Schüler in Deutschland in den Kompetenzbereichen Naturwissenschaften und Problemlösen in PISA-2003 wurden in dieser Untersuchung zunächst die zum Einsatz gekommenen Naturwissenschafts- und Problemlöseaufgaben einer ausführlichen Analyse auf niedrig- und hoch-inferentem Niveau unterzogen. Die Ergebnisse zeigen charakteristische Unterschiede zwischen den Aufgabenprofilen der beiden untersuchten Kompetenzbereiche: Erwartungskonform zeichnen sich die Naturwissenschaftsaufgaben durch ein hohes Maß an Fachsprachlichkeit aus, während bei den Problemlöseaufgaben überwiegend Alltagsvokabular verwendet wird. Zudem erfordern die Naturwissenschaftsaufgaben zur Lösung in erster Linie fachspezifisches deklaratives Sachwissen, während bei den Problemlö-

<sup>2</sup> Aufgrund des zu erwartenden selektiven Dropouts in der Hauptschule und der PISA-2003 Ergebnisse, wonach der Niveauunterschied zwischen Naturwissenschaften und Problemlösen im unteren Leistungsbereich besonders ausgeprägt ist, erfolgte eine stärkere Berücksichtigung von Hauptschulklassen bei der Auswahl der Stichprobe.

seaufgaben Handlungswissen im Vordergrund steht, welches hauptsächlich dem Allgemeinwissen zuzuordnen ist. Der Aufgabenstamm der Naturwissenschaftsaufgaben besteht überwiegend aus Texten, die zu einem großen Teil auch lösungsirrelevante Informationen enthalten und als sprachlich komplexer eingeschätzt werden, als dies bei den Problemlöseaufgaben der Fall ist. Die lösungsrelevanten Bedingungen der Naturwissenschaftsaufgaben werden im Unterschied zu den Problemlöseaufgaben zudem kaum visuell hervorgehoben oder explizit genannt. Diese Befunde decken sich zum Teil mit den Aufgabenanalysen von Kulgemeyer (2009), der ebenfalls einen hohen Anteil an irrelevanter Information in den Stimulus-Materialien der Naturwissenschaftsaufgaben bei PISA-2003 nachweisen kann.

Die aus der Aufgabenanalyse resultierenden spezifischen Profile der Naturwissenschafts- und Problemlöseaufgaben wurden anschließend wechselseitig aufeinander übertragen: Originale Problemlöseaufgaben wurden mit dem Profil der Naturwissenschaftsaufgaben „maskiert“ und umgekehrt. Die so erhaltenen abgewandelten Aufgaben wurden dann, zusammen mit den Originalaufgaben, 15-jährigen Schülerinnen und Schülern im Rahmen eines Multiple-Matrix-Designs zur Bearbeitung vorgelegt. Die erhaltene Datenstruktur folgt einem 2-dimensionalen Modell (Naturwissenschaft vs. Problemlösen). Wie der Mittelwertvergleich der Dimension „Problemlösen“ zeigte, sind die abgewandelten Aufgaben signifikant schwerer als die originalen Aufgaben. Umgekehrt zeigt sich beim Mittelwertvergleich der Dimension „Naturwissenschaft“, dass die abgewandelten Aufgaben leichter sind als die originalen Aufgaben. Dieser Befund deutet darauf hin, dass die Erfassung der naturwissenschaftlichen Kompetenz – stärker als die Erfassung der Problemlösekompetenz – von Unterschieden auf weiteren Variablen, insbesondere dem Vorwissen, determiniert wird.

Kritisch anzumerken ist, dass der nachgewiesene Einfluss der unterschiedlichen Aufgabenprofile von Naturwissenschafts- und

Problemlöseaufgaben bei PISA-2003 keinen Rückschluss auf das absolute Abschneiden der Schülerinnen und Schüler in Deutschland im internationalen Vergleich zulässt. Da die vorliegende Untersuchung allein auf einer deutschen, und überdies vergleichsweise kleinen, Stichprobe gründet, muss vor monokausalen Erklärungsansätzen für die Ausprägung der naturwissenschaftlichen Kompetenz von Schülerinnen und Schülern in Deutschland („die Aufgaben sind schuld“) entschieden gewarnt werden. Gleichwohl verdeutlichen die Ergebnisse auch die prinzipielle Andersartigkeit der Konstrukte „Naturwissenschaftliche Kompetenz“ und „Problemlösekompetenz“. Dieser Andersartigkeit soll in weiteren Untersuchungen nachgegangen werden. Hierzu werden Teilkompetenzen der Naturwissenschaftlichen Kompetenz und der Problemlösekompetenz erhoben und die Zusammenhänge dieser Teilkompetenzen mit Hilfe von Strukturgleichungsmodellen untersucht.

## Danksagung

Diese Veröffentlichung wurde ermöglicht durch eine Sachbeihilfe der Deutschen Forschungsgemeinschaft (Kennz.: RU 1437/4-2 und LE 645/12-2) im Schwerpunktprogramm „Kompetenzmodelle zur Erfassung individueller Lernergebnisse und zur Bilanzierung von Bildungsprozessen“ (SPP 1293).

## Literatur

- American Association for the Advancement of Science (AAAS). (1993). *Benchmarks for science literacy*. New York, NY: Oxford University Press.
- Anderson, J. (1989). *Kognitive Psychologie*. 2. Auflage. Berlin: Spektrum.
- Bamberger, R. & Vanecek, E. (1984). *Lesen-Verstehen-Lernen-Schreiben – Die Schwierigkeitsstufen von Texten in deutscher Sprache*. Frankfurt a.M.: Diesterweg.
- Bodner, G. M. & Herron, J. D. (2002). Problem-solving in chemistry. In J. K. Gilbert, O. de Jong, R. Justi, D. F. Treagust & J. H. van Driel (Eds.), *Chemical education: Towards research-based practise* (pp. 235–266). Dordrecht: Kluwer.
- Bond, T. G. & Fox, C. M. (2001). *Applying the rasch model*. Mahwah, NJ: Erlbaum.
- Bybee, R.W. (1997). Toward an understanding of scientific literacy. In W. Gräber & C. Bolte (Eds.), *Scientific Literacy* (pp. 37–68) Kiel: IPN.
- Bybee, R. W. (2000). Teaching science as inquiry. In J. Minstrell & E. H. van Zee (Eds.), *Inquiry into inquiry learning and teaching in science* (pp. 20–46). Washington, DC: American Association for the Advancement of Science.
- Clausen, M., Reusser, K. & Klieme E. (2003). Unterrichtsqualität auf der Basis hochinferenter Unterrichtsbeurteilungen: Ein instruktionspsychologischer Vergleich zwischen Deutschland und der deutschsprachigen Schweiz. *Unterrichtswissenschaft*, 31, 122–141.
- Cordova, D. I. & Lepper, M. R. (1996). Intrinsic motivation and the process of learning: Beneficial effects of contextualization, personalization, and choice. *Journal of Educational Psychology*, 88, 715–730.
- Dunbar, K. (1995). How scientists really reason: scientific reasoning in real-world laboratories. In R. J. Sternberg & J. Davidson (Eds.), *The nature of insight* (pp. 365–395). Cambridge, MA: MIT Press.
- Fischer, H. E. & Draxler, D. (2001). Aufgaben und naturwissenschaftlicher Unterricht. *Der Mathematische und Naturwissenschaftliche Unterricht*, 54, 388–393.
- Fleischer, J., Wirth, J., Rumann, S. & Leutner, D. (2010). Strukturen fächerübergreifender und fachlicher Problemlösekompetenz – Analyse von Aufgabenprofilen. *Zeitschrift für Pädagogik*, 56. Beiheft, 239–248.
- Funke, J. (2003). *Problemlösendes Denken*. Stuttgart: Kohlhammer.
- Green, K. (1984). Effects of item characteristics on multiple-choice item difficulty. *Educational and Psychological Measurement*, 44, 551–561.
- Haladyna, T. & Downing, S. (1989). A taxonomy of multiple-choice item-writing rules. *Applied Measurement in Education*, 2, 37–50.
- Haladyna, T. M. (2004). *Developing and validating multiple-choice test items*. Mahwah, NJ: Erlbaum.
- Hammann, M. (2006). Kompetenzförderung und Aufgabenentwicklung. *Der mathematische und naturwissenschaftliche Unterricht*, 59, 85–95.
- Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6, 65–70.
- Jordan, A. et al. (2006). *Klassifikation für Mathematikaufgaben: Dokumentation der Aufgabenkategorisierung im COACTIV-Projekt*. Berlin: Max Planck Institut für Bildungsforschung.
- Kauertz, A., Fischer, H., Mayer, J., Sumfleth, E. & Walpuski, M. (2010). Standardbezogene Kompetenzmodellierung in den Naturwissenschaften der Sekundarstufe. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 135–154.
- Klauer, K.J. (1987). Kriteriumsorientierte Tests. Göttingen: Hogrefe.
- Klieme, E., Funke, J., Leutner, D., Reimann, P. & Wirth, J. (2001). Problemlösen als fächerübergreifende Kompetenz? Konzeption und erste Resultate aus einer Schulleistungsstudie. *Zeitschrift für Pädagogik*, 47, 179–200.
- KMK. (2004). *Einheitliche Prüfungsanforderungen in der Abiturprüfung Chemie* (Beschluss vom 05. Februar 2004). Verfügbar unter: [http://www.kmk.org/fileadmin/veroeffentlichungen\\_beschluesse/1989/1989\\_12\\_01-EPA-Chemie.pdf](http://www.kmk.org/fileadmin/veroeffentlichungen_beschluesse/1989/1989_12_01-EPA-Chemie.pdf) (11.10.2011).
- KMK. (2005a). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Biologie für den mittleren Schulabschluss* (Beschluss von 16. Dezember 2004). München: Wolters Kluwer.
- KMK. (2005b). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Chemie für den mittleren Schulabschluss* (Beschluss von 16. Dezember 2004). München: Wolters Kluwer.
- KMK. (2005c). *Beschlüsse der Kultusministerkonferenz – Bildungsstandards im Fach Physik für den mittleren Schulabschluss* (Beschluss von 16. Dezember 2004). München: Wolters Kluwer.
- Kuhn, D., Amsel, E. & O'Loughlin, M. (1988). *The development of scientific thinking skills*. San Diego, CA: Academic Press.

- Kulgemeier, C. (2009). *PISA-Aufgaben im Vergleich – Strukturanalyse der Naturwissenschaftsitems aus den PISA-Durchläufen 2000 bis 2006*. Norderstedt: Books on Demand GmbH.
- Landis, J. R. & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, 33, 159–174.
- Langer, I. & Schulz von Thun, F. (2007). Messung komplexer Merkmale in Psychologie und Pädagogik: Ratingverfahren. Münster: Waxmann.
- Leutner, D., Fleischer, H.E., Kauertz, A., Schabram, N. & Fleischer, J. (2008). Instruktionspsychologische und fachdidaktische Aspekte der Qualität von Lernaufgaben und Testaufgaben im Physikunterricht. In J. Thonhauser (Hrsg.), *Aufgaben als Katalysatoren von Lernprozessen* (S. 169–181). Münster: Waxmann.
- Leutner, D., Fleischer, J. & Wirth, J. (2006). Problemlösekompetenz als Prädiktor für zukünftige Kompetenz in Mathematik und in den Naturwissenschaften. In: M. Prenzel, J. Baumert, W. Blum, R. Lehmann, D. Leutner, M. Neubrand, R. Pekrun, J. Rost & U. Schiefele (Hrsg.), *PISA 2003. Untersuchungen zur Kompetenzentwicklung im Verlauf eines Schuljahres* (S. 119–137). Münster: Waxmann.
- Leutner, D., Klieme, E., Meyer, K. & Wirth, J. (2004). Problemlösen. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs*, (S. 147–175). Münster: Waxmann.
- Leutner, D., Fleischer, J., Wirth, J., Greiff, S. & Funke, J. (im Druck). Analytisches und dynamisches Problemlösen im Lichte internationaler Schulleistungsvergleichsstudien. *Psychologische Rundschau*.
- Martinez, M. E. (1999). Cognition and the question of test item format. *Educational Psychologist*, 34, 207–218.
- Marx, A. & Rinkens, H.-D. (2008). Anforderungsmerkmale der MT21-Testitems und ihre Weiterentwicklung aus mathematikdidaktischer Sicht. 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. 425–452). Münster: Waxmann.
- Melle, I., Parchmann, I. & Sumfleth, E. (2004) Kerncurriculum Chemie – Ziele, Rahmenbedingungen und Ansatzpunkte für die Sekundarstufe II. In A. Pitton (Hrsg.), *Chemie- und physikdidaktische Forschung und naturwissenschaftliche Bildung* (S. 129–132). Münster: Lit.
- OECD. (2003). *The PISA 2003 assessment framework – Mathematics, reading, science and problem solving knowledge and skills*. Paris: OECD.
- OECD. (2004). *Problem solving for tomorrow's world. First measurements of cross-curricular competencies from PISA 2003*. Paris: OECD.
- OECD. (2005). *PISA 2003 – Technical report*. Paris: OECD.
- Polya, G. (1945). *How to solve it: A new aspect of mathematical method*. Princeton, NJ: Princeton University Press.
- Rost, J., Walter, O., Carstensen, C. H., Senkbeil, M. & Prenzel, M. (2004). Naturwissenschaftliche Kompetenz. In PISA-Konsortium Deutschland (Hrsg.), *PISA 2003. Der Bildungsstand der Jugendlichen in Deutschland – Ergebnisse des zweiten internationalen Vergleichs* (S. 111–146). Münster: Waxmann.
- Rost, J. (2004). *Lehrbuch Testtheorie und Testkonstruktion*. Bern: Huber.
- 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.
- Schnotz, W. & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19, 469–508.
- Seidel, T., Prenzel, M., Duit, R. & Lehrke, M. (2003). *Technischer Bericht zur Videostudie „Lehr-Lern-Prozesse im Physikunterricht“*. Kiel: IPN.
- Wirtz, M. & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wu, M. L., Adams, R. J., Wilson, M. R. & Haldane, S. A. (2007). *ACER ConQuest version 2.0: Generalised item response modelling software*. Camberwell, VIC: ACER Press.

### **Kontakt**

Prof. Dr. Stefan Rumann

Didaktik der Chemie, Universität Duisburg-Essen

Schützenbahn 70

45127 Essen

*stefan.rumann@uni-due.de*

### **Autoreninformation**

Dr. rer. nat. Stefan Rumann ist Professor für Chemiedidaktik an der Universität Duisburg-Essen. Seine Forschungsschwerpunkte liegen im Bereich des naturwissenschaftlichen Problemlösens sowie in der chemiebezogenen Professionsentwicklung im tertiären Ausbildungssektor.

Jens Fleischer, Dipl. Psychologe, ist wissenschaftlicher Mitarbeiter am Lehrstuhl für Lehr-Lernpsychologie der Universität Duisburg-Essen.

Dr. rer. nat. Holger Stawitz promovierte im Arbeitskreis von Stefan Rumann und ist zurzeit Studienreferendar für die Fächer Chemie und Sozialwissenschaften in Nordrhein-Westfalen.

Dr. rer. nat. Joachim Wirth ist Professor für Lehr-Lernforschung an der Ruhr-Universität Bochum. Seine Forschungsschwerpunkte liegen im Bereich der Diagnostik fächerübergreifender Kompetenzen wie selbstreguliertes Lernen oder Problemlösen sowie in der lernförderlichen Gestaltung computerbasierter Simulationen und multimedialer Lernumgebungen.

Dr. phil. Detlev Leutner ist Professor für Lehr-Lernpsychologie an der Universität Duisburg-Essen. Seine Forschungsschwerpunkte liegen im Bereich des selbstregulierten Lernens, dem Lernen mit Multimedia, der Evaluation von Ausbildungsprogrammen sowie der Forschungsmethodik und Diagnostik.

