

Kulgemeyer, Christoph; Schecker, Horst

Physikalische Kommunikationskompetenz. Empirische Validierung eines normativen Modells

Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 18 (2012), S. 29-54



Quellenangabe/ Reference:

Kulgemeyer, Christoph; Schecker, Horst: Physikalische Kommunikationskompetenz. Empirische Validierung eines normativen Modells - In: Zeitschrift für Didaktik der Naturwissenschaften : ZfDN 18 (2012), S. 29-54 - URN: urn:nbn:de:0111-pedocs-317489 - DOI: 10.25656/01:31748

<https://nbn-resolving.org/urn:nbn:de:0111-pedocs-317489>

<https://doi.org/10.25656/01:31748>

in Kooperation mit / in cooperation with:



IPN

Leibniz-Institut für die Pädagogik der
Naturwissenschaften und Mathematik

<https://www.leibniz-ipn.de>

Nutzungsbedingungen

Gewährt wird ein nicht exklusives, nicht übertragbares, persönliches und beschränktes Recht auf Nutzung dieses Dokuments. Dieses Dokument ist ausschließlich für den persönlichen, nicht-kommerziellen Gebrauch bestimmt. Die Nutzung stellt keine Übertragung des Eigentumsrechts an diesem Dokument dar und gilt vorbehaltlich der folgenden Einschränkungen: Auf sämtlichen Kopien dieses Dokuments müssen alle Urheberrechtshinweise und sonstigen Hinweise auf gesetzlichen Schutz beibehalten werden. Sie dürfen dieses Dokument nicht in irgendeiner Weise abändern, noch dürfen Sie dieses Dokument für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, aufführen, vertreiben oder anderweitig nutzen.

Mit der Verwendung dieses Dokuments erkennen Sie die Nutzungsbedingungen an.

Terms of use

We grant a non-exclusive, non-transferable, individual and limited right to using this document.

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

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

Kontakt / Contact:

peDOCS

DIPF | Leibniz-Institut für Bildungsforschung und Bildungsinformation

Informationszentrum (IZ) Bildung

E-Mail: pedocs@dipf.de

Internet: www.pedocs.de

Mitglied der


Leibniz-Gemeinschaft

CHRISTOPH KULGEMEYER UND HORST SCHECKER

Physikalische Kommunikationskompetenz – Empirische Validierung eines normativen Modells

Empirical validation of a model for physics-related communication competence

ZUSAMMENFASSUNG

In diesem Beitrag werden die Ergebnisse einer quantitativen Studie zur Überführung eines normativen Modells physikalischer Kommunikationskompetenz in ein deskriptives Modell dargestellt. Die Studie erkundet einen bisher empirisch kaum bearbeiteten Kompetenzbereich. Zu dem Ausgangsmodell (Kulgemeyer & Schecker, 2009) wurde ein schriftliches Testverfahren entwickelt und in zwei Phasen pilotiert. An der Hauptstudie nahmen 216 Schülerinnen und Schüler aus 10. Klassen des Gymnasiums teil. Die Datenauswertung (u. a. mehrdimensionale Rasch-Analysen) diente zur Überprüfung der zunächst präskriptiv beschriebenen Dimensionen des Modells und deren Komponenten. Das empirisch validierte Modell wird vorgestellt. Es enthält u. a. eine Beschreibung von Kompetenzstufen.

Schlüsselwörter: Bildungsstandards, Kompetenz, Kommunikation, Kommunikationskompetenz

ABSTRACT

We present the results of a quantitative study about the empirical validation of a model for communication competence in physics. Starting from a normative model we describe the development of a paper-and-pencil-test. This test was validated in two pilot-studies and then used in a main study with 216 students (gymnasium, grade 10). The model was refined so that it now describes students' actual competence structures. We can show that levels of communication competence in physics can be scaled according to psycholinguistic processes.

Keywords: Educational Standards, Competence, Communication, Communication Competence

1 Einleitung

Der Beitrag stellt die Ergebnisse der empirischen Validierung eines Modells physikalischer Kommunikationskompetenz

vor, d. h. Prozess und Resultate der Überführung eines normativen Ausgangsmodells in ein empirisch überprüfbares deskriptives Modell (zur Unterscheidung s. Schecker & Parchmann, 2006). Das Mo-

dell und seine theoretischen Grundlagen haben wir in dieser Zeitschrift bereits detailliert vorgestellt (Kulgemeyer & Schecker, 2009). Das Modell wird hier daher lediglich zusammenfassend präsentiert. Die für die empirische Überprüfung des Modells eingesetzten Testinstrumente und ihr Entwicklungsprozess werden in diesem Beitrag jedoch ebenso ausführlich beschrieben wie die Methodik der Analyse der Binnenstruktur des Modells. Bei der Datenanalyse kamen mehrdimensionale Rasch-Analysen und auch klassische statistische Verfahren zum Einsatz.

2 Ein Kompetenzmodell für physikalische Kommunikation

Im Sinne der deutschen nationalen Bildungsstandards für die drei Naturwissenschaften (Mittlerer Schulabschluss, KMK, 2005) ist ein Modell zu formulieren, das Kommunikationskompetenz domänenspezifisch für die Physik beschreibt. Mit der Beschreibung von Kommunikation als naturwissenschaftlichem Kompetenzbereich stehen die deutschen Bildungsstandards jedoch international keineswegs allein. So finden sich beispielsweise in den amerikanischen National Science Education Standards (CSMEE, 1996) und in diversen australischen Curricula (Hafner, 2007) auf Naturwissenschaftskommunikation bezogene Bildungsziele. Eine Explizierung dessen, was unter der als domänenspezifisch angenommenen Kompetenz zu verstehen sei, wird jedoch weder in den internationalen Curricula noch in den deutschen Bildungsstandards vor-

genommen. Ein entsprechendes Modell für die Naturwissenschaftsdidaktik blieb daher zunächst ein Forschungsdesiderat. Der scheinbar verwandte Bereich der Argumentation ist hingegen schon seit längerem Gegenstand umfänglicher Forschungen (z. B. Bricker & Bell, 2008; Driver, Newton & Osborne, 2000). Einige Aspekte von Argumentation können durchaus vor dem Hintergrund kommunikativer Prozesse beschrieben werden. Argumentation orientiert sich jedoch wesentlich an Logik und zielt zumeist auf die argumentative Untermauerung eines vertretenen Standpunkts zu einem Sachverhalt ab (Osborne, Erduran & Simon, 2004). Argumentation lässt sich in wichtigen Teilen auch dem Kompetenzbereich Bewertung der Bildungsstandards zuordnen. Kommunikation orientiert sich hingegen an der für einen Adressaten angemessenen Aufbereitung naturwissenschaftlicher Sachverhalte oder Methoden. Das Vorwissen des Adressaten und die potenziellen Verständnishürden sind dafür besonders wichtig. Man kann knapp formulieren, dass Argumentation vorrangig der *Überzeugung* dienen soll und Kommunikation dem *Verständnis*. Kommunikation kann also von Argumentation abgegrenzt werden und erfordert ein eigenständiges Modell, auch wenn es Überlappungsbereiche gibt. Kulgemeyer & Schecker (2009) stellen ein Modell für Kommunikation vor, das aus theoretischen Annahmen entwickelt wurde und die Struktur der Fähigkeiten normativ beschreibt, die notwendig sind, um physikalisch zu kommunizieren. Das Modell besteht aus drei grundlegenden Dimensionen: *Aspekt, Perspektive*

und *kognitiver Beiwert*, die im Folgenden kurz beschrieben werden (für Details s. Kulgemeyer & Schecker, 2009).

Die beiden Dimensionen *Perspektive* und *Aspekt* beziehen sich auf ein konstruktivistisches Kommunikationsmodell (Mehrten, 1995; Rusch, 1999). Der grundlegende Gedanke dieses Modells ist, dass drei konstituierenden Bestandteile in einem sich gegenseitig beeinflussenden Verhältnis zueinander stehen (siehe Abbildung 1):

- *Kommunikator* (derjenige, der Information übermitteln möchte),
- *Adressat* (derjenige, an den die Kommunikationsbemühungen gerichtet sind) und
- *Sachinhalt* (der Zusammenhang, über den kommuniziert werden soll).

Dabei ist der Adressat nicht als passiver Rezipient von Information zu sehen, sondern selbst aktiv am Prozess der Kommunika-

tion beteiligt: Er nimmt die Kommunikationsbemühungen des Kommunikators als ein Angebot wahr, das er akzeptieren oder auch ablehnen bzw. ignorieren kann. Die Gelingensbedingung erfolgreicher Kommunikation bemisst sich dabei am Verständnis der zu kommunizierenden Information durch den Adressaten. Der Kommunikator muss das Angebot möglichst attraktiv gestalten, damit der Adressat gewillt ist, dieses auch anzunehmen.

2.1 Dimension *Perspektive*

Der Kommunikator muss also den zu kommunizierenden Inhalt sowohl sachlich angemessen als auch attraktiv für den Adressaten gestalten. Diese beiden unterschiedlichen Sichtweisen auf die Gestaltung des Kommunikationsinhalts, die *adressatengemäße* und die *sachgerechte*, bilden im Kompetenzmodell die beiden Komponenten der Dimension *Perspek-*

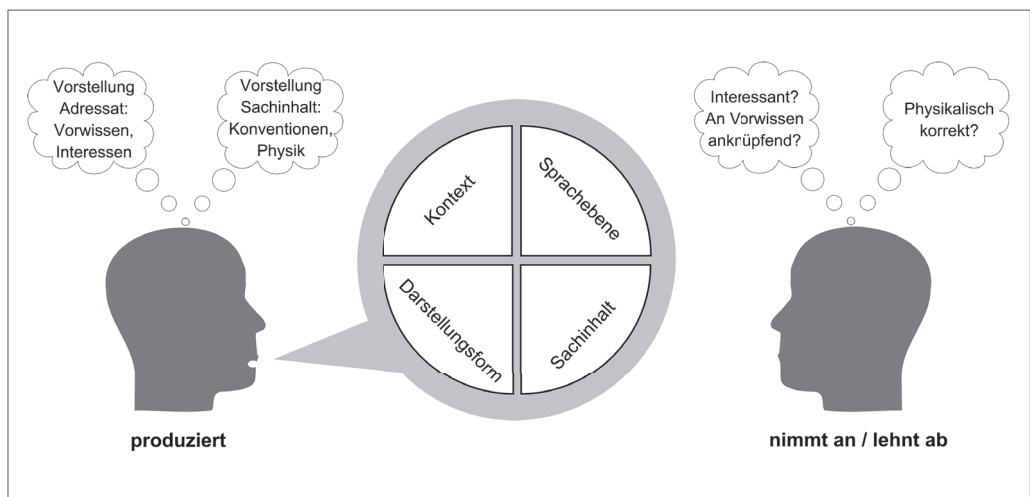


Abb. 1: Konstruktivistisches Modell physikalischer Kommunikation nach Kulgemeyer (2010).

Tab. 1: Beispiele für Indikatorfragen zur Bestimmung des kognitiven Beiwerts

PROZESS	INDIKATORFRAGE
Wahrnehmung des Kommunikationsrahmen	Wird ein Adressat in der Aufgabe benannt oder muss eine Kommunikationssituation betrachtet werden?
Linearisierung	Wird die Produktion mindestens eines Satzes verlangt?
	Wird die Produktion einer grafischen Darstellungsform verlangt?

tive. Aus sachgerechter Sicht ist z. B. zu meist die Darstellung eines Sachverhalts in Fachsprache anzustreben, doch für einen Laien als Adressaten ist dies oftmals keine angemessene Form der Präsentation (Kulgemeyer & Schecker, 2009).

2.2 Dimension *Aspekt*

Nach dem Modell von Kulgemeyer & Schecker (2009) hat der Kommunikator vier grundlegende Variablen bzw. *Aspekte*, die er in einer Kommunikationssituation sachgerecht und adressatengemäß verändern kann. Dies sind nach Kulgemeyer (2010) am Beispiel der Kommunikation mit einem fachlichen Laien über Farbenentstehung:

1. der Teilbereich des *Sachinhalts*, der zur Kommunikation ausgewählt wird – z. B. Dispersion,
2. der *Kontext*, der zur Illustration verwendet wird – z. B. ein Regenbogen,
3. die grafische *Darstellungsform* – z. B. das Foto eines Regenbogens und
4. die Sprachebene, d. h. der sprachliche *Code* – z. B. Alltagssprache.

Diese Variablen bilden die vier Komponenten der Dimension *Aspekt* des Kompetenzmodells.

2.3 Dimension *kognitiver Beiwert*

Zu jedem Kompetenzmodell soll nach Klieme et al. (2003) eine Dimension gehören, die Kompetenzstufen beschreiben kann. Die Dimension *kognitiver Beiwert* soll diese Graduierung leisten. Die Annahme besteht darin, dass die Anzahl der kognitiven Prozesse, die zum Lösen einer Testaufgabe vertieft verarbeitet werden müssen, ein Maß für die Schwierigkeit des Items darstellt: je mehr Prozesse vertieft behandelt werden, desto schwieriger ist das Item. Kulgemeyer & Schecker (2009) haben eine Anzahl von Indikatorfragen formuliert, um jeder Testaufgabe zur physikalischen Kommunikationskompetenz einen kognitiven Beiwert zuzuweisen (Beispiele in Tabelle 1). Diese Fragen orientieren sich an den Prozessen, die zur Sprachproduktion notwendigerweise vorgenommen werden müssen. Die theoretische Grundlage bildet dabei Dietrich (2007), der die psycholinguistischen Erkenntnisse zur Sprachproduktion beschreibt. Bei der Berechnung des kognitiven Beiwerts werden jedoch nur die Prozesse berücksichtigt, die mit fachlicher Information umgehen. Z. B. sollten sich allgemeine Prozesse wie zur lautlichen Äußerung nicht zwischen verschiedenen Domänen differenzieren lassen; Prozesse, die jedoch auf das Lexikon

zugreifen, schon. Es ergeben sich elf Indikatorfragen, die jeweils positiv oder negativ beantwortet werden können. Bei jeder positiven Antwort steigt der kognitive Beiwert um einen Punkt, sodass er maximal Elf betragen kann. Zu manchen Prozessen gehören zwei Indikatorfragen, da sich diese für die Produktion von grafischen Darstellungsformen und linearen Texten unterscheiden.

2.4 Das normative Modell

Das resultierende normative Modell (Abbildung 2) besteht aus zwei Dimensionen,

die die Struktur des Kompetenzbereichs abbilden sollen (*Perspektive* und *Aspekt* eines Kommunikationsaktes) und einer Dimension, die Kompetenzstufen beschreibt (*kognitiver Beiwert*). Um der Forderung von Klieme et al. (2003) gerecht zu werden, dass jedes Kompetenzmodell eine stufende Dimension haben sollte, ist es notwendig, dass die Komponenten dieser der Dimension mindestens eine ordinale Skala bilden. Die Aufgaben wurden gezielt für die einzelnen Zellen des Modells passend neu entwickelt. In Abbildung 3 ist eine Beispielaufgabe für die Komponenten Code/Text und sachgerecht bei einem kognitiven Beiwert von 2 zu finden.

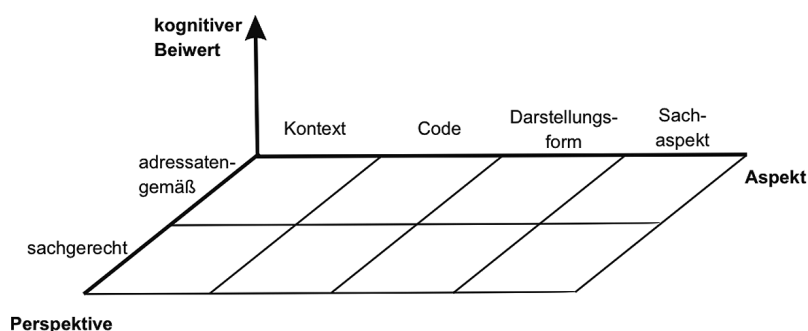


Abb. 2: Das resultierende Kompetenzmodell (aus Kulgemeyer & Schecker (2009)).

In der Tabelle findest Du drei Aussagen, die ein Physiker getroffen hat. Wenn er mit seinen Fachkollegen spricht, dann formuliert er allerdings anders als wenn er mit seiner Familie redet. Mit seinen Fachkollegen spricht er Fachsprache, mit seiner Familie Alltagssprache. **Entscheide, welche dieser Aussagen in Alltagssprache und welche in Fachsprache formuliert sind!**

	Alltags-sprache	Fach-sprache
Um den Energieverbrauch zu verringern, muss jeder für sich Anstrengungen unternehmen.	⊗	○
Körper gleicher Masse können mehr Energie speichern, wenn sie eine höhere spezifische Wärmekapazität haben.	○	⊗
Energie bleibt erhalten, sie kann nur von einer Energieform, z. B. kinetischer Energie, in eine andere umgewandelt werden.	○	⊗

Abb. 3: Musteraufgabe (inkl. Lösung) für die Komponente „Code“ der Dimension „Aspekt“ und die Komponente „sachgerecht“ der Dimension „Perspektive“ (aus Kulgemeyer, 2010, 354).

3 Forschungsfragen

Im vorliegenden Beitrag werden Ergebnisse zu den folgenden Forschungsfragen dargestellt:

1. Lässt sich ein valides, reliables und objektives Testverfahren für physikalische Kommunikationskompetenz entwickeln?
2. Entspricht die empirisch vorgefundene Struktur der Fähigkeiten dem normativen Modell? Wie muss das Modell weiterentwickelt werden?

In der zugrundeliegenden Studie (Kulgemeyer, 2010) kamen quantitative und qualitative Testinstrumente zum Einsatz. Der vorliegende Beitrag fokussiert auf die Ergebnisse eines schriftlichen Testverfahrens für physikalische Kommunikationskompetenz. Der Test enthält theoriegeleitet entwickelte Aufgaben, die in jeder Dimension des Kompetenzmodells jeweils genau eine Komponente ansprechen (siehe auch Abbildung 3). Die theoriegeleitete Itementwicklung ist in Kulgemeyer & Schecker (2009) dargestellt. Bei der Itementwicklung wurde unter anderem darauf geachtet, dass für die Komponente jeder Dimension die Verteilung der Items auf die anderen Dimensionen keine erhebliche Varianz aufweist. Damit soll vermieden werden, dass Varianz in anderen Dimensionen die Analysen einer Dimension des Modells unzulässig konfundiert (Kulgemeyer, 2010, 91–94). Aus testökonomischen Gründen war es allerdings nicht möglich, einen Test zu entwickeln, der jede Zelle des Kompetenzmodells mit

mehr als fünf Items besetzt, wie es beispielsweise Einhaus (2007) fordert – daraus würde eine Testlänge von wenigstens 120 Items resultieren (Kulgemeyer, 2010, 94), die selbst in einem Rotationsdesign, wie es z. B. Einhaus (2007) verwendet, eine erhebliche Anzahl an Probandinnen und Probanden erfordern würde.

Durch die Beschränkung der Testlänge wird also eine Grenze der Überprüfbarkeit des Modells gezogen; kleine Effekte sind demnach nicht auflösbar. Das Ziel der Studie ist jedoch eine grundlegende Untersuchung des Modells auf seine Tauglichkeit, sodass dieser Mangel als vertretbar eingeschätzt wurde. Sie ist also als erste Erkundung des empirisch bis dahin praktisch unbearbeiteten Kompetenzbereichs Kommunikation angelegt worden. Auf eine in die Tiefe gehende Untersuchung einzelner Dimensionen des Kompetenzbereichs und ihrer Komponenten musste verzichtet werden, um zu erforschen, ob der gewählte Ansatz grundsätzlich aussichtsreich zur Behandlung von Kommunikationskompetenz ist. Folgestudien können insbesondere vom untersuchten Methodenrepertoire profitieren. Die in dieser Studie gefundenen ersten Ergebnisse sollen dann als Hypothesen behandelt und weitergehend überprüft werden.

Die beiden Forschungsfragen 1 und 2 erfordern unterschiedliche methodische Ansätze. In Kapitel 4 werden das Vorgehen und die Ergebnisse bei der Testentwicklung dargestellt (Forschungsfrage 1), in Kapitel 5 die Methodik der Analyse der Kompetenzmodellstruktur sowie die Ergebnisse der entsprechenden Analysen (Forschungsfrage 2). Bei bei-

den Forschungsfragen sind angesichts der gewählten Breite der Erkundung von naturwissenschaftsbezogener Kommunikationskompetenz Einschränkungen bezüglich einer Verallgemeinerbarkeit der Ergebnisse vorzunehmen. Dies wird in der Darstellung der Ergebnisse diskutiert (Abschnitt 5 und Abschnitt 6).

4 Forschungsfrage 1: Entwicklung eines Testinstruments

Der schriftliche Test wurde in einem vier-schrittigen Verfahren entwickelt:

1. Theoriegeleitete Itementwicklung auf Basis des normativen Kompetenzmodells
2. Präpilotierung der Items (erste Einschätzung, ob eine valide Messung möglich erscheint)
3. Pilotstudie zur Validierung des Tests sowie zur ersten Einschätzung der Reliabilität
4. Hauptstudie zur Erstellung einer Raschskalierten Version des Tests

Zwischen den einzelnen Schritten wurden die jeweils vorliegenden Ergebnisse dazu verwendet, die Validität der Items und die Reliabilität des Tests zu verbessern. Außerdem wurde im Rahmen der Testentwicklung an den jeweiligen Probanden auch ein Test mit Items zum angrenzenden physikalischen Fachwissen, zur Lesekompetenz, zur Verbalisierungsfähigkeit sowie zur Fähigkeit zur Perspektivenübernahme eingesetzt. Mithilfe dieser Skalen wurde

eine Konstruktvalidierung im Sinne einer Multitrait-Multimethod-Matrix vorgenommen, über die jedoch an anderer Stelle berichtet wird (Kulgemeyer, 2010).

4.1 Präpilotierung

Fricke (2008) konnte in einer Vorläuferstudie bei Studierenden des Sachunterrichts zeigen, dass der verwendete Typus von Kommunikations-Aufgaben prinzipiell valide Messungen erlaubt. Mithilfe von Interviews hat Fricke dazu ermittelt, dass die Überlegungen der Probanden bei der Lösung der Aufgaben im Bereich der Kommunikationskompetenz und nicht im Bereich des Fachwissens liegen. Dazu hat sie Probandinnen und Probanden unmittelbar nach Ausfüllen des schriftlichen Tests interviewt und die Gründe für Aufgabenlösungen erfragt. Es zeigte sich, dass zumeist Gründe genannt wurden, die im Sinne des Aufgabencharakters sind, z. B. Überlegungen zu möglichen Adressaten und nicht das eigene Verständnis fachlicher Inhalte.

In der *Präpilotierung* wurden zwei Testhefte an 62 Schülerinnen und Schülern des Klassenstufen 9 bis 12 (48 von Gymnasien, 14 von einer Hauptschule) erprobt, um ein möglichst breites Leistungsvermögen einzubeziehen. Auf Basis der Ergebnisse wurde der Test wie folgt überarbeitet:

1. Items geringer Trennschärfe (d. h. Trennschärfe $< 0,3$; vgl. dazu Weise (1975)) wurden überarbeitet oder gestrichen.
2. Zudem wurde darauf geachtet, eine breite Streuung an Lösungshäufigkeiten

zu erreichen, sodass für spätere Versionen des Tests Boden- und Deckeneffekte unwahrscheinlicher werden.

3. Ein weiteres Kriterium war, dass das Kompetenzmodell weiterhin vollständig von den Items abgedeckt werden sollte; bei der Streichung von Items wurde also darauf geachtet, möglichst kein Ungleichgewicht bei der Besetzung der Zellen zu erzeugen.

4.2 Pilotstudie mit Fokus auf Validität der Items

Für die *Pilotstudie* wurden die nach der Präpilotierung ausgewählten Items ($N=21$) in ein Testheft zusammengeführt, das innerhalb von 30 Minuten bearbeitet werden kann. Der Test wurde dann an 91 Schülerinnen und Schülern der intendierten Zielgruppe (Zehntklässler an Gymnasien) erprobt. Um die Validität des Tests zu erhöhen, wurden in der Pilotstudie zunächst die Stammtexte der Items aufgrund von Textverständlichkeitskriterien überarbeitet (Kulgemeyer, 2009).

Außerdem wurde eine Interviewstudie durchgeführt, um zu analysieren, ob die Probandinnen und Probanden die Items im Sinne der von den Aufgabenentwicklern intendierten Bedeutung verstehen. Dazu wurden zwei Methoden verwendet: Stimulated-Recall-Interviews ($N=14$) und Beobachtungen des „lauten Denkens“ ($N=17$) bei der Lösung der Aufgaben. Die Stimulated-Recall-Interviews erwiesen sich dabei als weit besser auswertbar als das laute Denken. Es zeigte sich, dass es ein großes Hindernis sein kann, Gedan-

kenprozesse zu verbalisieren, die gewohnterweise still vorgenommen werden (siehe dazu auch Johnston, Bottsford-Miller & Thompson, 2006). Bei den Stimulated-Recall-Interviews wurden zufällig ausgewählten Schülerinnen und Schülern direkt nach der schriftlichen Bearbeitung ihre Testlösungen wieder vorgelegt. Sie wurden dann befragt, wie sie zur Lösung gekommen seien und welche Fähigkeiten oder welches Wissen man ihrer Meinung nach benötige, um die korrekte Lösung zu finden. Zudem wurden ihnen die Musterlösung vorgestellt, wenn ihre Lösung davon abwich. Es ergaben sich für fünf Items Anregungen für Überarbeitungen, insbesondere in der Anpassung von Itemformulierungen bei abweichenden Interpretationen durch die Probanden. Die Auswertung des Stimulated-Recall wurde kategorienbasiert vorgenommen. Es wurden zunächst die geäußerten Überlegungen für im Sinne der Musterlösungen korrekte Lösungen und im Sinne der Musterlösungen falsche Lösungen gesammelt. Dann wurde das Material nach stringenten Überlegungsgängen analysiert, die in sich schlüssig sind, aber zu falschen Lösungen führen. Des Weiteren wurden die Überlegungsgänge, die in sich im Sinne des Validitätsbereichs des Tests falsch sind, aber dennoch zur richtigen Lösung führen, markiert. Items, bei denen diese beiden Arten von Lösungsbegründungen auftraten, wurden überarbeitet. Außerdem wurde analysiert, welche Fähigkeiten oder Kenntnisse die Probandinnen und Probanden als Grundlage korrekter Lösungen vermuten. Items wurden dann einem Überarbeitungsprozess unterzogen, wenn diese Anforderungen nicht

im Sinne von kommunikativer Kompetenz erkannt wurden. Die genauen Ergebnisse sind in Kulgemeyer (2010) dargestellt. Ebenso wie bei Fricke (2008) zeigte sich, dass die Items valide Messungen erlauben und sich die geäußerten Gedanken bei der Lösung auf die Komponenten von Kommunikationskompetenz bezogen, die der Aufgabenentwicklung zugrunde lagen. Die fünf in den Interviews problematischen Items wurden überarbeitet. Ein Beispiel für eine solche Überarbeitung bezieht sich auf das Beispielitem (Abbildung 3), bei dem die Probanden mehrere Sätze darauf einschätzen sollen, ob sie der Fachsprache entstammen oder nicht. Dabei zeigte sich, dass sie dazu tendieren, Sätze, die ausschließlich aus Worten bestehen, die sie kennen, nicht als Fachsprache zu bezeichnen – mit der Begründung, dass Fachsprache für sie selbst zu komplex und unverständlich sei. Eine mögliche Erklärung dafür könnte sein, dass dies aus einem niedrigen physikalischen Selbstkonzept resultiert: die eigene Kenntnis von Vokabeln schließt automatisch deren Fachsprachlichkeit aus. Der Grund konnte jedoch auf Basis der Daten nicht weitergehend untersucht werden; eine Interviewstudie zur Sicht von Schülerinnen und Schülern auf Fachsprache wäre eine lohnende Anschlussstudie. Eine Umformulierung konnte dieses Problem für die Testentwicklung beseitigen: aus „Energie bleibt erhalten, sie kann nur von einer Energieform in eine andere umgewandelt werden“ wurde „Energie bleibt erhalten, sie kann nur von einer Energieform, z. B. kinetischer Energie, in eine andere umgewandelt werden“. Die Fachsprachlichkeit wird hier durch das Hinzufügen einer Fachvo-

kabel betont. Dadurch werden die Differenzen in den formulierten Sätzen größer, dies sollte die Validität der Aufgaben erhöhen. Selbstverständlich sind Fachvokabeln nicht das einzige Kennzeichen von Fachsprache. Doch bei der Konstruktion der Testaufgaben waren sie als Merkmal relativ einfach variierbar, während die Variation textgrammatischer oder syntaktischer Merkmale erst durch ein Expertenrating validierbar wäre.

Auf Basis der quantitativen Ergebnisse wurden zudem Aufgaben mit geringer Trennschärfe (wiederum $< 0,3$) gestrichen, sofern dadurch nicht die vollständige Abdeckung des Kompetenzmodells mit Items verhindert wurde. Es konnte zudem quantitativ gezeigt werden, dass eine Korrelation zwischen dem Gesamtscore des Kommunikationskompetenztests und der letzten Physiknote vorliegt ($r = -0,248^*$, Korrelation nach Pearson signifikant auf dem Niveau von 0,05). Dies ist ein erster Hinweis auf Konstruktvalidität (konvergente Validität). Es ist allerdings zu betonen, dass die Physiknote, die aus einem konventionellen Physikunterricht resultiert, vermutlich in erster Linie anhand von Leistungen im Fachwissen und der Erkenntnisgewinnung bestimmt wird. Dieses Ergebnis ist also mit deutlichen Einschränkungen versehen.

4.3 Hauptstudie und Rasch-Skalierung des Testinstruments

In der *Hauptstudie* wurde der überarbeitete Test an 241 Probanden aus 13 Physikkursen des 10. Jahrgangs an Gymnasien

durchgeführt. Dabei erwies sich ein Kurs als nicht auswertbar, da die Testdurchführung durch einen ausgetragenen Konflikt zwischen der Lehrkraft und den Schülerinnen und Schülern beeinträchtigt wurde. Es konnten somit letztlich 216 Probanden berücksichtigt werden. Auf Basis dieser Stichprobe wurde der Test zur physikalischen Kommunikationskompetenz (genannt *PhyKo*) skaliert. Dabei konnte für den Test (19 Items) sowohl die Itemhomogenität (mit den üblichen Kriterien Infit-MnSq im Intervall 0,8 bis 1,2 bzw. Abweichung $|T| < 2,0$, siehe z.B. Bond & Fox, 2001; Einhaus, 2007) als auch die Personenhomogenität (mithilfe eines grafischen Modelltests) nachgewiesen werden. Ein eindimensionales Rasch-Modell erwies sich im Rahmen dieser Kriterien als ausreichend, um die Testlösungen aller Items zur Kommunikationskompetenz zu beschreiben.

Das Resultat der Hauptstudie ist also ein Rasch-skaliertes Test, der als ausreichend valide, reliabel und objektiv bezeichnet werden kann. Es konnte weder ein Boden- noch ein Deckeneffekt festgestellt werden, die Testergebnisse in den Personenfähigkeiten spannen von 356 bis 714 TIMSS-Punkten. Dazu wurden die Personenparameter aus dem Rasch-Modell aus Anschaulichkeitsgründen linear auf die TIMSS-Skala transformiert (Mittelwert: 500 TIMSS-Punkte, Standardabweichung: 100 TIMSS-Punkte). Zu den drei Hauptgütekriterien wurde folgendes festgestellt:

Validität: Zur Gewährleistung der Validität wurde sichergestellt, dass der Test das Kompetenzmodell ausreichend ab-

deckt (Inhaltsvalidität). Die Dimension des kognitiven Beiwerts konnte jedoch aus testpraktikablen Gründen nicht systematisch variiert werden, da der Testumfang ansonsten zu groß gewesen wäre; dies schränkt die Generalisierung der Ergebnisse ein. Es konnte in den Stimulated-Recall-Interviews sowie in den Beobachtungen des lauten Denkens nachgewiesen werden, dass die Probanden zur Lösung der Items auf Gedankengänge im Sinne der Kommunikationskompetenz zurückgreifen (Inhaltsvalidität). Die Perspektive der Itementwickler auf die Items wurde durch Weiterentwicklungen mit der Perspektive der Probanden abgeglichen. In Kulgemeyer (2010) findet sich auch eine Konstruktvalidierung durch eine Multi-trait-Multimethod-Matrix.

Reliabilität: Der *PhyKo*-Test hat sich als item- und personenhomogen erwiesen. Der *PhyKo*-Test erreicht zudem ein Cronbachs α von 0,61. Für ein erstmals theoretisch operationalisiertes Modell wird dieser Wert als befriedigend erachtet. Er gibt jedoch auch einen ersten Hinweis darauf, dass Kommunikationskompetenz eine komplexe Struktur aufweist – auch wenn eine testweise durchgeführte Faktorenanalyse nicht zu einer weiteren Untergliederung der Skala führte.

Objektivität: Der Test wurde anonymisiert und standardisiert durchgeführt sowie an einer Stichprobe erprobt, die homogen ist und der Zielgruppe der Bildungsstandards entspricht (Durchführungsobjektivität). Zudem ist sie vom Umfang her vergleichbar mit ähn-

lichen Studien zur Kompetenzmodellierung (z. B. Einhaus 2007, Schmidt 2008 oder Bernholt, Parchmann & Commons 2009). Das Manual zur Testauswertung stellt befriedigende Werte zur Interraterreliabilität sicher. Sowohl die Übereinstimmung bei der Beurteilung der einzelnen Items war mit einem Cohens $\kappa > 0,8$ als auch die Korrelation der Skalenergebnisse beider Rater mit einer Korrelation von $r = 0,931$ im sehr guten Bereich der Auswertungsobjektivität.

5 Forschungsfrage 2: Analysen der Kompetenzmodellstruktur

5.1 Methodik der empirische Überprüfung des Kompetenzmodells

Um das Kompetenzmodell auf Basis der empirischen Daten weiter zu entwickeln, wird in dieser Studie ein Verfahren auf Basis mehrdimensionaler Rasch-Modelle genutzt, das in sehr ähnlicher Form bereits bei Schmidt (2008) und Einhaus (2007) verwendet wurde. Einhaus (2007) hat dieses heuristische Verfahren mithilfe umfassender Simulationsstudien erprobt und validiert. Das Verfahren hat sich dabei als verlässlich erwiesen und bietet gute Möglichkeiten der Interpretation von Kompetenzstrukturen.

Die Ausgangsannahme besteht darin, dass die Komponenten der Dimension eines Kompetenzmodells als unterscheidbare Fähigkeiten aufgefasst werden können. Diese Hypothese wird mithilfe eines Kriteriensystems untersucht. An dieser Stelle ergeben sich leider durch die in der Rasch-

Analyse und der Kompetenzmodellierung unterschiedlich verwendete Terminologie eine Unschärfe: Die Dimensionen des Rasch-Modells sind nicht identisch mit den Dimensionen des Kompetenzmodells. Der Begriff der Dimension ist doppelt gegeben, die Dimensionen des Rasch-Modells entsprechen den Komponenten einer Dimension des Kompetenzmodells. Wenn beispielsweise eine Dimension des Kompetenzmodells aus zwei Komponenten besteht – wie in dieser Studie die Dimension *Perspektive* aus den Komponenten *adressatengemäß* und *sachgerecht* – wird u. a. untersucht, ob ein eindimensionales Rasch-Modell oder ein zweidimensionales Rasch-Modell die Daten besser beschreibt. Einhaus (2007) hat gezeigt, dass insgesamt sechs Kriterien eine Grundlage bilden, um zu entscheiden, ob Fähigkeiten

1. parallel vorliegen, sich also auf eine gemeinsame Grundfähigkeit zurückführen lassen,
2. sich qualitativ grundlegend unterscheiden oder
3. in einem hierarchisch gestuften Verhältnis zueinander stehen.

Die Konsequenz für die Weiterentwicklung des Kompetenzmodells wäre im ersten Falle, dass die Komponenten psychometrisch gerechtfertigt zusammengefasst werden können. Im zweiten Falle sollten sie als getrennten Komponenten bestehen bleiben. Wenn alle Komponenten einer Dimension sich als hierarchisch gestufte Fähigkeiten erweisen, so verfügt diese Dimension über Kompetenzstufen. Dabei werden unter anderem Personen-

parameter (als Maß für die Fähigkeitsausprägung einer Person) und Itemparameter (als Maß für die Schwierigkeit eines Items) verschiedener Modelle miteinander verglichen. Die Kriterien, um den Zusammenhang der Fähigkeiten zu beurteilen, sind in Abbildung 4 aufgeführt. Sie bilden drei Gruppen:

K1 (Modellgüte): In dieser Kriteriengruppe wird der Vergleich der Modellgüte der Rasch-Modelle mit dem *Bayes Information Criterion* (BIC) vorgenommen (Herleitung siehe Rost (2004) und Schmidt (2008)). Alternative Gütekriterien findet man z. B. mit dem AIC und dem CAIC (Rost, 2004, S. 342-344). Alle beruhen jedoch auf dem Grundgedanken, die Einfachheit eines Modells über die Parameteranzahl mit der Likelihoodfunktion des Modells – und somit seiner Übereinstimmung mit den empirischen Daten – zu verrechnen.

K2 (Itemparameter): In dieser Kriteriengruppe werden die Itemparameter des eindimensionalen Modells mit der Zuordnung der Items zu den Komponenten verglichen. Gehört z. B. die Hälfte der Items zur Komponente A und die Hälfte zur Komponente B, so wird zunächst die Korrelation der Komponentenanzuordnung zu den Itemparametern (Kriterium K2a) und dann eine Varianzaufklärung der Itemparameter durch die Komponenten berechnet (Kriterium K2b). Wenn die Komponenten die Schwierigkeit der Items und somit die Itemparameter beeinflussen, ergeben sich hier weiterführende Ergebnisse. Methodisch umgesetzt wird dies durch

Spearman's Rangkorrelation und univariate Varianzaufklärungen.

K3 (Personenparameter): In dieser Kriteriengruppe werden einerseits die Personenparameter des eindimensionalen und des mehrdimensionalen Rasch-Modells miteinander verglichen und andererseits Vergleiche der Personenparameter unterschiedlicher Dimensionen des mehrdimensionalen Modells vorgenommen. Dabei wird zunächst eine Varianzaufklärung der Personenparameter des eindimensionalen Modells durch die Personenparameter des mehrdimensionalen Modells vorgenommen (Kriterium K3a). Wenn die Dimensionen des mehrdimensionalen Modells nicht voneinander unterscheidbar sind, so werden für alle Dimensionen hohe Varianzaufklärungen erwartet, dies kann jedoch je nach Komponentenrelation (parallel, unabhängig, gestuft) variieren. Des Weiteren werden in dieser Gruppe Korrelationen zwischen den Personenparametern des mehrdimensionalen Rasch-Modells miteinander verglichen (Kriterium K3b). Außerdem werden die einzelnen Fähigkeiten in einer Art grafischem Test mit Blasendiagrammen auf ihre Beziehung zueinander getestet (Kriterium K3c). Dabei werden die Personenparameter jeweils zweier Dimensionen des mehrdimensionalen Modells in Leistungsquartile unterteilt. Es ergibt sich eine Kreuztabelle, in der für jede Quartilskombination die Anzahl der Fälle dargestellt wird (z. B. Anzahl der Fälle, die in beiden Komponenten im 1. Leistungsquartil liegen). Im Blasendiagramm wird diese Kreuztabelle veran-

schaulich, indem die Anzahl der Fälle durch die Blasengröße dargestellt wird und die Leistungsquartile beider Komponenten auf Abszisse bzw. Ordinate gelegt werden. Wie Einhaus (2007) gezeigt hat, ergeben sich je nach der Relation zwischen den Komponenten zueinander charakteristische Blasenordnungen.

Es soll hier betont werden, dass dieses Kriteriensystem heuristischen Charakter hat. Nicht alle Kriterien sind unabhängig von-

einander: K1b liefert ähnliche Informationen wie K3b und K2b ähnliche wie K2a. Dennoch ergibt sich aus der Gesamtschau der Kriterien ein Bild, das begründete Hinweise für die Frage nach Unabhängigkeit, Parallelität oder Stufung der Fähigkeiten in den Komponenten liefert. Eindeutige Ergebnisse über alle sieben Kriterien hinweg sind aus fachdidaktischer Sicht für eine komplexe und bisher unerschlossene Kompetenz wie physikbezogene Kommunikationsfähigkeit jedoch nicht zu erwarten.

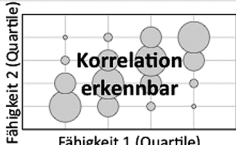
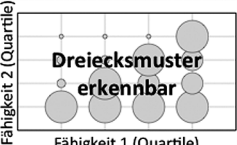
Die Komponenten beschreiben...			
	...qualitativ unterschiedliche Fähigkeiten (QUF)	...parallele Fähigkeiten (PAR)	...gestufte Fähigkeiten (GES)
K1	<i>Modell mit geringstem BIC</i>		
	mehrdimensional	eindimensional	eindimensional
K2a	<i>Korrelation der Itemparameter des eindimensionalen Modells mit den Komponenten</i>		
	keine	keine	hoch
K2b	<i>Varianzaufklärung der Itemparameter des eindimensionalen Modells durch die Komponenten</i>		
	niedrig ($< 10\%$)	niedrig ($< 10\%$)	hoch ($> 10\%$)
K3a	<i>Varianzaufklärung der Personenparameter des eindimensionalen Modells durch die Personenparameter des mehrdimensionalen Modells</i>		
	niedrig für jede Dimension ($< 15\%$)	hoch für jede Dimension ($> 15\%$)	für Dimensionen unterschiedlich ($> 15\%$; Differenz $> 10\%$)
K3b	<i>Korrelation zwischen Personenparametern der Dimensionen</i>		
	keine ($< 0,1$)	hoch ($> 0,3$)	niedrig ($0,1 < r < 0,3$)
K3c	<i>Verteilung der Personen in den Quartilsdimensionen</i>		
			

Abb. 4: Kriterien K1 bis K3c zur Differenzierung des Verhältnisses von Komponenten eines Kompetenzmodells nach Einhaus (2007).

5.2 Ergebnisse der Analysen der Kompetenzmodellstruktur

Die Ergebnisse der Untersuchung der Struktur physikalischer Kommunikationskompetenz werden in diesem Beitrag gegliedert nach den Dimensionen des Kompetenzmodells dargestellt. Dabei wird die Stichprobe von 216 Schülerinnen und Schülern der Klasse 10 aus der Hauptstudie der Testentwicklung verwendet. Um im vorigen Abschnitt beschriebene Verfahren zu illustrieren, geschieht dies bei der Analyse der ersten Dimension (Perspektive) ausführlicher. In allen Analysen wird auf den entwickelten PhyKo-Test Bezug genommen. Der PhyKo-Test besteht aus 19 Items. Mit dieser Anzahl ist es nicht möglich, die Komponenten aller Dimensionen systematisch und mit ausreichender Anzahl gegeneinander zu variieren. Die Dimensionen sollten jedoch unabhängig voneinander sein, sodass dies auch nicht zwingend notwendig ist. Dennoch sind die Ergebnisse unter dieser Einschränkung zu verstehen. Es handelt sich um die erste Erkundung eines empirisch bislang nicht bearbeiteten Kompetenzbereichs, bei der die prinzipielle empirische Umsetzbarkeit zu überprüfen werden sollte. Eine genaue Aufführung der Anzahl der Items mit Zuordnung zu den Zellen des Kompetenzmodells ist in der Diskussion der Ergebnisse (Abschnitt 6) der vorliegenden Arbeit aufgeführt.

5.2.1 Analyse der Dimension *Perspektive*

Die Komponenten der Dimension *Perspektive* (sachgerecht und adressatenge-

mäß kommunizieren) scheinen gestuft zueinander zu sein, d. h. zu ihrer Bearbeitung sind Fähigkeiten notwendig, die in einem hierarchischen Verhältnis zueinander stehen. Dabei steht *adressatengemäß* auf der höheren Stufe: tendenziell ist also die Fähigkeit, sachgerecht kommunizieren zu können, die Voraussetzung für adressatengemäßes Kommunizieren. Nachvollziehbar ist dieser Schluss insbesondere in Abbildung 5. Hier ist zu erkennen, dass erst das Erreichen eines hohen Leistungsquartils im sachgerechten Kommunizieren ermöglicht, dass auch im adressatengemäßen Kommunizieren ein hohes Quartil erreicht wird – umgekehrt gilt dies jedoch nicht.

Zur Analyse wurde ein eindimensionales Rasch-Modell mit einem zweidimensionalen verglichen. Für das erste Modell wurden alle Items in eine statistische Dimension zusammengefasst, für das zweite die jeweils zu den beiden Komponenten adressatengemäß und sachgerecht gehörenden Items in zwei getrennte Dimensionen (sieben Items zum sachgerecht Kommunizieren und zwölf Items zum adressatengemäßen). Die Ergebnisse der Analysen im Kriteriensystem nach Einhaus (2007) sind in Tabelle 2 dargestellt. Ein eindimensionales Rasch-Modell beschreibt die Daten demnach besser als ein zweidimensionales (Kriterium K1). Werden die Itemparameter aus dem eindimensionalen Rasch-Modell, also der Maßstab für die Itemschwierigkeit, mit der Zuordnung der Items zu den beiden Komponenten sachgerecht und adressatengemäß korreliert (Kriterium K2a), so ergibt sich eine signifikante Rangkorrelation nach

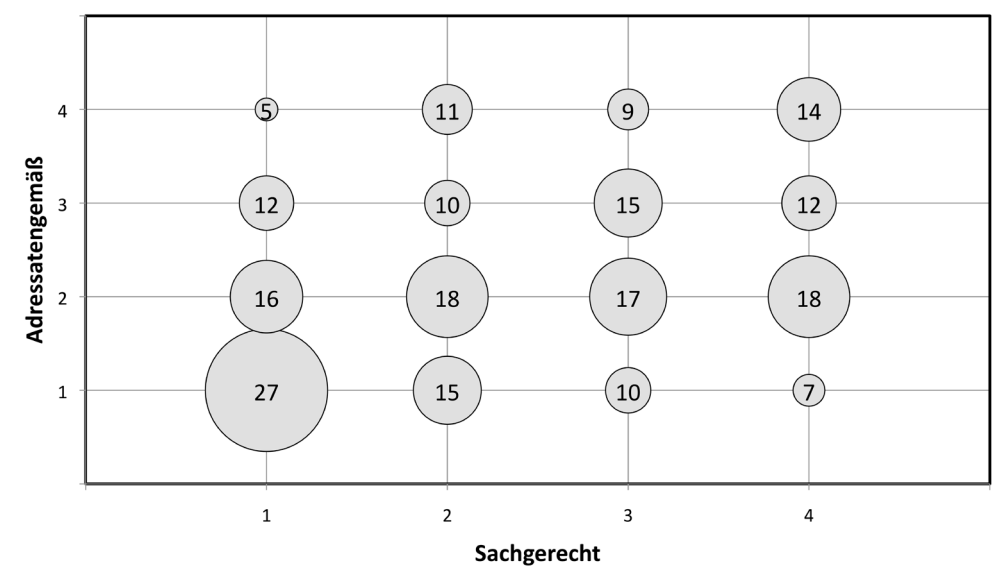


Abb. 5: Besetzung der Leistungsquartile für adressatengemäßes und sachgerechtes Kommunizieren.

Tab. 2: Ergebnisse der kriterienorientierten Analyse der Dimension *Perspektive* mit einem zweidimensionalen Rasch-Modell

Dim. 1: sachgerecht; Dim. 2: adressatengemäß		
KRITERIUM	ERGEBNIS	FOLGERUNG
K1 (BIC)	1dim: 4962,648 2dim: 4965,340	parallele Fähigkeiten/ gestufte Fähigkeiten
K2a (Kor. Itemp.)	$r=0,538^*$	gestufte Fähigkeiten
K2b (Var. Itemp.)	29,4%	gestufte Fähigkeiten
K3a (Var. Persp.)	56,4% vs. 73,0%	gestufte Fähigkeiten
K3b (Kor. Persp.)	$r=0,249^{**}$	gestufte Fähigkeiten
K3c (Blasend.)	Abbildung 4	gestufte Fähigkeiten
	Gesamt:	gestufte Fähigkeiten

Spearman von $\rho=0,538^*$, d.h. dass die Itemschwierigkeit von der Zugehörigkeit zu einer der beiden Komponenten abhängt. Bestätigt wird dies durch die Varianzaufklärung (Kriterium K2b), beides spricht für gestufte Fähigkeiten.

Für Kriterium K3a wurde überprüft, inwieweit die Varianz in den Personenparametern des eindimensionalen Modells durch die Varianz der beiden Personenparameter des zweidimensionalen Modells aufgeklärt werden. Dabei ergeben

sich insgesamt hohe, aber voneinander bedeutend abweichende Werte. Dies spricht im Kriteriensystem für gestufte Fähigkeiten. Die Personenparameter sind ein Maß für die Fähigkeit der Personen, d.h. in diesem Falle ihr Leistungsvermögen im Kommunizieren insgesamt (eindimensionales Modell) und in den beiden Perspektiven adressatengemäß und sachgerecht (die beiden Dimensionen des zweidimensionalen Modells). Für Kriterium K3b werden die Personenparameter beider Dimensionen des zweidimensionalen Rasch-Modells miteinander korreliert, also die Fähigkeiten der Personen im adressatengemäßen und sachgerechten Kommunizieren miteinander in Beziehung gesetzt. Dabei ergibt sich eine hochsignifikante, aber geringe Korrelation. Beide Fähigkeitsausprägungen stehen also in einem geringen Zusammenhang. Insgesamt sprechen die Ergebnisse klar für gestufte Fähigkeiten, wobei sachgerechtes Kommunizieren auf der niedrigeren Stufe liegt und somit die Voraussetzung für adressatengemäßes Kommunizieren darstellt.

Für die empirische Weiterentwicklung des Kompetenzmodells kann aus diesen Ergebnissen geschlossen werden, dass die Trennung von adressatengemäßer und sachgerechter Kommunikation in zwei verschiedenen Komponenten einer Dimension sinnvoll ist und beibehalten werden kann. Für beide Komponenten könnten zudem verschiedene Fähigkeiten, die in einem gestuften Verhältnis zueinander stehen, die Grundlage zu sein. Im Abschluss teil dieser Arbeit wird dieses Ergebnis vor dem Hintergrund der beschränkten Itemzahlen weitergehend diskutiert.

5.2.2 Analyse der Dimension *Aspekt*

Die Dimension *Aspekt* des Kompetenzmodells besteht aus vier Komponenten: Kontext, Code, Darstellungsform und Sachaspekt. Die Komponente *Code* wurde mit fünf Items des Tests besetzt, die Komponente *Darstellungsform* ebenfalls und die Komponente *Sachaspekt* mit sieben Items. Lediglich die Komponente *Kontext* wurden nur durch zwei Items angesprochen, die Ergebnisse in diesem Bereich sollen unter dieser deutlichen Einschränkung verstanden werden.

Der Umfang des Tests reicht nicht aus, um ein aussagekräftiges vierdimensionales Rasch-Modell zu berechnen. Zur Analyse der Dimension *Aspekt* wurden deshalb vier zweidimensionale und ein dreidimensionales Rasch-Modell berechnet. Diese Analysen ergeben jedoch ein konsistentes Bild über die Struktur dieser Dimension des Kompetenzmodells. Wiederum sind die geringen Itemzahlen jedoch als Einschränkung dieser Ergebnisse hervor zu heben.

Bei den zweidimensionalen Modellen wurde jeweils eine der vier Komponenten der Dimension *Aspekt* (Kontext, Code, Darstellungsform und Sachaspekt) in einer Dimension den übrigen Komponenten gegenübergestellt. Es ergibt sich dabei, dass sich der Aspekt Kontext von den anderen Aspekten differenzieren lässt – die Fähigkeit, den Aspekt Kontext zu bearbeiten, erweist sich als qualitativ unterschiedlich zur Fähigkeit, die übrigen Aspekte auszuwählen. Die Aspekte Darstellungsform und Code lassen sich allerdings nicht voneinander trennen, die Analysen weisen auf parallele Fähigkeiten hin, die

zur Lösung dementsprechender Aufgaben vorhanden sein müssen. Der Aspekt Sachaspekt ist den Ergebnissen zufolge in einem gestuften Verhältnis zu den anderen Komponenten.

Auf Basis der Ergebnisse der Analyse mithilfe zweidimensionaler Rasch-Modelle wurde ein dreidimensionales Rasch-Modell berechnet, das die Aspekte Sachaspekt und Kontext in getrennten Dimensionen erfasst und die dritte Dimension aus den kombinierten Komponenten Code und Darstellungsform bildet. Dies rechtfertigt sich dadurch, dass Code und Darstellungsform sich nicht trennen zu lassen scheinen, die übrigen Komponenten jedoch schon. Es ergibt sich ein Modell, das die Daten besser beschreibt als das eindimensionale Rasch-Modell.

Kriterium K1: Mit einem Devianzmaß von 4825,400 ergibt sich für das dreidimensionale Rasch-Modell ein BIC von 4959,782. Es gilt also $BIC(1dim) > BIC(3dim)$, die Messwerte passen besser zum dreidimensionalen Modell. Das ist ein Hinweis auf qualitativ unterschiedliche Fähigkeiten.

Kriterium K2a: Die Korrelation zwischen den Itemparametern des eindimensionalen Modells und der Zuordnung der Items zu den Dimensionen des dreidimensionalen Rasch-Modells beträgt $\rho = 0,317$. Dies ist ein Hinweis auf qualitativ unterschiedliche Fähigkeiten oder parallele Fähigkeiten.

Kriterium K2b: Durch die Zuordnung der Items zu den Dimensionen des dreidimensionalen Rasch-Modells werden 15,7 % der Varianz der Itemparameter des eindimensionalen Modells aufgeklärt. Diese mittlere Varianzaufklärung weist auf gestufte Fähigkeiten hin.

Kriterium K3a: Die Aspekte klären einen unterschiedlich hohen Anteil der Varianz auf. Dies spricht für gestufte Fähigkeiten. Dieses Ergebnis steht allerdings wieder unter der Einschränkung einer geringen Itemanzahl für den Aspekt Kontext (siehe Tab. 3).

Kriterium K3b: Die Korrelation nach Pearson zwischen den Personenparametern der beiden Dimensionen ergibt sich wie in Tabelle 4 zu sehen. Dies spricht für gestufte Fähigkeiten zwischen Code/Darstellungsform und Sachaspekt sowie

Tab. 3: Kriterium K3a für die Dimension Aspekt

	KONTEXT	CODE/ DARSTELLUNGSFORM	SACHASPEKT
Aufgeklärte Varianz	8,7 %	66,9 %	55,6 %

Tab. 4: Kriterium K3b die Dimension Aspekt

	KONTEXT	CODE/DARSTELLUNGSFORM
Sachaspekt	0,049	0,294**
Kontext	–	0,041

qualitativ unterschiedliche Fähigkeiten zwischen Kontext und den anderen beiden Dimensionen.

Kriterium K3c: Die Analyse der Blasendiagramm führt lediglich bei der Abgrenzung von Darstellungsform/Code und Sachaspekt zu auswertbaren Ergebnissen. Da die Komponente Kontext mit lediglich zwei Items besetzt ist, können keine Quartile gebildet werden. Abbildung 6 zeigt jedoch, dass Sachaspekt auf einer höheren Kompetenzstufe zu liegen scheint als Darstellungsform/Code (Dreiecksmuster nach Einhaus (2007)).

Die Analysen weisen auch hier darauf hin, dass die Fähigkeit, den richtigen Kontext auszuwählen, sich qualitativ von den anderen beiden Fähigkeiten – Code/Darstellungsform bzw. Sachaspekte auszuwählen – unterscheidet. Die Fähigkeit, den rich-

tigen Sachaspekt auszuwählen, steht in einem gestuften Verhältnis zur Fähigkeit, den richtigen Code bzw. die richtige Darstellungsform zu finden ist. Dies bestätigt sich im dreidimensionalen Rasch-Modell. Dabei erweist sich die Fähigkeit, den richtigen Code bzw. die richtige Darstellungsform zu finden, als Voraussetzung für die Auswahl des Sachaspekts. Dies könnte man wie folgt deuten: Es lässt sich erst dann der richtige Sachaspekt zur Kommunikation als Argument auswählen bzw. erst dann erkennen, wo die Verständnisprobleme eines Adressaten liegen, wenn der Umgang mit Fachsprache und fachlichen grafischen Darstellungsformen möglich ist.

Auch dass sich die Komponenten Code und Darstellungsform nicht trennen lassen, ist ein im Rahmen gängiger Theorie interpretierbares Ergebnis. Schon bei PISA wird der Umgang mit diskontinu-

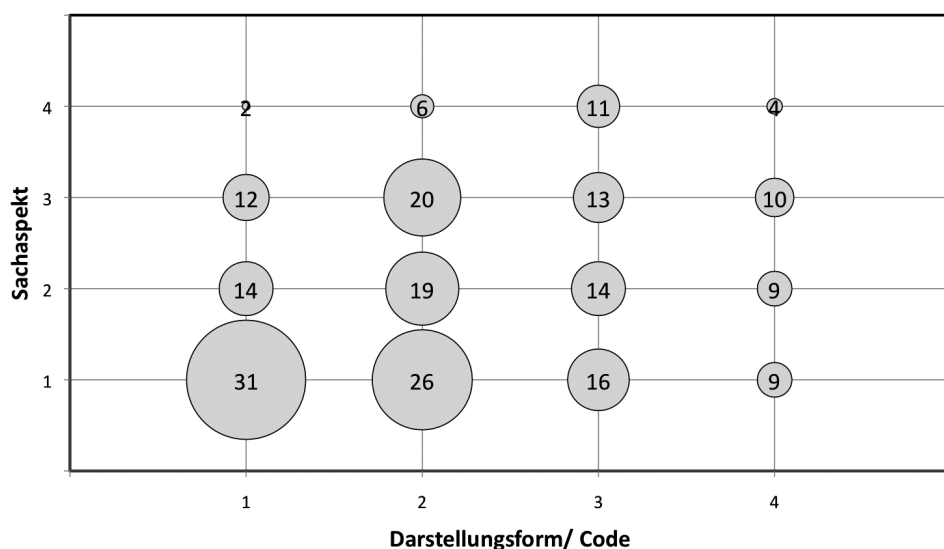


Abb. 6: Besetzung der Leistungsquartile für die Komponente Sachaspekt und die zusammengefassten Komponenten Darstellungsform und Code

ierlichen Texten und kontinuierlichen Texten als gemeinsamer Teil der Lesekompetenz getestet. Im Zusammenhang mit Physikunterricht wird Text oftmals als besondere Darstellungsform aufgefasst, deren Information in grafische Darstellungsformen überführbar ist (Leisen, 2005). Diese Sicht lässt sich nach den hier gefundenen Ergebnissen bestätigen: Die Fähigkeit, mit einem fachlichen Text umzugehen, und die Fähigkeit, eine fachliche grafische Darstellungsform zu verwenden, lassen sich nicht trennen. Sie können als eine zusammenhängende Fähigkeit aufgefasst werden. Gesprochene und geschriebene Sprache sind zwar jeweils für sich ein besonderes Symbolsystem, aber ebenfalls für sich nicht das einzige mögliche für Fachinformation.

In einem weiterführenden Kompetenzmodell lassen sich demzufolge die Komponenten Code und Darstellungsform zusammenfassen, die anderen beiden Komponenten müssen jedoch getrennt bleiben. Die Dimension *Aspekt* besteht demnach nunmehr lediglich aus drei Komponenten: Kontext, Code/ Darstellungsform und Sachaspekt.

5.2.3 Analyse der Dimension *kognitiver Beiwert*

Für die Analysen werden die Items in drei Gruppen des kognitiven Beiwerts sortiert: Gruppe 1 umfasst die aus theoretischer Sicht leichten Items der kognitiven Beiwerte 1 bis 3, Gruppe 2 die mittelschwierigen Items der kognitiven Beiwerte 4 und 5 und Gruppe 3 die schwierigen Items der kognitiven Beiwerte 6 und 7. Die Dimen-

sion *kognitiver Beiwert* besteht demnach aus drei Komponenten, die den Gruppen kognitiver Beiwerte entsprechen. Die Setzung erfolgte aus pragmatischen Gründen, ein Standardsettingverfahren ist nicht vorgenommen worden. Ziel war es, Gruppen zu bilden, die mit einer ausreichenden Anzahl an Items besetzt werden können und gleichzeitig eine ausreichende Auflösung an Kompetenzstufen abbilden. Drei Stufen werden dafür als Mindestanzahl angenommen. In den Bildungsstandards werden ebenfalls drei Anforderungsbereiche ausgewiesen, die eine Stufung implizieren. Bei vier Gruppen von kognitiven Beiwerten wäre die Anzahl der Items pro Stufe deutlich gesunken. Bei drei Gruppen können mindestens fünf Items pro Komponente der Dimension für die Analysen herangezogen werden.

Vor den Analysen mithilfe der Rasch-Modelle wurden in einem ersten Schritt die Itemparameter des eindimensionalen Rasch-Modells mit dem gruppierten kognitiven Beiwert korreliert, damit ein erster Eindruck über die mögliche Stufung der Skala gewonnen werden kann.

Da der kognitive Beiwert als ordinale Skala angenommen wird, wurde dazu der Rangkorrelationskoeffizient nach Spearman berechnet ($\rho = 0,631$). Die Korrelation ist signifikant auf dem 1 %-Niveau und positiv. Dies ist zu erwarten, wenn der kognitive Beiwert ein Maß für die Itemschwierigkeit darstellen soll. Ergänzend dazu wurde eine univariate Varianzanalyse durchgeführt, um den Einfluss des kognitiven Beiwerts auf die Schwierigkeit der Items genauer zu untersuchen.

Die Voraussetzung der Varianzenhomogenität ist dabei gegeben, da der Levene-Test nicht signifikant ist ($p=0,310$). Es zeigt sich ein höchst signifikanter Einfluss des gruppierten kognitiven Beiwerts auf den Itemparameter ($p=0,001$), durch den $R^2=65,2\%$ der Varianz aufgeklärt werden. Der Scheffé-Test zeigt, dass sich die Gruppen 1 und 3 signifikant voneinander unterscheiden, die Gruppen 1 und 2 jedoch nicht. Insgesamt sprechen diese Ergebnisse im eindimensionalen Rasch-Modell dafür, dass der kognitive Beiwert ein ex-ante-Maß für die Itemschwierigkeit ist.

Zur genaueren Klärung dessen wird wiederum das Verfahren nach Einhaus (2007) gewählt. Dazu werden hier drei zweidimensionale Rasch-Modelle mit dem eindimensionalen Modell verglichen (Tabelle 5). In den ersten beiden Modellen wird nur überprüft, ob eine Stufung zwischen Gruppe 1 auf der einen und Gruppe 2 und 3 auf der anderen bzw. Gruppe 1 und 2 auf der einen und Gruppe 3 auf der anderen Seite besteht. Wenn allerdings Gruppe 3 aus besonders schwierigen Items besteht, würde in beiden Fällen eine Stufung nachgewiesen werden, obwohl zwischen den Gruppen 1 und 2 nicht notwendigerweise auch eine Stufe existiert. Deshalb wird im letzten Modell Gruppe 3 gestrichen und nur die Beziehung zwischen den ersten beiden Itemgruppen untersucht. Dies rechtfertigt sich auch daraus, dass sich in der Analyse des eindimensionalen Modells zeigte, dass der Abstand von Gruppe 2 zu Gruppe 3 deutlich größer ist als der zwischen den Gruppen 1 und 2.

Die Analysen im eindimensionalen Rasch-Modell deuten darauf hin, dass die Dimension *kognitiver Beiwert* des Kompetenzmodells zur physikalischen Kommunikation gestufte Fähigkeiten beschreibt. Es gibt jedoch auch Hinweise darauf, dass die Gruppen 1 und 2 des kognitiven Beiwerts nicht sauber voneinander getrennt werden können; der post-hoc-Test der Varianzaufklärung zeigt keine signifikanten Unterschiede. Auch die Trennung der beiden Gruppen im zweidimensionalen Rasch-Modell ist nicht so eindeutig feststellbar wie bei Einbeziehung von Gruppe 3. Dies spricht dafür, dass die Stufe von Gruppe 2 zu Gruppe 3 größer ist als die von Gruppe 1 zu Gruppe 2 – auf einer ordinalen Skala kommt dieser Effekt jedoch nicht zur Geltung. Das Modell für physikalische Kommunikationskompetenz verfügt also über die von Klieme et al. (2003) geforderte Dimension, die Kompetenzstufen beschreibt. Dies kann als im Rahmen der Einschränkungen durch Itemanzahl und Anzahl der Personen als bestätigt angesehen werden.

6 Diskussion und Ausblick

Vorgelegt werden die Ergebnisse einer Studie zur quantitativen Erfassung physikbezogener Kommunikationskompetenz und der inneren Struktur dieser Kompetenz. Das Konstrukt wird breit gefasst und differenziert im Ausgangsmodell zwischen adressaten- und sachbezogenen Perspektiven sowie vier Aspekten eines Kommunikationsaktes (Kontext, Code,

Tab. 5: Ergebnisse der kriterienorientierten Analyse der Dimension *kognitiver Beiwert* mit einem zweidimensionalen Rasch-Modell. Verglichen werden die drei Modelle

Dim. 1: Gruppe 1; Dim. 2: Gruppen 2 + 3			Dim. 1: Gruppen 1 und 2; Dim. 2: Gruppe 3		
KRIT.	ERGEBNIS	FOLGERUNG FÜR FÄHIGKEITEN	KRIT.	ERGEBNIS	FOLGERUNG FÜR FÄHIGKEITEN
K1	1dim: 4962,648 2dim: 4948,651	unterschiedlich	K1	1dim: 4962,648 2dim: 4963,379	parallel/gestuft
K2a	$r=0,478^*$	gestuft	K2a	$r=0,606^*$	gestuft
K2b	21,5 %	gestuft	K2b	60,5 %	gestuft
K3a	44,9 % vs. 72,7 %	gestuft	K3a	89,0 % vs. 35,1 %	gestuft
K3b	$r=0,283^{**}$	gestuft	K3b	$r=0,228^{**}$	gestuft
K3c	Dreieck	gestuft	K3c	Dreieck	gestuft
	Gesamt:	gestufte Fähigkeiten		Gesamt:	gestufte Fähigkeiten

Dim. 1: Gruppe 1; Dim. 2: Gruppe 2		
KRIT.	ERGEBNIS	FOLGERUNG FÜR FÄHIGKEITEN
K1	1dim: 4288,152 2dim: 4270,801	unterschiedlich
K2a	$r=0,399$	gestuft
K2b	15,9 %	gestuft
K3a	2,7 % vs. 1,9 %	unterschiedlich
K3b	$r=0,180^{**}$	gestuft
K3c	Dreieck	gestuft
	Gesamt:	gestufte Fähigkeiten

Darstellungsform, und Sachaspekt). Aus den Analysen der Testdaten lässt sich eine Unterscheidbarkeit der beiden Perspektiven adressatengemäß und sachgerecht

belegen. Die Graduierung von Aufgabenschwierigkeiten lässt sich durch den Parameter „kognitiver Beiwert“ erfassen. Mit der Annahme dreier Kompetenzstufen konnten 65,2 % der Varianz in der Itemschwierigkeit erklärt werden. Fähigkeiten zu den beiden Aspekten Code und Darstellungsform lassen sich empirisch nicht trennen. Bei den Fähigkeiten in den verbleibenden drei Aspekten Kontext, Sachaspekt und Code/Darstellungsform scheint die Fähigkeit zur Wahl eines für die Kommunikation geeigneten Sachverhalts besser zu gelingen, wenn Fähigkeiten zum Umgang mit Fachsprache (Code) und grafischen Darstellungsformen vorliegen. Die Wahl eines geeigneten Erklärungskontextes ist von den beiden anderen Aspekten unabhängig.

Mit den Ergebnissen dieser Studie liegt nun erstmals ein empirisch gestütztes Modell physikalischer Kommunikation vor, das in Anschlussstudien Verwendung finden kann (Abb. 7). Die Studie kann durch ihren erkundenden Charakter keine allgemeingültigen Ergebnisse der Kompetenzstruktur präsentieren, aber dennoch empirisch und theoretisch gut begründbar feststellen, dass das Modell aussichtsreich ist, um den Kompetenzbereich in seiner Struktur zu beschreiben. Es hat sich gezeigt, dass die theoretisch abgeleitete Modellstruktur in wesentlichen Punkten auch die empirischen Befunde erklären kann.

Dennoch sind diese Ergebnisse zunächst nur als Hinweise auf die grundsätzliche Eignung des Modells zu interpretieren. Die Einschränkungen ergeben sich insbesondere aus der Begrenzung der im Rahmen einer Einzelstudie außerhalb eines nationalen Bildungsmonitorings realistisch erreichbaren Probandenzahlen, was wiederum Grenzen bei der Anzahl einsetzbarer Testitems bewirkt. Dies soll im Folgenden näher diskutiert werden.

Um die Ergebnisse kritisch einschätzen zu können, muss betrachtet werden, ob die Items bei der Analyse einer Dimension in den anderen beiden Dimensionen gleich verteilt über alle Komponenten vorliegen. Würden beispielsweise alle Items der Komponente Kontext (Dimension *Aspekt*) den höchsten kognitiven Beiwert besitzen, so würde dies die Analyse der Dimension *Aspekt* erschweren. Es ergäbe sich möglicherweise eine Kompetenzstufe mit der Komponente Kontext auf dem höheren Level, obwohl der Unterschied lediglich auf den höheren kognitiven Beiwert der

Items zurück zu führen ist. Anzustreben ist, dass die Items sich auf die Komponenten der anderen Dimensionen gleichmäßig verteilen. Die Wahrscheinlichkeit, dass ein Item zur Komponente Kontext gehört, darf nicht davon abhängen, welchen kognitiven Beiwert dieses Item hat. Ein Eindruck über die Verteilung der Items ist in Tabelle 6 gegeben. Dort ist die Anzahl der Items, die nach der Rasch-Skalierung im Test verblieben sind, in Kreuztabellen zu den Dimensionen des Kompetenzmodells angegeben. Da der Test aus nur 19 Items besteht, lässt sich eine Gleichverteilung nicht statistisch überprüfen.

Ein χ^2 -Test könnte eine Aussage über Gleichverteilungen treffen, benötigt aber mindestens fünf Items pro Zelle des Kompetenzmodells – diese Anzahl liegt nicht vor. Da der χ^2 -Test jedoch auch bei Verletzungen dieser Mindestanzahl pro Zelle relativ robuste Interpretationen ermöglicht (Bortz, 2005, S. 177), soll er hier zumindest als heuristisches Maß für eine Gleichverteilung verwendet werden. In Tabelle 6 lässt sich erkennen, dass in der Verteilung der Komponenten der Dimension *kognitiver Beiwert* auf die Komponenten der Dimension *Perspektive* ein signifikanter χ^2 -Wert vorliegt (d.h. $\alpha < 0,05$), die Verteilung ist also nicht gleichmäßig. Es könnte also sein, dass die bei der Analyse der Dimension *Perspektive* resultierende Stufung zwischen den Komponenten adressatengemäß und sachgerecht aus einer Konfundierung durch den kognitiven Beiwert der Items resultiert. Die Analyseergebnisse deuten jedoch relativ klar auf eine Stufung hin und der χ^2 -Wert ist nur knapp signifikant. In einer weiteren Analyse der Dimension wurde deshalb die

Tab 6: Anzahl der Items pro Zelle des Kompetenzmodells. Zur Illustration eines Maßes für die Unabhängigkeit der Dimension angegeben ist außerdem der χ^2 -Wert nach Pearson, die Anzahl der Freiheitsgrade df sowie die asymptotische Signifikanz α

		Dimension Aspekt			
		Kontext	Code	Darstellungsform	Sachaspekt
Dimension Perspektive	adressatengemäß	2	2	3	5
	sachgerecht	0	3	2	2

$$\chi^2=2,546; df=3 \alpha=0,467$$

		Dimension Aspekt			
		Kontext	Code	Darstellungsform	Sachaspekt
Dimension Kognitiver Beiwert	Gruppe 1	2	2	3	0
	Gruppe 2	0	3	1	5
	Gruppe 3	0	0	1	2

$$\chi^2=10,133; df=6 \alpha=0,119$$

		Dimension kognitiver Beiwert		
		Gruppe 1	Gruppe 2	Gruppe 3
Dimension Perspektive	adressatengemäß	5	7	0
	sachgerecht	2	2	3

$$\chi^2=6,175; df=2 \alpha=0,046$$

Varianzaufklärung der Itemparameter des eindimensionalen Rasch-Modells durch die Zuordnung zu Komponenten der Dimension *Perspektive* wiederholt und der kognitive Beiwert als Kovariate behandelt. Es zeigt sich dabei weiterhin ein (nunmehr nur noch knapp) signifikanter Einfluss der Komponenten auf die Varianzaufklärung ($\alpha=0,05$), ebenso knapp signifikant ($\alpha=0,05$) ist eine partielle Korrelation von Itemparameter und Zuordnung zu den Komponenten der Dimension *Perspektive*, bei der der kognitive Beiwert als Kontrollvariable verwendet wird. Beides spricht dafür, dass die ange-

nommene Kompetenzstufe aufrechterhalten werden kann – doch Eindeutigkeit kann bei dieser Analyse nicht hergestellt werden. Es ist also eine Problemstellung für weiterführende Studien, ob eine Kompetenzstufe zwischen sachgerechter und adressatengemäßer Kommunikation unabhängig vom kognitiven Beiwert existiert. Umgekehrt könnte die Analyse des kognitiven Beiwerts durch die Dimension *Perspektive* konfundiert sein. Auch hier ergeben sich allerdings keine Änderungen in der Annahme hoher Signifikanz der Varianzaufklärungen des Itemparameters durch den kognitiven Bei-

wert, wenn die Komponenten der Dimension *Perspektive* (hilfsweise) als Kovariate behandelt werden und weiterhin eine signifikante partielle Korrelation zwischen den gruppierten kognitiven Beiwerten und den Itemparametern, wenn die Komponenten der Dimension *Perspektive* als Kontrollvariable hinzugefügt werden. Auch hier kann die in diesem Artikel vorgestellte Studie jedoch nur vorläufige Ergebnisse liefern.

Die Analysen der anderen Dimensionen werden nicht zu stark durch eine Ungleichverteilung der Items in der Matrix des Kompetenzmodells erschwert.

In Abbildung 7 ist ein nach den Ergebnissen überarbeitetes Modell physikalischer Kommunikationskompetenz abgebildet, in dem die Komponenten *Darstellungsform* und *Code* der Dimension *Aspekt* zusammengefasst wurden, da sie sich in den Analysen nicht voneinander trennen ließen. In anschließenden Studien sollte überprüft werden, ob sich diese Strukturen bestätigen lassen, dazu sollten nunmehr allerdings Ausschnitte des Kompetenzmodells in Einzelstudien untersucht werden, da zur sy-

stematischen Variation der Komponenten der Dimensionen ein Vielfaches an Items der hier beschriebenen Studie notwendig sind. Von der erfolgten breiten Testung der Möglichkeiten des Modells ist nunmehr der Schritt zu genauen Untersuchungen von Modellausschnitten gerechtfertigt. Das Modell scheint den Ergebnissen zufolge das Potential zu haben, auch in Large-Scale-Assessments einsetzbar zu sein und somit eine weitere Erweiterung und Präzisierung der hier vorgestellten Studie vorzunehmen. Neben dem Modell wurde in diesem Beitrag auch ein dazugehöriges schriftliches Testinstrument vorgestellt. Mit seiner Hilfe kann auch der Zusammenhang von physikalischer Kommunikationskompetenz und physikalischem Fachwissen untersucht werden. Der Zusammenhang zwischen adressatengemäßer Kommunikationsfähigkeit und physikalischem Fachwissen erweist sich dabei als komplex. Fachwissen ist die Voraussetzung dafür, eine durchschnittliche adressatengemäße Kommunikationskompetenz zu erreichen – es gibt jedoch Hinweise darauf, dass ein sehr hohes Fach-

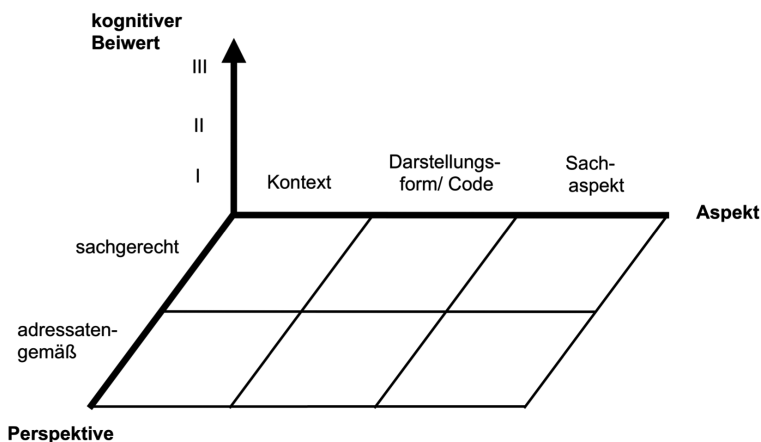


Abb. 7: Empirisch validiertes Modell für physikalische Kommunikationskompetenz.

wissen ein Hindernis für das Erreichen überdurchschnittlicher adressatengemäßer Kommunikationskompetenz ist. Über die Ergebnisse dieser Studie wird an anderer Stelle berichtet (Kulgemeyer, 2010). In einer Anschlussstudie soll überprüft werden, ob dieser Effekt durch eine geeignete Förderung von Kommunikationskompetenz vermieden werden kann. Das schriftliche Testinstrument ist in Kulgemeyer (2010) vollständig abgedruckt.

In Anlehnung an das in diesem Beitrag präsentierte Modell wurde physikalische Kommunikationskompetenz auch mit qualitativen Methoden untersucht und dabei insbesondere die motivational-sozial-volitionale Facette von Kompetenz im Weinert-schen Sinne analysiert. Die Ergebnisse dabei weisen darauf hin, dass der in diesem Beitrag verwendete schriftliche Test nur für die kognitive Facette ein valides Maß darstellt. Es zeigt sich dabei auch, dass Variationen in den Komponenten der Dimension Aspekt eine wichtige Rolle dabei spielen, Erklärungen einem bestimmten Adressaten geeignet vorzunehmen (Kulgemeyer, 2010; Kulgemeyer & Schecker, 2010). In anschließenden Studien sollte der Frage nachgegangen werden, inwiefern mit schriftlichen Testverfahren einem umfassenden Kompetenzbegriff überhaupt Rechnung getragen werden kann.

Literatur

- Bernholt, S., Parchmann, I. & Commons, M. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 219–245.
- Bond, T. & Fox, C. (2001). *Applying the Rasch Model. Fundamental Measurement in the Human Sciences*. London: Lawrence Erlbaum Associates.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler*. Heidelberg: Springer.
- Bricker, L.A. & Bell, P. (2008). Conceptualizations of argumentation from learning sciencestudies and the learning sciences and their implications for the practices of science education. *Science Education*, 92, 437–498.
- CSMEE (Center for Science, Mathematics and Engineering Education) (Hrsg.) (1996). *National science education standards*. Washington: National Academy Press.
- Davies, M. v. (2000). Winmira – a program system for analyses with the rasch model, with the latent class analysis and with the mixed rasch model. *Programma*.
- Dietrich, R. (2007). *Psycholinguistik* (2. Aufl.). Stuttgart; Weimar: Metzler.
- Driver, R., Newton, P., & Osborne, J. (2000). Establishing the norms of scientific argumentation in classroom. *Science Education*, 84(3), 287–312.
- Einhaus, E. (2007). *Schülerkompetenzen im Bereich Wärmelehre. Entwicklung eines Testinstruments zur Überprüfung und Weiterentwicklung eines normativen Modells fachbezogener Kompetenzen*. Berlin: Logos.
- Fricke, A. (2008). *Kompetenzbereich Kommunikation der Bildungsstandards Physik – Erprobung von Messverfahren. 1. Staatsexamensarbeit*. Universität Bremen.
- Hafner, R. (2007). Standards in science education in Australia. In D. Waddington, P. Nentwig & S. Schanze (Eds.), *Making it comparable. Standards in science education* (pp.23–60). Münster: Waxmann.
- Hartmann, U. (2008). *Perspektivenübernahme als eine Kompetenz historischen Verstehens*. Dissertation, Georg-August-Universität zu Göttingen.
- Johnston, C., Bottsford-Miller, N. & Thompson, S. (2006). *Using think aloud method (cognitive labs) to evaluate test design for students with disabilities and English language learners*. Forschungsbericht Nr. 44, Nationale Center on Educational Outcomes, Minneapolis.

- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H.-E. & Vollmer, H. (2003). *Zur Entwicklung nationaler Bildungsstandards – Expertise*. Bundesministerium für Bildung und Forschung (BMBF).
- KMK (Hrsg.) (2005). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss*. München: Luchterhand.
- Kulgemeyer, C. (2009). *PISA-Aufgaben im Vergleich: Strukturanalyse der Naturwissenschaftsitems aus den PISA-Durchläufen 2000 bis 2006*. Norderstedt: BoD.
- Kulgemeyer, C. & Schecker, H. (2009). Kommunikationskompetenz in der Physik: Zur Entwicklung eines domänenspezifischen Kompetenzbegriffs. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 131–153.
- Kulgemeyer, C. (2010). *Physikalische Kommunikationskompetenz. Modellierung und Diagnostik*. Berlin: Logos.
- Kulgemeyer, C. & Schecker, H. (2010). Kompetenzdiagnostik mit qualitativen Methoden am Beispiel eines Rollenspiels. Zum Verhältnis von Kognition und Volition bei Kommunikationskompetenz. *PhyDid B – Didaktik der Physik – Beiträge zur DPG-Frühjahrstagung 1*. <http://www.phydid.de/index.php/phydid-b/article/view/167/147> (Abgerufen am 1.8.2011).
- Leisen, J. (2005). Wechsel der Darstellungsformen. Eine wichtige Strategie im kommunikativen Physikunterricht. *Naturwissenschaften im Unterricht Physik* 16(87), 10–11.
- Merten, K. (1995). Konstruktivismus als Theorie für die Kommunikationswissenschaft. *Medien Journal* 19(4), 3–20.
- Osborne, J., Erduran, S. & Simon, S. (2004). Enhancing the quality of argumentation in school science. *Journal of Research in Science Teaching*, 41(10), 994–1020.
- Rost, J. (2004). *Testtheorie – Testkonstruktion*. Bern: Hans Huber.
- Rusch, G. (1999). Eine Kommunikationstheorie für kognitive Systeme. In G. Rusch & S. Schmidt (Eds.), *Konstruktivismus in der Medien- und Kommunikationswissenschaft* (pp. 150–184). Frankfurt a. M.: Suhrkamp.
- Schecker, H. & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften* 12, 45–66.
- Schmidt, M. (2008). *Kompetenzmodellierung und -diagnostik im Themengebiet Energie der Sekundarstufe I. Entwicklung und Erprobung eines Testinventars*. Berlin: Logos.
- Weinert, F. (2001). Vergleichende Leistungsmessung in Schulen – eine umstrittene Selbstverständlichkeit. In F. Weinert (Ed.), *Leistungsmessung in Schulen* (pp. 17–31). Weinheim: Beltz.
- Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.

KONTAKT

Dr. Christoph Kulgemeyer

Universität Bremen, Institut für Didaktik der Naturwissenschaften, Abt. Physikdidaktik
Fachbereich 1 Physik/Elektrotechnik
Postfach 330440
D-28334 Bremen
kulgemeyer@physik.uni-bremen.de

AUTORENINFORMATION

Christoph Kulgemeyer ist zur Zeit Referendar für die Fächer Physik und Deutsch am Hermann-Böse-Gymnasium Bremen und als Gastwissenschaftler in der Arbeitsgruppe von Horst Schecker tätig. Er hat seine Dissertation über die Modellierung physikalischer Kommunikationskompetenz verfasst. Ab 1.5.2012 wird er als Postdoc in die Arbeitsgruppe von Horst Schecker zurückkehren.

Horst Schecker arbeitet an der Modellierung physikalischer Kompetenzstrukturen bei Schülern sowie daran orientierten Lernaufgaben und Leistungstests. Langfristige Forschungsgebiete sind Schülervorstellungen und Lernprozesse im Physikunterricht sowie Studien zur Lernwirksamkeit virtueller Lehr- und Lernmedien. Horst Schecker ist seit 2005 Sprecher des Vorstands der Gesellschaft für Didaktik der Chemie und Physik (GDGP).