

JAN CHRISTOPH HADENFELDT UND KNUT NEUMANN

Die Erfassung des Verständnisses von Materie durch Ordered Multiple Choice Aufgaben

Assessing students' understanding of matter using ordered multiple choice items

ZUSAMMENFASSUNG

Die Vermittlung des Verständnisses naturwissenschaftlicher Konzepte, wie z. B. dem Materiekonzept, setzt entsprechende Modelle der Verständnisentwicklung voraus. Zur Validierung dieser Modelle ist es notwendig, das Verständnis der jeweiligen Konzepte reliabel und valide erfassen zu können. Briggs, Alonzo, Schwab und Wilson (2006) schlagen in diesem Zusammenhang einen neuen Aufgabentyp vor, die sogenannten Ordered Multiple Choice (OMC) Aufgaben. Mit diesem Aufgabentyp können in einer Aufgabe unterschiedliche Niveaus des Verständnisses im Sinne unterscheidbarer, hierarchisch geordneter Ausprägungen einer latenten Fähigkeit erfasst werden. In der in diesem Beitrag vorgestellten Studie wurde untersucht, inwieweit OMC Aufgaben geeignet sind, das Verständnis von Materie zu erfassen. $N=294$ Schülerinnen und Schüler der Jahrgangsstufen 6 bis 12 eines Gymnasiums in Schleswig-Holstein wurden zehn OMC Aufgaben und entsprechende offene Aufgaben vorgelegt. Die Ergebnisse legen nahe, dass OMC Aufgaben unterschiedliche Niveaus im Verständnis von Materie über verschiedene Jahrgangsstufen hinweg erfassen können und damit zur Validierung eines Modells der Entwicklung des Verständnisses von Materie geeignet sind.

Schlüsselwörter: Physik, Chemie, Materie, Ordered Multiple Choice Aufgaben, Testinstrument, Kompetenzentwicklung

ABSTRACT

Fostering students' progression in understanding scientific concepts such as the concept of matter requires a validated model of the teaching and learning of the respective concepts. Validation of such models in turn requires instruments which can reliably and validly assess students' understanding of the concepts. In this context Briggs et al. (2006) recently suggested a new type of assessment item, the so called Ordered Multiple Choice (OMC) items. This type of item allows assessing different levels in students' understanding as distinct, hierarchically ordered magnitudes of a latent trait. This study investigates the extent to which OMC items may be utilized to assess students' understanding of matter. Ten OMC items and corresponding open-ended versions were administered to a sample of $N=294$ students from grades 6 to 12 at a grammar-school in Schleswig-Holstein, Germany. The results suggest that OMC items are in fact suitable to assess different levels in students' understanding of the concept of matter across different grades, which in turn

suggests, that OMC items can be utilized for validation of a respective model of the teaching and learning of matter from 6 to 12.

Keywords: physics, chemistry, matter, ordered multiple choice items, instrument development, learning progression.

1 Einleitung

Der Erwerb einer naturwissenschaftlichen Grundbildung ist ein zentrales Ziel des Unterrichts in den naturwissenschaftlichen Fächern (Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK], 2005a, 2005b, 2005c; vgl. Bernholt, Nentwig & Neumann, 2012). Als wesentliches Element einer naturwissenschaftlichen Grundbildung gilt ein strukturiertes Fachwissen auf der Basis zentraler Konzepte des jeweiligen Fachs, wie z. B. Materie (vgl. KMK 2005a, 2005b, 2005c; siehe auch Duschl, Schweingruber & Shouse, 2007). Die systematische Vermittlung eines solchen Fachwissens im Unterricht in den naturwissenschaftlichen Fächern setzt zunächst Modelle der Entwicklung des Verständnisses dieser zentralen Konzepte voraus. In jüngster Zeit wurden derartige Modelle für verschiedene Konzepte vorgeschlagen (Alonzo & Steedle, 2009; Bernholt, Commons, & Parchmann, 2009; Mohan, Chen, & Anderson, 2009; Neumann, Viering, Boone & Fischer, 2012). Mit den theoretischen Herausforderungen der Entwicklung solcher Modelle ist auch die Frage verbunden, wie diese Modelle angemessen auf ihre empirische Gültigkeit überprüft werden können (vgl.

Koeppen, Hartig, Klieme, & Leutner, 2008). Für die empirische Validierung von Modellen der Verständniseentwicklung werden Instrumente benötigt, mit denen die Stufe der Verständniseentwicklung, auf der sich eine Schülerin bzw. ein Schüler befindet, reliabel und valide erfasst werden kann. Vor diesem Hintergrund haben Briggs et al. (2006) kürzlich einen neuen Aufgabentyp vorgeschlagen: die sogenannten Ordered Multiple Choice (OMC) Aufgaben. Im Gegensatz zu traditionellen Multiple Choice Aufgaben können bereits mit einer einzelnen OMC Aufgabe unterschiedliche Niveaus des Verständnisses eines Konzepts erfasst werden. Dadurch kann bei gleicher Effizienz in der Auswertung mehr diagnostische Information gewonnen und damit eine Validität bei der Erfassung konzeptuellen Verständnisses erreicht werden, die dichter an der Validität offener Aufgaben liegt (Briggs et al., 2006). Allerdings liegen bisher nur wenige Studien vor, die diese Annahmen für die Erfassung naturwissenschaftlicher Konzepte bestätigen (z. B. Alonzo & Steedle, 2009).

Mit dem Ziel, OMC Aufgaben für die Validierung eines Modells der Entwicklung des Verständnisses von Materie (siehe Hadenfeldt & Neumann, 2011) nutzbar zu machen, wurde daher in der in diesem

Beitrag vorgestellten Studie untersucht, inwieweit OMC Aufgaben geeignet sind, das Verständnis des Materiekonzepts reliabel und valide zu erfassen. Dabei wurde insbesondere auch untersucht, inwieweit OMC Aufgaben zur gleichen Einschätzung bezüglich der Stufe in der Entwicklung des Verständnisses von Materie kommen wie entsprechende offene Aufgaben.

2 Theoretischer Hintergrund

Das Materiekonzept gilt als zentrales Konzept der Naturwissenschaften (KMK, 2005c; American Association for the Advancement of Science [AAAS], 2001; NRC, 1996). Die Bildungsstandards für den mittleren Schulabschluss im Fach Physik benennen Materie als eines von vier Basiskonzepten. Im Fach Chemie wird es durch vier Basiskonzepte zu unterschiedlichen Anteilen beschrieben (vgl. KMK, 2005b).

2.1 Die Entwicklung des Verständnisses von Materie

Der Schwerpunkt der bisherigen Forschung zum Verständnis von Materie lag auf der Beschreibung verschiedener Schülervorstellungen von Materie in einzelnen Altersgruppen (vgl. Duit, 2004). Eine erste Systematisierung findet sich bei Andersson (1990). Er gruppiert die Schülervorstellungen von Materie grob in zwei Kategorien: Die Kategorie *Alltägliche Vorstellungen* umfasst dabei die Alltagsvorstellungen von Schülerinnen und Schülern zu chemischen Reaktionen, Aggregatzu-

ständen, der Änderung von Aggregatzuständen sowie der Erhaltung von Materie. Die Kategorie *Teilchenvorstellungen* umfasst die Vorstellungen, die Schülerinnen und Schüler von Atomen, Molekülen und Teilchensystemen besitzen. In den darauf folgenden Jahren wurden weitere Schülervorstellungen von chemischen Reaktionen (z. B. Ahtee & Varjola, 1998), von Aggregatzuständen oder der Änderung von Aggregatzuständen (z. B. Johnson, 1998), von der Erhaltung von Materie (z. B. Gómez, Bennarock & Marín, 2006) sowie von der Struktur und dem Aufbau von Materie (z. B. Nakhleh & Samarapungavan, 1999) dokumentiert (vgl. u. a. Liu, 2001; Johnson, 2002; Kind, 2004; Barke, 2006). Die Vorstellungen von Schülerinnen und Schülern zum Materiekonzept sind demnach ausführlich dokumentiert. Wie sich jedoch das Verständnis im Umgang mit dem Materiekonzept im Verlauf der Schulzeit entwickelt, ist bisher nur in geringerem Umfang empirisch untersucht. Bereits Andersson (1990) postulierte bezogen auf die von ihm identifizierten Bereiche eine Entwicklung des Verständnisses von Materie von einer zunächst eher makroskopischen hin zu einer mikroskopischen durch das Verständnis der Teilchennatur von Materie geprägten Vorstellung (vgl. Liu & Lesniak, 2005). Krnel, Watson und Glazar (1998) vermuten in ähnlicher Weise, dass sich das Verständnis von Materie mit zunehmender Fähigkeit zur Unterscheidung zwischen Objekten und Stoffen entwickelt. In einer Querschnittsstudie mit insgesamt $N=84$ Probanden verschiedener Altersgruppen zwischen 3 und 13 Jahren finden sie

diese Annahme bestätigt (Krnel, Glazar & Watson, 2003). Johnson (1998) untersucht die Entwicklung des Verständnisses von Materie in einer dreijährigen Längsschnittuntersuchung mit $N=36$ Probanden im Alter von 11 bis 14 Jahren. Aufgrund der in klinischen Interviews geäußerten Vorstellungen vom Siedevorgang postuliert Johnson (1998) vier aufeinanderfolgende Stadien in der Entwicklung der Teilchenvorstellungen der Schülerinnen und Schüler: (1) Materie als homogene Masse; (2) Materie enthält kleinste Teilchen; (3) Materie besteht aus kleinsten Teilchen, denen makroskopische Eigenschaften zugeordnet werden; (4) Materie besteht aus kleinsten Teilchen, Eigenschaften von Materie werden auf das kollektive Verhalten der Teilchen zurückgeführt. Dieses oder diesem sehr ähnliche Systeme bildeten für viele nachfolgende Untersuchungen die Grundlage für die Kategorisierung von Schülervorstellungen zu Materie (vgl. dazu Buck, Johnson, Fischler, Peuckert, & Seifert, 2001; Johnson, 2002; Papageorgiou & Johnson, 2005; Gómez et al., 2006; Margel, Eylon, & Scherz, 2008; Löfgren & Helldén, 2009).

In neueren Studien liegt der Schwerpunkt auf der detaillierten Untersuchung der Veränderung von Schülervorstellungen von Materie im Längsschnitt. So berichten z. B. Adbo und Taber (2009) die Ergebnisse des ersten Erhebungszeitpunktes einer gerade begonnenen Längsschnittuntersuchung mit $N=18$ Schülerinnen und Schülern. Adadan, Irving und Trundle (2009) untersuchen, welchen Einfluss Unterricht auf Grundlage eines *Reform-Based Teaching* auf die Materievorstellungen von

Schülerinnen und Schülern hat. Löfgren und Helldén (2009) untersuchen in einer Längsschnittstudie über 10 Jahre, wie sich die Fähigkeit von Schülerinnen und Schülern, das Materiekonzept zur Erklärung von Alltagssituationen heranzuziehen, im Verlauf der Schulzeit verändert. Peper et al. (2007) untersuchen in mehreren aufeinanderfolgenden Unterrichtseinheiten, wie Schülerinnen und Schüler das Materiekonzept nutzen, um (Alltags-)Situatio- nen zu beschreiben bzw. zu erklären und skizzieren ausgehend von ihren Befunden mögliche Verläufe in der Entwicklung des Verständnisses von Materie. Alle diese Untersuchungen weisen darauf hin, dass sich Schülervorstellungen von Vorstellungen, die sich an dem konkret Beobachtbaren orientieren, hin zu abstrakten, mit differenzierten Teilchenvorstellungen verbundenen Vorstellungen entwickeln.

Ein Vergleich der bisherigen Untersuchungen zur Entwicklung des Verständnisses von Materie ergibt das in Abbildung 1 dargestellte Modell (für eine ausführliche Darstellung siehe Hadenfeldt & Neumann, 2011). Darin wurden die vier von Johnson (1998) postulierten Verständnisniveaus um ein fünftes ergänzt, um vor allem die Befunde neuerer Studien zu berücksichtigen (u. a. Adbo & Taber, 2009; Park & Light, 2009; Stevens et al., 2010). Es wird angenommen, dass sich die Entwicklung des Verständnisses von Materie im Verlauf der Schulzeit entlang dieser fünf Niveaus vollzieht. Dies bedeutet jedoch nicht, dass sich das Verständnis sprunghaft entwickelt. Vielmehr wird darunter eine kontinuierliche Entwicklung der Fähigkeit verstanden, das entsprechende Verständnis in einer zu-

nehmenden Zahl von Kontexten anwenden zu können (vgl. dazu auch Steedle & Shavelson, 2009). Zudem wird nicht angenommen, dass sich die Entwicklung ausschließlich im Sinne eines zunehmend ausdifferenzierten wissenschaftlichen Verständnisses vollzieht (vgl. Johnson, 1998; Nakhleh & Samarapungavan, 1999). Vielmehr wird auf Basis der bisherigen Befunde, dass die Integration neuen Wissens häufig mit dem Auftreten von Fehlvorstellungen einhergeht (vgl. Talanquer, 2006), angenommen, dass auf jedem Verständnisniveau auch entsprechend elaborierte Fehlvorstellungen möglich sind.

2.2 Erfassung des Materieverständnisses durch Ordered Multiple Choice Aufgaben

Das beschriebene Modell der Entwicklung des Materieverständnisses ist bisher noch hypothetischer Natur und bedarf einer empirischen Prüfung. Dazu werden Aufgaben benötigt, die das Verständnis von Materie erfassen und eine Zuordnung zu den im Modell beschriebenen Niveaus zulassen. Die Konstruktion von Aufgaben, die statt auswendig gelerntem Faktenwissen, tiefgehendes konzeptuelles Verständnis erfassen sollen, bewegt sich im Spannungsfeld zwischen Objektivität, Reliabilität und Validität. Offenen Aufgaben wird die Eigenschaft zugeschrieben, eine besonders valide Erfassung konzeptuellen Verständnisses zu ermöglichen, weil die

NIVEAU	BESCHREIBUNG
5	<i>Systemische Teilchenvorstellungen</i> Materie ist aus Teilchen aufgebaut, die wiederum aus weiteren Teilchen aufgebaut sind. Das kollektive Verhalten dieser Teilchenensembles bestimmt die Eigenschaften der Materie.
4	<i>Differenzierte Teilchenvorstellungen</i> Materie ist aus Teilchen (Atomen) aufgebaut, die wiederum aus weiteren Teilchen aufgebaut sind.
3	<i>Einfache Teilchenvorstellungen</i> Materie ist aus Teilchen aufgebaut, die als „letztes Teilungsstück“ aufgefasst werden und denen makroskopische Eigenschaften zugeordnet werden.
2	<i>Hybridvorstellungen</i> Materie enthält Teilchen als in einen Stoff eingebettete Entitäten. Zwischen den Teilchen ist der eigentliche Stoff.
1	<i>Alltagsvorstellungen</i> Materie ist alles, was Raum einnimmt und eine Masse hat. Materie ist ein Kontinuum, das portionierbar ist, aber nicht weiter aus Teilchen aufgebaut ist. Materie kann erschaffen und vernichtet werden.

Abb. 1: Modell der Verständniseentwicklung von Materie (Hadenfeldt & Neumann, 2011).

Schülerinnen und Schüler die Möglichkeit haben, ihre komplexen Überlegungen darzustellen (Briggs et al., 2006). Aus den Antworten der Schülerinnen und Schüler lässt sich jedoch nicht direkt auf ihr Verständnis zurückschließen. Die Antworten müssen interpretiert werden. Dies beeinträchtigt die Objektivität bzw. Reliabilität der Einschätzung. Die Erfassung kognitiver Fähigkeiten wie Wissen oder Verständnis mit Hilfe von Multiple Choice Aufgaben gilt umgekehrt als besonders objektiv bzw. reliabel (Wainer & Thissen, 1993). Zudem ermöglicht die kürzere Bearbeitungszeit die Administration einer höheren Zahl von Aufgaben in der gleichen Zeit, was üblicherweise die Reliabilität der Messung erhöht (vgl. Bortz, 2006). Allerdings wird die Eignung von Multiple Choice Aufgaben für die Erfassung konzeptuellen Verständnisses in Frage gestellt (Berlak, 1992; Ruiz-Primo & Shavelson, 1996; Shavelson, Carey & Webb, 1990). Ein Kritikpunkt ist dabei, dass selbst Schülerinnen und Schüler die über kein Verständnis verfügen, eine gewisse (von der Zahl der Antwortoptionen abhängige) Wahrscheinlichkeit besitzen, die Aufgabe richtig zu lösen. Zudem ist häufig unklar, inwieweit die Antwortoptionen Schülerinnen und Schüler überhaupt gleichwertig erscheinen oder ob diese einzelne Antwortoptionen ausschließen können, weil sie ihnen unsinnig erscheinen. Im Hinblick auf eine Modellprüfung ist es zudem von Nachteil, dass eine Multiple Choice Aufgabe nur erfassen kann, ob Schülerinnen und Schüler ein spezifisches Verständnisniveau erreicht haben oder nicht. Offene Aufgaben hingegen erlauben eine Ein-

schätzung, welchem Niveau im Modell das bei der Bearbeitung ausgedrückte Verständnis entspricht.

In diesem Spannungsfeld schlagen Briggs et al. (2006) einen Aufgabentyp vor, den sie als Ordered Multiple Choice (OMC) bezeichnen. Dieser Aufgabentyp soll die Vorteile von offenen und Multiple Choice Aufgaben kombinieren. So können in einer Aufgabe mehrere Verständnisniveaus im Sinne unterscheidbarer, hierarchisch geordneter Ausprägungen einer latenten Fähigkeit erfasst werden. Jede Antwortoption einer OMC Aufgabe bildet ein bestimmtes Verständnisniveau ab. Die richtige Antwortoption markiert das höchste in der Aufgabe erfasste Verständnisniveau, während die anderen Antwortoptionen niedrigeren Verständnisniveaus entsprechen (vgl. Abbildung 2). Dies stellt hohe Ansprüche an die Entwicklung der Aufgaben. Üblicherweise sollen alle Antwortoptionen einer traditionellen Multiple Choice Aufgabe für Schülerinnen und Schüler, die nicht in der Lage sind, die Aufgabe korrekt zu beantworten, gleich plausibel sein (Liu, 2010). Die Antwortoptionen von OMC Aufgaben sind hingegen so gestaltet, dass je nach Ausprägung des Verständnisses, eine der Antwortoptionen (das können und sollen auch explizit Distraktoren sein) für Schülerinnen und Schüler plausibler ist, als die anderen. Dem zu Grunde liegt die Annahme, dass sich jedes Verständnisniveau nicht nur durch bestimmte wissenschaftliche Vorstellungen, sondern auch durch typische alternative Vorstellungen auszeichnet (vgl. Briggs et al., 2006). Geeignete alternative Vorstellungen für die Konstruktion der Distrak-

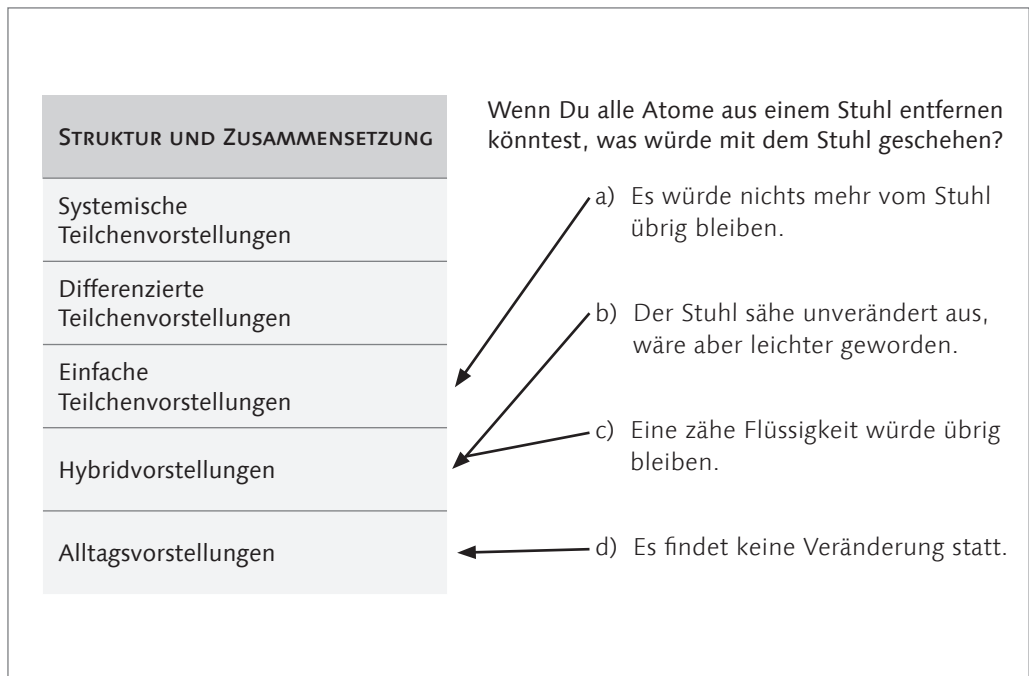


Abb. 2: Beispiel einer OMC Aufgabe.

toren zu verwenden, ist eine wesentliche Herausforderung bei der Entwicklung von OMC Aufgaben (Alonzo & Steedle, 2009). Dabei ist insbesondere darauf zu achten, dass die vermeintliche unterschiedliche Elaboriertheit der Antwortoptionen Schülerinnen und Schüler nicht dazu verleitet, bestimmte Antwortoptionen von vornherein auszuschließen (Liu, 2010). OMC Aufgaben wurden bereits zur Erfassung von konzeptuellem Verständnis in zwei Themenbereichen eingesetzt: „The Earth in the Solar System“ (Briggs et al., 2006) und „Force and Motion“ (Alonzo & Steedle, 2009). Alonzo und Steedle (2009) konnten dabei zeigen, dass OMC Aufgaben Verständnis sogar besser abbilden können, als offene Aufgaben (bezogen auf Schüleräußerungen in Interviews). Voraussetzung dafür ist ein eindimensionales Entwicklungsmodell, in

dem aufeinander aufbauende Niveaus beschrieben werden, von denen angenommen wird, dass sie einen sukzessiven Verständnis-aufbau beschreiben (vgl. Lin, Chu, & Meng, 2010). Das im letzten Abschnitt beschriebene Entwicklungsmodell für ein Verständnis von Materie (vgl. Abbildung 1) erfüllt diese Voraussetzungen. Ob OMC Aufgaben geeignet sind, das Verständnis von Materie – insbesondere Stufen der Entwicklung des Verständnisses von Materie – reliabel und valide zu erfassen, ist bisher jedoch unklar. Daher soll die hier beschriebene Studie zur Klärung der folgenden Fragestellung beitragen:

Inwieweit lässt sich das Verständnis von Schülerinnen und Schülern im Umgang mit dem Materiekonzept durch OMC Aufgaben reliabel und valide erfassen?

3. Methode

Ziel der vorliegenden Studie ist, zu prüfen, inwieweit ein auf OMC Aufgaben basierendes Testinstrument das Verständnis von Schülerinnen und Schülern im Umgang mit dem Materiekonzept reliabel und valide erfassen kann. Das heißt, wie präzise lässt sich durch den Test Verständnis erfassen und inwieweit lässt das Testergebnis Rückschlüsse auf die Ausprägung des Verständnisses zu. Zur Klärung der Frage nach der *Reliabilität* eines Tests werden häufig entsprechende Koeffizienten, wie zum Beispiel Cronbachs Alpha verwendet (vgl. auch Bortz und Döring (2006) für eine ausführlichere Darstellung). *Validität* hingegen lässt sich nicht auf einen Koeffizienten reduzieren (Hartig, Frey, & Jude, 2007). Grundsätzlich bezieht sich Validität auf das Ausmaß, in dem der Test das misst, was er zu messen vorgibt (Messick, 1995; vgl. Wild & Möller, 2009). Aus fachdidaktischer Sicht ist dieser Begriff mit einer Reihe bedeutsamer Fragestellungen verbunden, zum Beispiel: Repräsentieren die eingesetzten Fragen das zu untersuchende Konstrukt angemessen? Sind die Antwortmöglichkeiten so formuliert, dass tatsächlich die intendierten kognitiven Fähigkeiten auf Seiten der Schülerinnen und Schüler bei der Bearbeitung des Tests zum Einsatz kommen?

In der Literatur finden sich verschiedene Konzeptionen von Validität (vgl. Hartig et al., 2007). Messick (1995) vereint diese zu einer Konzeption, wie sie auch in den Standards der American Psychological Association (American Psychological Association, 2001) verwendet wird. Unter dem Begriff

Konstruktvalidität werden dabei sechs Aspekte subsummiert. Als *inhaltliche Validität* wird die Eignung der Aufgaben verstanden, das zu untersuchende Konstrukt inhaltlich zu erfassen. *Kognitive Validität* bezieht sich auf die stattfindenden kognitiven Prozesse bei der Bearbeitung von Aufgaben und inwieweit sie den theoretisch zu erwartenden Prozessen entsprechen. Unter *struktureller Validität* wird der Grad verstanden, zu dem die empirisch gewonnenen Daten die theoretisch erwartete Konstruktstruktur abbilden. *Kriteriumsorientierte Validität* beschreibt den Zusammenhang zu anderen Konstrukten und Variablen (konvergente und divergente Validität). Ein weiterer Aspekt von Validität ist die *Generalisierbarkeit* der Testergebnisse auf Personengruppen, die nicht zu den Probanden gehörten und auf weitere, für das zu erfassende Konstrukt relevante Aufgaben, die nicht Teil des Tests waren. Schließlich bildet ein weiterer Aspekt von Validität die *Konsequenzen* von Aussagen, die anhand der Testergebnisse getroffen werden können. Dieser Aspekt umschließt sowohl intendierte, wie auch nichtintendierte Konsequenzen. Alle sechs Aspekte tragen zu einem umfassenden Bild von Validität bei und müssen daher bei der Beantwortung einer solchen Fragestellung berücksichtigt werden. Ziel dieser Studie ist es daher, bezogen auf die eben vorgestellten sechs Aspekte Evidenz für Validität herzustellen.

3.1 Anlage der Untersuchung

Zur Prüfung, inwieweit OMC Aufgaben geeignet sind, das Verständnis von Mate-

rie und die Entwicklung im Verständnis von Materie reliabel und valide zu erfassen, wurden im Rahmen dieser Studie zunächst OMC Aufgaben zur Erfassung des Verständnisses von Materie entwickelt. Da ein besonderes Augenmerk darauf lag, inwieweit OMC Aufgaben Verständnis von Materie ähnlich valide wie offene Aufgaben erfassen können (vgl. Abschnitt 2.2), wurden anschließend aus diesen Aufgaben offene Aufgaben generiert. Aus diesen Aufgaben wurden Testhefte zusammengestellt, die in einem ersten Teil ausgewählte offene Aufgaben und in einem zweiten Teil die entwickelten OMC Aufgaben enthielten.

Diese Testhefte wurden $N=294$ Schülerinnen und Schülern der Klassenstufe 6 bis 12 eines Kieler Gymnasiums zur Bearbeitung vorgelegt. An diesem Gymnasium wird Physik ab Jahrgangsstufe 7 und Chemie ab Jahrgangsstufe 8 unterrichtet. Die Entscheidung, Schülerinnen und Schüler nur einer Schule zu befragen, wurde bewusst getroffen, um den Einfluss schulspezifischer Merkmale wie z. B. des Schulcurriculums zu reduzieren. Die Schülerinnen und Schüler waren zunächst aufgefordert, den ersten Teil mit den offenen Aufgaben zu bearbeiten und anschließend den zweiten Teil mit den OMC Aufgaben. Sobald mit der Bearbeitung der OMC Aufgaben begonnen wurde, durften die Schülerinnen und Schüler nicht mehr zurückblättern. Dies wurde gezielt durch den Testleiter kontrolliert. Zur Prüfung konvergenter bzw. divergenter Validität wurden zusätzlich noch die Noten der Schülerinnen und Schüler in den naturwissenschaftlichen Fächern und in den Sprachen in einem

Kurzfragebogen erfasst. Die Befragung der Schülerinnen und Schüler dauerte insgesamt 45 Minuten.

3.2 Instrument

Da zunächst noch nicht die Prüfung des Entwicklungsmodells, sondern lediglich die Eignung von OMC Aufgaben zur Erfassung unterschiedlicher Niveaus im Verständnis von Materie im Vordergrund stand, wurden zunächst zehn OMC Aufgaben entwickelt. Zur Sicherung inhaltlicher Validität wurde bei der Entwicklung theoriegeleitet vorgegangen (vgl. Abschnitt 2.2). Aufbauend auf einem Modell der Verständnisentwicklung im Umgang mit dem Materiekonzept (vgl. Hadenfeldt & Neumann, 2011) wurden curriculare Analysen durchgeführt, um die Inhalte zu identifizieren, die in der Sekundarstufe I unterrichtet werden. Als Grundlage dienten dabei neben dem Lehrplan des Landes Schleswig-Holstein für die Fächer Chemie und Physik (Ministerium für Bildung, 1997b, 1997a) die schulinternen Fachcurricula der beiden Fächer. Darauf aufbauend wurden Aufgabenstellungen entwickelt, die Schülerinnen und Schülern der Sekundarstufe I in Schleswig-Holstein aus dem Alltag oder Schulalltag bekannt sein sollten.

Auf Grundlage der Literatur zu alternativen Schülervorstellungen (u. a. Liu, 2001; Kind, 2004; Smith, Wisser, Anderson, Krajcik, & Coppola, 2006; Nakhleh, Samarapungavan, & Saglam, 2005; Talanquer, 2009) wurden anschließend Antwortoptionen entwickelt, die die im Modell vor-

gesehenen Verständnisniveaus abbilden. Jede OMC Aufgabe ist dabei so aufgebaut, dass es nur eine wissenschaftlich korrekte Antwortoption gibt, die innerhalb der Aufgabe dem höchsten Verständnisniveau zugeordnet ist. Dies muss nicht dem höchsten Niveau des Modells entsprechen (vgl. Abschnitt 2.2). Die Distraktoren einer OMC Aufgabe bilden Verständnisniveaus ab, welche sich unterhalb dieses Verständnisniveaus befinden. Diese müssen jedoch nicht vollständig abgebildet werden. Das heißt, eine OMC Aufgabe kann z. B. ausschließlich die Verständnisniveaus 4, 3 und 2 abbilden (vgl. Alonzo & Steedle, 2009). Bei der Konstruktion der Aufgaben für diese Studie lag der Fokus auf den Verständnisniveaus 1 bis 4. Die Zuordnung der Antwortoptionen zu den Verständnisniveaus im Modell wurde durch ein Expertenrating mit drei Fachdidaktikerinnen bzw. Fachdidaktikern (2 x Physik, 1 x Chemie) abgesichert. Die mittlere Übereinstimmung der Experten wurde mit einem gewichteten Cohens Kappa (vgl. Wirtz & Caspar, 2002) zu $\kappa = .86$ bestimmt, was eine sehr gute Übereinstimmung darstellt. Für die Konstruktion der entsprechenden offenen Aufgaben wurden die Aufgabenstellungen ohne die Antwortoptionen verwendet. Aus den offenen und OMC Aufgaben wurden anschließend wie folgt Testhefte zusammengestellt: Aus den zehn offenen Aufgaben wurden, zusammen mit drei weiteren offenen Aufgaben zur Konstruktion von zusätzlichen OMC Aufgaben in Folgestudien, dreizehn erste Testteile zu je drei Aufgaben erstellt. Dabei wurden die Aufgaben in einem rotierenden Verfahren auf die Testteile verteilt,

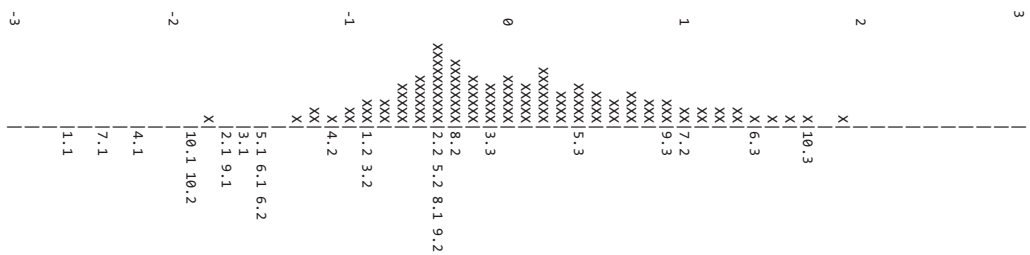
wobei jeweils benachbarte Testhefte zwei gemeinsame offene Aufgaben enthielten. Diese wurden anschließend zusammen mit einem zweiten Testteil, der alle zehn OMC Aufgaben enthielt, zu vollständigen Testheften kombiniert.

4. Auswertung und Ergebnisse

Um Evidenz bezüglich Reliabilität und Validität von OMC Aufgaben zu generieren, wurden die gewonnenen Daten in drei Schritten ausgewertet: Zuerst wurde überprüft, inwieweit die Aufgaben eine unidimensionale latente Fähigkeit erfassen und die Aufgaben in der Lage sind, die Entwicklung des Verständnisses im Verlauf der Schulzeit zu erfassen. Im zweiten Schritt wurde untersucht, inwieweit die Schülerinnen und Schüler über die verschiedenen Aufgaben hinweg ein konsistentes Antwortverhalten zeigen (vgl. Abschnitt 2.2). Im dritten und letzten Schritt wurde das Antwortverhalten der Schülerinnen und Schüler bei den offenen Aufgaben mit dem Antwortverhalten bei den entsprechenden OMC Aufgaben verglichen.

4.1 Analyse von Aufgabenschwierigkeit und Personenfähigkeit

Zur Untersuchung, inwieweit die OMC Aufgaben eine latente Fähigkeit (das Verständnis der Struktur und Zusammensetzung von Materie) erfassen, wurden die Daten unter Verwendung des Raschmodells (Rasch, 1960) analysiert.



Oberhalb der Horizontalen: Personenfähigkeit (X = 2,6 Schülerinnen und Schüler)
 Unterhalb der Horizontalen: Generalized Item Thresholds

Abb. 3: Wright Map.

Insbesondere wurde geprüft, inwieweit die einzelnen Aufgaben die theoretisch postulierten Verständnisniveaus erfassen können. Dazu wurden die einzelnen Antwortoptionen zunächst durch den Zahlenwert des ihnen zugeordneten Verständnisniveaus kodiert. Nicht bearbeiteten Aufgaben wurde der Zahlenwert 1 zugewiesen. Dies entspricht dem niedrigsten im Modell vorgesehenen Verständnisniveau, das durch ein vorschulisches, durch alltägliche Erfahrungen geprägtes Verständnis charakterisiert ist, über das jede Schülerin bzw. jeder Schüler verfügen sollte.

Anschließend wurde ein Partial Credit Raschmodell (Bond & Fox, 2001) an die Daten angepasst. Neun der zehn OMC Aufgaben erfüllten die üblichen Kriterien für eine gute Modellpassung ($.8 < \text{weighted MNSQ} < 1.2$ bzw. $T < 2.0$) (vgl. Bond & Fox, 2001). Die WLE-Reliabilität für diese Aufgaben lag bei .71. Dies ist für Instrumente, die komplexe Konstrukte, wie z. B. Kompetenz erfassen, ein zufriedenstellender Wert (vgl. Kauertz, 2008). Die in Abbildung 3 gezeigte, sogenannte Wright Map, zeigt die Verteilung der Personenfähig-

keiten und die generalisierten Schwellenwerte (Generalized Item Thresholds), d.h. den Wert, ab dem eine Schülerin bzw. ein Schüler mit höherer Wahrscheinlichkeit dem nächsthöheren Verständnisniveau zuzuordnen ist.

Ein Vergleich der Verteilung der Schwellenwerte mit der Verteilung der Schülerfähigkeiten legt nahe, dass die Aufgaben das Verständnis von Materie prinzipiell gut abbilden. Da in dieser Studie nur zehn Aufgaben eingesetzt wurden, ist eine lückenlose Abdeckung nicht zu erwarten gewesen. Die zufriedenstellende Reliabilität zeigt jedoch, dass bereits zehn Aufgaben für eine reliable Erfassung des Verständnisses bzw. von Unterschieden im Verständnis geeignet sind. Eine statistische Prüfung der Abhängigkeit der Schwellenwerte vom Verständnisniveau zeigt zudem einen signifikanten Einfluss des Verständnisniveaus auf die Schwellenwerte ($\tau = .69$, $p < .001$; $F(2,22) = 22.98$; $p < .001$; $R^2 = .676$), was darauf hinweist, dass die Schwellenwerte die theoretisch angenommenen Niveaus des Verständnisses abbilden. Als nächstes wurde untersucht, ob die OMC Aufgaben geeignet sind, eine Ent-

wicklung im Verständnis der Schülerinnen und Schüler über die Jahrgangsstufen hinweg zu beschreiben. Dazu wurde untersucht, inwieweit Schülerinnen und Schüler höherer Jahrgangsstufen eine höhere Personenfähigkeit besitzen (vgl. Abbildung 4). Dabei konnte ein signifikanter Einfluss der Jahrgangsstufe auf die Personenfähigkeit nachgewiesen werden ($\tau = .42$, $p < .001$; $F(6,287) = 28.75$; $p < .001$; $R^2 = .375$). Statistisch bedeutsame Zusammenhänge zwischen Halbjahresnoten und Personenfähigkeit konnten jedoch nicht ermittelt werden (vgl. Tabelle 1).

4.2 Identifikation von Mustern bei der Bearbeitung der Aufgaben

Mit Hilfe einer latenten Klassenanalyse (vgl. Bortz & Döring, 2006) wurde anschließend nach Mustern im Antwortverhalten der Schülerinnen und Schüler gesucht. Dabei stand die Frage im Vordergrund, welchen Einfluss das den Antwortoptionen zugeschriebene Verständnisniveau auf das Antwortverhalten hat. Gibt es z. B. eine Klasse von Schülerinnen und Schülern, die über alle Aufga-

ben hinweg Antwortoptionen wählen, die dem Verständnisniveau 3 zugeordnet sind (vgl. Abbildung 1), oder wechselt dies von Aufgabe zu Aufgabe? Theoretisch können drei Modelle angenommen werden. In einem Modell mit zwei Klassen kann sich eine Klasse von Schülerinnen und Schülern für die richtige Antwort entscheiden, die anderen variieren in der Wahl der Distraktoren (Modell 1). In der Regel wurden pro OMC Aufgabe Antwortoptionen auf drei verschiedenen Verständnisniveaus entwickelt. Diese lassen sich im Antwortmuster wiederfinden (Modell 2). Insgesamt werden durch die OMC Aufgaben vier von fünf Verständnisniveaus des Entwicklungsmodells für Materie (vgl. Abbildung 1) abgedeckt. Diese haben einen Einfluss auf die Anzahl der Klassen (Modell 3).

Ein Vergleich der Fit-Werte für die unterschiedlichen Modelle legt eine Beschreibung der Antwortmuster durch ein 3-Klassen-Modell nahe (vgl. Tabelle 2). Schülerinnen und Schüler der Klasse III entschieden sich (von zwei Aufgaben abgesehen) mit großer Wahrscheinlichkeit für die richtige Antwortoption (vgl. Tabelle 3). Schülerinnen und Schüler der

Tab. 1: Zusammenhang zwischen Personenfähigkeit und Halbjahresnoten pro Jahrgang

Klassenstufe	Mathematik	Deutsch	Englisch	Biologie	Physik	Chemie
6	-.11	-.15	-.06	-.05	-	-
7	-.30	-.12	-.24	-.50	-.28	-
8	-.12	.06	-.10	-.12	-.23	.23
9	-.33	-.12	-.27	-.28	-.40	-.02
10	.03	.07	.28	-.02	-.17	-.11
11	-.03	.04	.09	-.15	-.13	-.04
12	.18	.16	-.04	.02	.07	.45

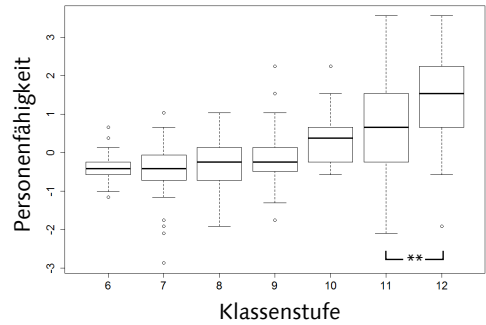
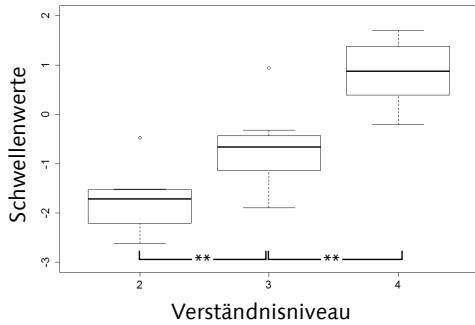


Abb. 4: Schwellenwerte über Verständnisniveau (links); Personenfähigkeit über Klassenstufe (rechts).

Klasse II bevorzugten je nach Aufgabe die richtige Antwortoption oder Distraktoren, die den mittleren Verständnisniveaus zugeordnet waren. Schülerinnen und Schüler dieser Klassen schlossen in der Regel Antwortoptionen, die dem niedrigsten Verständnisniveau zugeordnet wurden, aus (die Wahrscheinlichkeit für die Wahl einer solchen Antwortoption betrug bis auf eine Ausnahme unter 5%). Schülerinnen und Schüler der Klasse I zeichneten sich durch ein inkonsistentes Antwortmuster aus. Mit großer Wahrscheinlichkeit wählten sie Antwortoptionen, die dem Verständnisniveau 1 zugeordnet waren, oder entschieden sich für keine der Antworten. Je nach Aufgabe entschieden sie sich aber auch mit einer Wahrscheinlichkeit von teilweise über 30% für Antwortoptionen auf Niveau 2, 3 oder 4.

In Tabelle 3 wird angegeben, wie sich die Schülerinnen und Schüler der einzelnen Klassen in ihrem Antwortverhalten unterscheiden. Bezogen auf Aufgabe 4 wählen z. B. Schülerinnen und Schüler der Klasse I mit einer Wahrscheinlichkeit von 24%

Antwortoptionen, die dem Verständnisniveau 1 zugeordnet wurden und mit einer Wahrscheinlichkeit von 69% Antwortoptionen, die dem Verständnisniveau 3 zugeordnet wurden. Klasse I ist mit einem Anteil von 58% die größte der drei Klassen, gefolgt von Klasse III mit einem Anteil von 29% und Klasse II mit einem Anteil von 13%.

4.3 Vergleich von offenen und OMC Aufgaben

Zuletzt wurden die Antworten der Schülerinnen und Schüler auf die offenen Fragen mit der Wahl der Antwortoption bei OMC Aufgaben verglichen. Dazu wurden die

Tab. 2: Vergleich der unterschiedlichen Modelle zur latenten Klassenanalyse

Modell	Anzahl der Klassen	BIC-Wert
1	2	5424
2	3	5380
3	4	5460

Tab. 3: Ausgewählte Ergebnisse der latenten Klassenanalyse

Aufgabe 4	Niveau 1 Anteil in %	Niveau 2 Anteil in %	Niveau 3 Anteil in %	Aufgabe 7	Niveau 1 Anteil in %	Niveau 2 Anteil in %	Niveau 3 Anteil in %
Klasse I	24	7	69	Klasse I	41	53	6
Klasse II	4	29	67	Klasse II	4	76	17
Klasse III	0	1	99	Klasse III	1	36	63

offenen Antworten einem Verständnisniveau gemäß dem Entwicklungsmodell (vgl. Abbildung 1) zugeordnet. Die Zuordnung fand durch zwei geschulte Kodierer statt, die alle Antworten unabhängig voneinander kodierten. Die prozentuale Übereinstimmung in den Zuordnungen lag bei 94 %. Als weiteres Übereinstimmungsmaß wurde Cohens Kappa zu $\kappa = .91$ bestimmt, welches ebenfalls für eine sehr gute Übereinstimmung spricht. Auf dieser Grundlage wurde ermittelt, inwieweit man durch den Einsatz der offenen Aufgaben zu einer ähnlichen Einschätzung des Verständnisses der Schülerinnen und Schüler wie durch die OMC Aufgaben gelangt. Dies war in 78,8 % aller Antwortpaare der Fall. Cohens Kappa wurde zu $\kappa = .65$ bestimmt. Dabei zeigte sich insbesondere bei den ersten sieben Aufgaben eine sehr gute Übereinstimmung (vgl. Tabelle 4). Um zu prüfen, inwieweit die Verständnisniveaus, die den Schülerinnen und Schülern durch OMC Aufgaben und offenen Aufgaben zugewiesen wurden, voneinander abweichen, wurde zudem die Differenz beider Verständnisniveaus (VN) gebildet.

$$VN_{OMC} - VN_{offen}$$

Über alle Aufgaben ergab sich eine mittlere Differenz von $M = .14$ mit einer mitt-

leren Standardabweichung von $SD = .70$. Das heißt, in den 21,2 % aller Fälle, bei denen VN_{OMC} und VN_{offen} nicht übereinstimmen, wird den Schülerinnen und Schülern durch OMC Aufgaben im Mittel ein leicht höheres Verständnisniveau zugeschrieben. Auch hier liegt für die ersten sieben Aufgaben eine geringe Differenz der Mittelwerte vor, was zusammen mit der dazugehörigen Standardabweichung ebenfalls für eine sehr gute Übereinstimmung spricht. Detaillierte Ergebnisse pro Aufgabe sind in Tabelle 4 angegeben.

5. Diskussion

Unter Berücksichtigung der Komplexität des untersuchten Konstrukts konnte eine zufriedenstellende Reliabilität ermittelt werden (vgl. Abschnitt 4). Inwieweit es gelungen ist, Evidenz für Validität nach Messick (1995, vgl. Abschnitt 3) herzustellen, wird im Folgenden diskutiert.

Die *inhaltliche Validität* eines Testinstruments umfasst nach Messick (1995) die präzise Definition der Domäne bzw. des Konstrukts sowie die angemessene Abbildung der Domäne bzw. des Konstrukts durch Aufgaben. Das Konstrukt wurde im Vorfeld der vorliegenden Studie durch eine

umfangreiche Analyse der einschlägigen Forschungsliteratur und den Abgleich mit gängigen Curricula sichergestellt. Daraus wurde ein Modell für das Verständnis von Materie entwickelt, das auch den Aspekt der Entwicklung berücksichtigt (vgl. Hadenfeldt & Neumann, 2011). Die Passung zwischen Modell und Aufgaben wurde in der vorliegenden Studie durch ein standardisiertes Vorgehen bei der Aufgabenentwicklung sichergestellt (vgl. Wild & Möller, 2009). Abschließend wurden die Aufgaben einer Gruppe von Experten zur Bewertung vorgelegt (vgl. Abschnitt 3.1). Die sehr gute Übereinstimmung von $\kappa = .86$ spricht für die inhaltliche Validität des Testinstruments.

Kognitive Validität beschreibt den Anteil der intendierten kognitiven Prozesse bei der Bearbeitung der Aufgaben. D.h., bezogen auf die vorliegende Studie unter anderem, inwieweit die Wahl einer Antwortoption vom Verständnis von Materie oder z. B. von bestimmten Schlüsselwörtern in den Antwortoptionen abhängig ist. Ein starkes Argument für die kognitive Validität der Erfassung des Verständnisses von Materie liefert der Vergleich

der offenen Aufgaben mit den OMC Aufgaben (vgl. Tabelle 4). Die OMC Aufgaben erfassten in mindestens 60 % aller Fälle das gleiche Verständnisniveau wie die korrespondierenden offenen Aufgaben. D.h. die Schülerinnen und Schüler formulierten die Antworten in offenen Aufgaben auf dem Niveau für das sie sich auch in den korrespondierenden OMC Aufgaben entschieden haben. Bei fünf der OMC Aufgaben lag die Übereinstimmung sogar bei mehr als 80 %. Zusätzlich zeigen die Ergebnisse der latenten Klassenanalyse, dass die Aufgaben im Prinzip geeignet sind, zwischen Schülerinnen und Schülern mit geringem und hohem Verständnis zu unterscheiden. Zusammen mit dem bedeutsamen Zusammenhang zwischen Beschulung und Personenfähigkeit (vgl. Abbildung 4) legt dies nahe, dass die Schülerinnen und Schüler bei der Bearbeitung der OMC Aufgaben tatsächlich kognitive Prozesse vollzogen haben, die dem Verständnisniveau der von ihnen gewählten Aufgaben entsprachen. *Strukturelle Validität* beschreibt den Grad, zu dem die empirisch gewonnenen Daten die theoretisch erwartete Struktur eines Konstrukts abbilden. Evidenz für

Tab. 4: Vergleich OMC Aufgaben – offene Aufgaben

Aufgaben	A1	A2	A3	A4	A5	A6	A7	A8	A9	A10
Prozentuale Übereinstimmung des Verständnisniveaus ermittelt durch OMC – und korrespondierender offener Aufgabe	.89	.79	.77	.90	.84	.81	.92	.60	.67	.68
Mittelwert ($VN_{OMC} - VN_{offen}$)	-.12	.21	.28	.03	.12	-.10	.03	.06	.36	.56
Standardabweichung ($VN_{OMC} - VN_{offen}$)	.50	.55	.56	.32	.49	.54	.28	1.10	.86	.96

die strukturelle Validität des Testinstruments können die Ergebnisse der latenten Klassenanalyse liefern (vgl. Wild & Möller, 2009). Die Ergebnisse zeigen, dass Schülerinnen und Schüler durchaus je nach Aufgabe ein anderes Verständnis von Materie zeigen können. Dies ist konform zu der theoretischen Annahme, dass die Entwicklung des Verständnisses von Materie nicht in diskreten Schritten verläuft (Wilson, 2005; Steedle & Shavelson, 2009). Insbesondere die Entscheidung, Antwortoptionen auf Niveau 2 oder 3 zu wählen, ist stark von der jeweiligen Aufgabe abhängig. Solch eine Inkonsistenz im Antwortverhalten wurde auch von Johnson (2002) und Nakhleh, Samarapungavan und Saglam (2005) berichtet. Allerdings konnte in dieser Studie auch gezeigt werden, dass Schülerinnen und Schüler mit fortschreitender Beschulung in Chemie und Physik Antwortoptionen wählen, die höheren Verständnisniveaus zugeordnet sind (vgl. Abbildung 4). Dies entspricht der Erwartung, dass sich das Verständnis der Schülerinnen und Schüler von Materie im Verlauf der Schulzeit kontinuierlich gemäß dem Modell (vgl. Abbildung 1) entwickelt, was allerdings noch in einem echten Längsschnitt bestätigt werden muss.

Bezogen auf die eben genannten Aspekte von Validität konnte im Rahmen dieser Studie ausreichend Evidenz hergestellt werden. Für die noch folgenden Aspekte von Validität konnte im Rahmen dieser Studie nur teilweise oder keine Evidenz für Validität hergestellt werden. Mit Blick auf *kriteriumsorientierte Validität*, wurden im Rahmen dieser Studie die letzten Halbjahresnoten der Schülerinnen und

Schüler in den mathematisch-naturwissenschaftlichen Fächern (konvergente Validität) und in den sprachlichen Fächern (divergente Validität) erhoben. Es wurde erwartet, dass die Testleistung stärker mit den Halbjahresnoten in den mathematisch-naturwissenschaftlichen Fächern korreliert, als mit den Halbjahresnoten in den sprachlichen Fächern. Ein statistisch bedeutsamer Zusammenhang konnte jedoch nicht ermittelt werden. Dies kann zum einen daran liegen, dass die Zusammenhänge durchweg niedriger ausfallen, als bei der verhältnismäßig geringen Zahl von Schülerinnen und Schüler pro Jahrgang (im Mittel $n=42$ Schülerinnen und Schüler) erforderlich, um statistisch bedeutsame Zusammenhänge ermitteln zu können (bei einer Stichprobengröße von $n=42$ Schülerinnen und Schülern wäre mindestens ein mittlerer Zusammenhang, $r=.50$ notwendig, um ein Signifikanzniveau von $p < .05$ zu erreichen). Zum anderen wird bei der Vergabe von Schulnoten mehr als nur die Entwicklung im Verständnis von Materie berücksichtigt, z. B. der individuelle Lernfortschritt im Halbjahr und die Beteiligung im Unterricht.

Da in der vorliegenden Studie zehn Aufgaben nur insgesamt $N=294$ Schülerinnen und Schülern einer Schule mit relativ begrenztem Einzugsgebiet zur Bearbeitung vorgelegt wurden, liegt nur sehr eingeschränkt Evidenz für *Generalisierbarkeit* vor. Generalisierbarkeit hieße im Kontext dieser Studie, Evidenz dafür zu finden, dass 1) die Ergebnisse auf andere Schülergruppen übertragen werden können und 2) von der Bearbeitung der zehn OMC Aufgaben auf das Verständnis von

Materie im Allgemeinen geschlossen werden kann. Bezüglich der Übertragbarkeit auf andere Schülergruppen kann lediglich angeführt werden, dass keine Evidenz vorliegt, dass die Ergebnisse nicht auf andere Stichproben übertragbar wären. Evidenz für Übertragbarkeit müsste in einer weiteren Studie mit einer größeren Zahl von Schülerinnen und Schülern aus unterschiedlichen Schulen erwirkt werden. Bezüglich des zweiten Aspektes ist festzuhalten, dass die eingesetzten OMC Aufgaben nur für den Bereich *Struktur und Zusammensetzung* von Materie entwickelt wurden. Nach Liu und Lesniak (2005) ist die *Struktur und Zusammensetzung* nur ein Bereich von Materie, auf den bezogen Schülerinnen und Schüler ein Verständnis erwerben können. Weitere sind: *physikalische Eigenschaften und Veränderungen*, *chemische Reaktionen* und *Erhaltung* (vgl. auch Hadenfeldt & Neumann, 2011). Zwar ist nach Othman, Treagust & Chandrasegaran (2008) ein Verständnis der Struktur und Zusammensetzung ein Prädiktor für das Verständnis der weiteren Bereiche des Materiekonzepts, doch liegen bislang keine Studien vor, in denen OMC Aufgaben zur Erfassung von Verständnis in allen Bereichen eingesetzt wurden. Solche Studien sind ebenfalls Teil des dieser Studie übergeordneten Forschungsprojekts. Unter *Konsequenzen*, dem sechsten und letzten Aspekt von Validität nach Messick (1995), sollen sowohl intendierte, als auch explizit nichtintendierte Konsequenzen diskutiert werden, die aus den Testergebnissen gezogen werden können. Solche Aussagen können im Rahmen dieser Studie nur als Vermutungen formuliert

werden, da empirische Evidenz weder für positive, noch negative Konsequenzen vorliegt. Als positiv ist anzumerken, dass es im Rahmen dieser Studie gelungen ist, durch OMC Aufgaben eine Alternative zu Multiple Choice Aufgaben aufzuzeigen, durch die sich im Vergleich zu Multiple Choice Aufgaben mehr diagnostische Information generieren lässt. Für Lehrkräfte bieten OMC Aufgaben den Vorteil, auf effiziente Weise erfassen zu können, wie weit einzelne Schülerinnen und Schüler bereits in der Entwicklung ihres Materieverständnisses vorangeschritten sind und wo sie noch Schwierigkeiten haben (vgl. Briggs et al., 2006). Diese Erkenntnisse können dann bei der Unterrichtsplanung berücksichtigt werden (Lin et al., 2010). Inwieweit OMC Aufgaben in large-scale Testungen verwendet werden und hier gegebenenfalls zu einer Reduzierung der zu entwickelnden Aufgaben beitragen können, bleibt vorerst offen.

6. Zusammenfassung und Ausblick

Ziel der hier vorgestellten Studie war, zu prüfen, inwieweit OMC Aufgaben geeignet sind, das Verständnis von Schülerinnen und Schülern im Umgang mit dem Materiekonzept zu erfassen. Insgesamt konnte gezeigt werden, dass dieser Aufgabentyp erfolgreich zur reliablen und validen Erfassung von Verständnis im Umgang mit dem Materiekonzept eingesetzt werden kann. Insbesondere waren die eingesetzten OMC Aufgaben geeignet, sowohl Unterschiede im Verständnis

zu erfassen, als auch eine Entwicklung des Verständnisses im Laufe der Schulzeit zu beschreiben. Wie bereits in der Diskussion erwähnt, ist diese Studie Teil eines übergeordneten Forschungsprojekts, in dem die Entwicklung des Verständnisses von Materie im Verlauf der Schulzeit beschrieben werden soll. Wurden in dieser Studie Aufgaben vornehmlich für den Bereich *Struktur und Zusammensetzung von Materie* entwickelt, sollen in folgenden Studien weitere OMC Aufgaben entwickelt werden, durch die das Verständnis von Materie breiter erfasst werden kann. Expertenratings und Lautes-Denken-Studien (vgl. Ericsson & Simon, 1993) sollen dabei den Prozess der Aufgabenentwicklung begleiten. Diese Studien sollen, insbesondere im Hinblick auf kognitive Validität, kriteriumsorientierte Validität und Generalisierbarkeit, weitere Evidenz für die Validität von Testinstrumenten basierend auf OMC Aufgaben zur Erfassung des Verständnisses von Materie herstellen.

lung von Kompetenz im Umgang mit dem Materiekonzept“ entstanden.

Danksagung

Unser Dank gilt den Gutachterinnen und Gutachtern für die vielen hilfreichen Hinweise zur Überarbeitung des Manuskripts. Darüber hinaus danken wir Dr. Sascha Bernholt und Prof. Dr. Ilka Parchmann für die Unterstützung bei der Planung und Durchführung der Studie sowie die konstruktiven Rückmeldungen zu früheren Versionen des Manuskripts.

Dieser Beitrag ist im Rahmen des vom Bundesministerium für Bildung und Forschung geförderten Projektes „Entwick-

Literatur

- Adadan, E., Irving, K. E., & Trundle, K. C. (2009). Impacts of Multi-representational Instruction on High School Students' Conceptual Understandings of the Particulate Nature of Matter. *International Journal of Science Education*, 31(13), 1743–1775.
- Adadan, E., Trundle, K. C., & Irving, K. E. (2010). Exploring Grade 11 students' conceptual pathways of the particulate nature of matter in the context of multirepresentational instruction. *Journal of Research in Science Teaching*, 47(8), 1004–1035.
- Adbo, K., & Taber, K. S. (2009). Learners' Mental Models of the Particle Nature of Matter: A study of 16-year-old Swedish science students. *International Journal of Science Education*, 31(6), 757–786.
- Ahtee, M., & Varjola, I. (1998). Students' understanding of chemical reaction. *International Journal of Science Education*, 20(3), 305–316.
- Alonzo, A. C., & Steedle, J. T. (2009). Developing and assessing a force and motion learning progression. *Science Education*, 93(3), 389–421.
- American Association for the Advancement of Science (2001). *Atlas of Science Literacy*. New York: Oxford University.
- American Association for the Advancement of Science (2007). Getting Assessment Right. *2061 today*, 17(1), 2–7.
- American Psychological Association (2001). *Publication Manual of the American Psychological Association* (Vol. 5). Washington, DC: American Psychological Association.
- Andersson, B. (1990). Pupils' Conceptions of Matter and its Transformations (age 12–16). *Studies in Science Education*, 18, 53–85.
- Barke, H.-D. (2006). *Chemiedidaktik: Diagnose und Korrektur von Schülervorstellungen*. Berlin: Springer.
- Baumert, J., Bos, W., & Watermann, R. (1998). TIMSS/III. Schülerleistungen in Mathematik und den Naturwissenschaften am Ende der Sekundarstufe II im internationalen Vergleich.: Zusammenfassung deskriptiver Ergebnisse. Studien und Berichte. Berlin: Max-Planck-Institut für Bildungsforschung.
- Baumert, J., Lehmann, R., Lehrke, M., Schmitz, B., Clausen, M., & Hosenfeld, I., et al. (Eds.) (1997). TIMSS. Mathematisch-naturwissenschaftlicher Unterricht im internationalen Vergleich. Deskriptive Befunde. Opladen: Leske+Budrich.
- Berlak, H. (1992). The need for a new science of assessment. In H. Berlak et al. (Eds.), *Toward a new science of educational testing and assessment* (pp. 1–22). Albany: State University of New York Press.
- Bernholt, S., Commons, M. L., & Parchmann, I. (2009). Kompetenzmodellierung zwischen Forschung und Unterrichtspraxis. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 219–245.
- Bond, T. G., & Fox, C. M. (2001). *Applying the Rasch Model: Fundamental measurement in the human sciences*. Mahawah: Lawrence Erlbaum.
- Bortz, J. (2005). *Statistik für Human- und Sozialwissenschaftler: Mit 242 Tabellen* (6., vollst. überarb. und aktualisierte Aufl.). Berlin ; Heidelberg ; New York: Springer.
- Bortz, J., & Döring, N. (2006). *Forschungsmethoden und Evaluation: Für Human- und Sozialwissenschaftler* (4., überarbeitete Auflage.). Springer-Lehrbuch. Berlin, Heidelberg: Springer.
- Briggs, D. C., Alonzo, A. C., Schwab, C., & Wilson, M. (2006). Diagnostic Assessment with Ordered Multiple-Choice Items. *Educational Assessment*, 11(1), 33–63.
- Buck, P., Johnson, P., Fischler, H., Peuckert, J., & Seifert, S. (2001). The need for and the role of metacognition in teaching and learning the particle model. In H. Behrendt, H. Dahncke, R. Duit, W. Graeber, M. Komorek, & A. Kross (Eds.), *Research in Science Education – Past, Present, and Future* (pp. 225–234). Dordrecht, The Netherlands: Kluwer Academic Publishers.
- Duit, R. (2008). Students' and teachers' Conceptions and Science Education (STCSE) (Bibliographie). Retrieved 12.03.2010, from IPN.

- Duschl, R. A., Schweingruber, H. A., & Shouse, A. W. (2007). *Taking science to school: Learning and teaching science in grades K-8*. Washington D.C.: National Academies Press.
- Ericsson, K. A., & Simon, H. A. (1993). *Protocol analysis: Verbal reports as data* (Rev. ed.). Cambridge, Mass: MIT Press.
- Gómez, E. J., Benarroch, A., & Marín, N. (2006). Evaluation of the degree of coherence found in students' conceptions concerning the particulate nature of matter. *Journal of Research in Science Teaching*, 43(6), 577–598.
- Hadenfeldt, J. C., Neumann, K. (2011). Materie verstehen – aber wie?. *NiU Chemie*, 22(4+5), S.87–91.
- Hartig, J. (2008). Psychometric models for the assessment of competencies. In J. Hartig (Ed.), *Assessment of competencies in educational contexts* (pp. 69–90). Cambridge, Mass.: Hogrefe.
- Hartig, J., Frey, A., & Jude, N. (2007). Validität. In H. Moosbrugger & A. Kelava (Eds.), *Testtheorie und Fragebogenkonstruktion* (pp. 135–163). Springer.
- Johnson, P. (1998). Progression in children's understanding of a basic particle theory: A longitudinal study. *International Journal of Science Education*, 20(4), 393–412.
- Johnson, P. (2002). Children's understanding of substances, Part 2: explaining chemical change. *International Journal of Science Education*, 24(10), 1037–1054.
- Kauertz, A. (2008). *Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben. Studien zum Physik- und Chemielernen: Vol. 79*. Berlin: Logos.
- Kind, V. (2004). *Beyond Appearances: Students' misconceptions about basic chemical ideas*. Retrieved 27.05.2010, from http://www.rsc.org/images/Misconceptions_update_tcm18-188603.pdf
- Kircher, E., Girwidz, R., & Häußler, P. (2010). *Physikdidaktik*. Berlin, Heidelberg: Springer Berlin Heidelberg.
- Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current Issues in Competence Modeling and Assessment. *Zeitschrift für Psychologie / Journal of Psychology*, 216(2), 61–73.
- Krnel, D., & Glazar, S. S., Watson, R. (2003). The development of the concept of "Matter": A cross-age study of how children classify materials. *Science Education*, 87(5), 621–639.
- Krnel, D., Watson, R., & Glazar, S. A. (1998). Survey of research related to the development of the concept of 'matter'. *International Journal of Science Education*, 20(3), 257–289.
- Lin, J., Chu, K., & Meng, Y. (2010). Distractor Rationale Taxonomy: Diagnostic Assessment of Reading with Ordered Multiple-Choice Items. Paper presented at the annual meeting of AERA, Denver, Colorado.
- Liu, X. (2001). Synthesizing research on student conceptions in science. *International Journal of Science Education*, 23(1), 55–82.
- Liu, X. (2010). *Using and Developing Measurement Instruments in Science Education: A Rasch Modeling Approach (Hc)*: Iap.
- Liu, X., & Lesniak, K. M. (2005). Students' progression of understanding the matter concept from elementary to high school. *Science Education*, 89(3), 433–450.
- Löfgren, L., & Helldén, G. (2009). A Longitudinal Study Showing how Students use a Molecule Concept when Explaining Everyday Situations. *International Journal of Science Education*, 31(12), 1631–1655.
- Margel, H., Eylon, B.-S., & Scherz, Z. (2008). A longitudinal study of junior high school students' conceptions of the structure of materials. *Journal of Research in Science Teaching*, 45(1), 132–152.
- Messick, S. (1995). Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning. *American Psychologist*, 50(9), 741–749.
- Ministerium für Bildung, W. F. u. K. L. S.-H. (1997a). *Lehrplan für die Sekundarstufe I der weiterführenden allgemeinbildenden Schulen Hauptschule, Realschule, Gymnasium: Physik*. Glückstadt: Glückstädter Werkstätten.

- Ministerium für Bildung, W. F. u. K. L. S.-H. (1997b). *Lehrplan für die Sekundarstufe I der weiterführenden allgemeinbildenden Schulen Hauptschule, Realschule, Gymnasium: Chemie*. Glückstadt: Glückstädter Werkstätten.
- Mohan, L., Chen, J., & Anderson, C. W. (2009). Developing a multi-year learning progression for carbon cycling in socio-ecological systems. *Journal of Research in Science Teaching*, 46(6), 675–698.
- Nakhleh, M. B., & Samarapungavan, A. (1999). Elementary school children's beliefs about matter. *Journal of Research in Science Teaching*, 36(7), 777–805.
- Nakhleh, M. B., Samarapungavan, A., & Saglam, Y. (2005). Middle school students' beliefs about matter. *Journal of Research in Science Teaching*, 42(5), 581–612.
- National Assessment Governing Board, & American Institutes for Research (2008). *Science framework for the 2009 National Assessment of Educational Progress: National Assessment Governing Board, US Dept. of Education*.
- National Research Council (1996). *National Science Education Standards*. Washington, D.C: National Academy Press.
- Neumann, K., Kauertz, A., Lau, A., Notarp, H., & Fischer, H. E. (2007). Die Modellierung physikalischer Kompetenz und ihrer Entwicklung. *Zeitschrift für Didaktik der Naturwissenschaften*, 13, 103–123.
- Neumann, K., Fischer, H. E., & Sumfleth, E. (2008). Vertikale Vernetzung und kumulatives Lernen im Chemie- und Physikunterricht. In E.-M. Lankes (Hrsg.), *Pädagogische Professionalität als Gegenstand empirischer Forschung* (S. 141–152). Münster: Waxmann.
- Othman, J., Tregust, D. F., & Chandrasegaran, A. L. (2008). An investigation into the relationship between students' conceptions of the particulate nature of matter and their understanding of chemical bonding. *International Journal of Science Education*, 30(11), 1531–1550.
- Papageorgiou, G., & Johnson, P. (2005). Do particle ideas help or hinder pupils' understanding of phenomena? *International Journal of Science Education*, 27(11), 1299–1317.
- Park, E. J., & Light, G. (2009). Identifying Atomic Structure as a Threshold Concept: Student mental models and troublesomeness. *International Journal of Science Education*, 31(2), 233–258.
- Peper, M., Schmidt, S., Wilms, M., Oetken, M., & Parchmann, I. (2007). Modellvorstellungen entwickeln und anwenden: Einsatz von Medien, Alltagsphänomenen und Experimenten. *Naturwissenschaften im Unterricht Chemie*, 18, 17–22.
- PISA-Konsortium Deutschland (Ed.) (2006). *PISA 2003. Dokumentation der Ergebnisse*. Münster: Waxmann.
- PISA-Konsortium Deutschland (Ed.) (2007). *PISA 2006. Die Ergebnisse der dritten internationalen Vergleichsstudie*. Münster: Waxmann.
- Rasch, G. (1960). *Probabilistic models for some intelligence and attainment tests*. Copenhagen: Danish Institute for Educational Research.
- Ruiz-Primo, M. A., & Shavelson, R. J. (1996). Rhetoric and Reality in Science Performance Assessments: An Update. *Journal of Research in Science Teaching*, 33(10), 1045–1063.
- Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK] (2005a). *Bildungsstandards im Fach Biologie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK] (2005b). *Bildungsstandards im Fach Chemie für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München: Luchterhand.
- Sekretariat der Ständigen Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland [KMK] (2005c). *Bildungsstandards im Fach Physik für den Mittleren Schulabschluss. Beschluss vom 16.12.2004*. München: Luchterhand.

- Shavelson, R.J., Carey, N.B., & Webb, N.M. (1990). Indicators of science achievement: Options for a powerful policy instrument. *Phi Delta Kappan*, 71, 692–697.
- Smith, C., Wiser, M., Anderson, C. W., Krajcik, J., & Coppola (2006). Implications of research on children's learning for assessment: Matter and atomic molecular theory. *Measurement*, 14(1 & 2), 1–98.
- Steedle, J. T., & Shavelson, R. J. (2009). Supporting valid interpretations of learning progression level diagnoses. *Journal of Research in Science Teaching*, 46(6), 699–715.
- Stevens, S. Y., Delgado, C., & Krajcik, J. S. (2010). Developing a hypothetical multi-dimensional learning progression for the nature of matter. *Journal of Research in Science Teaching*, 47(6), 687–715.
- Talanquer, V. (2006). Commonsense chemistry: A model for understanding students' alternative conceptions. *Journal of Chemical Education*, 83(5), 811.
- Talanquer, V. (2009). On Cognitive Constraints and Learning Progressions: The case of "structure of matter". *International Journal of Science Education*, 31(15), 2123–2136.
- Wainer, H., & Thissen, D. (1993). Combining Multiple-Choice and Constructed-Response Test Scores: Toward a Marxist Theory of Test Construction. *Applied Measurement in Education*, 6(2), 103–118.
- Wild, E., & Möller, J. (Eds.) (2009). *Pädagogische Psychologie* (1st ed.). Berlin: Springer.
- Wilson, M. (2005). *Constructing Measures: An Item Response Modeling Approach*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität. Methoden zur Bestimmung und Verbesserung der Zuverlässigkeit von Einschätzungen mittels Kategoriensystemen und Ratingskalen*. Göttingen: Hogrefe.

KONTAKT

Jan Christoph Hadenfeldt
Leibniz-Institut für die Pädagogik
der Naturwissenschaften (IPN) Kiel
Didaktik der Physik
Olshausenstrasse 62
24098 Kiel
hadenfeldt@ipn.uni-kiel.de

AUTORENINFORMATION

Jan Christoph Hadenfeldt ist Studienrat an der Kieler Gelehrtenschule. Seit 2009 ist er teilabgeordneter Wissenschaftlicher Mitarbeiter in der Abteilung Didaktik der Physik am IPN Kiel. Dort untersucht er die Entwicklung des Verständnisses im Umgang mit dem Materiekonzept.

Dr. paed. Knut Neumann ist Professor für Didaktik der Physik am Leibniz-Institut für die Pädagogik der Naturwissenschaften und der Mathematik (IPN) in Kiel. Seine Forschungsschwerpunkte liegen im Bereich der Diagnose und Entwicklung physikalischer Kompetenz sowie Untersuchungen zur Unterrichtsqualität in Physik.