

Erhebung von repräsentationaler Kohärenzfähigkeit von Schülerinnen und Schülern im Themenbereich Strahlenoptik

Jochen Scheid¹  · Andreas Müller² · Rosa Hettmannsperger³ · Jochen Kuhn⁴

Eingegangen: 29. Mai 2015 / Angenommen: 2. Juni 2017 / Online publiziert: 15. Juni 2017

© Gesellschaft für Didaktik der Physik und Chemie (GDPC); Fachsektion Didaktik der Biologie im VBIO (FDdB im VBIO) and Springer-Verlag GmbH Germany 2017

Zusammenfassung Der Beitrag stellt die Theorie, die Operationalisierung und psychometrische Analyse eines Instruments zur Erhebung von repräsentationaler Kohärenzfähigkeit vor. Eine umfangreiche naturwissenschaftsdidaktische und pädagogisch-psychologische Literatur belegt, dass der Umgang mit multiplen Repräsentationen und deren kognitiven Verbindungen Grundlage für die Erlangung eines angemessenen Verständnisses naturwissenschaftlicher Experimente, Phänomene und Konzepte ist. Eine Reihe von Studien hat jedoch gezeigt, dass Lernende verschiedener Ausbildungsstufen geringe repräsentationale Fähigkeiten besitzen. Insbesondere spielt die repräsentationale Kohärenzfähigkeit, d. h. Informationen innerhalb oder zwischen multiplen Repräsentationen korrekt und zielführend auf einander zu beziehen, vergleichen, ändern oder ineinander überset-

zen zu können, eine zentrale Rolle für deren Nutzung als domänenspezifisches Denkwerkzeug. Ihr Erwerb stellt nach vorliegenden Befunden zugleich eine besonders ausgeprägte Lernschwierigkeit dar. Gleichzeitig fehlen derzeit einschlägige Erhebungsinstrumente für die meisten Themenbereiche der Physik und des Physikunterrichtes. Es wurde daher ein Erhebungsinstrument für repräsentationale Kohärenzfähigkeit im Bereich der Strahlenoptik entwickelt. Dies ist ein an multiplen Repräsentationen besonders reiches und in vielen Lehrplänen der Sekundarstufe I früh angesiedeltes Thema, sodass diesbezügliches Lernen und dessen verlässliche Prüfung eine hervorgehobene und ggf. auch weiter reichende Bedeutung haben. 488 Lernende zwischen 12 und 14 Jahren bearbeiteten den Test an rheinland-pfälzischen Gymnasien jeweils zu drei Zeitpunkten. Das Testinstrument (14 Items) erwies sich als ausreichend reliabel ($\alpha_C = 0,8$; $r_{it} \geq 0,3$) und wurde in einem Expertenrating als curricular valide eingeschätzt ($N = 11$ Lehrende, Intraklassen-Korrelation $0,5 < ICC < 0,7$). Auf dieser Grundlage kann die repräsentationale Kohärenzfähigkeit als eine Basiskompetenz für das Verständnis von wissenschaftlichen Experimenten und Konzepten quantitativ bestimmt werden und es steht ein diagnostisches Werkzeug für die fachdidaktische Forschung und die Unterrichtspraxis zur Verfügung.

Zusatzmaterial online Zusätzliche Informationen sind in der Online-Version dieses Artikels (doi: [10.1007/s40573-017-0065-4](https://doi.org/10.1007/s40573-017-0065-4)) enthalten.

✉ Jochen Scheid
scheid@uni-landau.de

- ¹ Campus Landau, FB7: Natur- und Umweltwissenschaften, Institut für naturwissenschaftliche Bildung (InB), AG Physikdidaktik, Universität Koblenz-Landau, Fortstraße 7, 76829 Landau, Deutschland
- ² Faculté des Sciences/Section Physique, Institut Universitaire de Formation des Enseignants (IUFE), Pavillon d'Uni Mail (IUFE), Université de Genève, Boulevard du Pont d'Arve 40, 1211 Genève, Schweiz
- ³ Fachbereich Erziehungswissenschaften, Institut für Pädagogik der Elementar- und Primarstufe, Goethe Universität Frankfurt am Main, Theodor-W.-Adorno-Platz 6, 60323 Frankfurt am Main, Deutschland
- ⁴ Fachbereich Physik, Didaktik der Physik, Universität Kaiserslautern, Erwin-Schrödinger-Str., Bau 46, Raum 525, 67663 Kaiserslautern, Deutschland

Schlüsselwörter Multiple Repräsentationen · Kohärenztest · Strahlenoptik

Assessment of Students' Representational Coherence Ability in the Field of Ray Optics

Abstract This contribution contains the theory, the operationalization and the empirical psychometrical analysis of an instrument for Representational Coherence Ability.

Science education and pedagogical-psychological research have shown the importance of multiple representations and their cognitive connections for an adequate understanding of scientific experiments, phenomena, and concepts. However, several further studies have shown that learners of different age groups have low levels of representational abilities. Especially Representational Coherence Ability as the ability to correctly connect, relate, compare and change information within one or several representations is central for the use of multiple representations as a tool for reasoning; but empirical results have shown that development of Representational Coherence Ability is a very distinctive learning difficulty. Also, for most topics in science education, there is a lack of appropriate assessment instruments. Therefore, in this study, an instrument to measure students' Representational Coherence Ability in the domain of ray optics was developed. Ray optics is very rich in multiple representations and placed relatively early in the curriculum of secondary level I. So learning and the reliable assessment of it can have a special meaning or wider influence for several subsequent learning processes. The study took place in Germany, Rhineland-Palatinate, where 488 grammar-school students, aged between 12 and 14, worked with the instrument at three measurement times. The instrument (14 items) showed an acceptable reliability ($\alpha_C = 0.8$; $r_{it} \geq 0.3$) and an expert rating with 11 teachers established its curricular validity ($0.5 < ICC < 0.7$ for intraclass correlation). On that base, it is possible to assess Representational Coherence Ability as a basic competence for the understanding of science experiments and concepts and thus to serve as a diagnostic tool for physics education research and classroom practice.

Keywords Multiple representations · Test for coherence · Ray optics

Einleitung und Forschungshintergrund

Dieser Abschnitt stellt zunächst die große Bedeutung multipler Repräsentationen für die Naturwissenschaften und

den naturwissenschaftlichen Unterricht dar. Anschließend wird auf notwendige Zusammenhänge mit allgemeinen kognitionspsychologischen bzw. enger auf das Lernen bezogenen Theorien eingegangen (Schnotz 2002, 2005; Schnotz und Bannert 2003; bzw. Ainsworth 1999, 2006a). Auf dieser Basis erfolgen die Diskussion des Begriffs der repräsentationalen Kohärenzfähigkeit und die Begründung der Notwendigkeit eines Instrumentes zu deren Messung.

Repräsentationen und Multiple Repräsentationen

Unter einer „Repräsentation“ versteht man eine stellvertretende Darstellung jenes Gegenstandes oder Sachverhaltes (des Referenten), auf den sich die Repräsentation bezieht (Peterson 1996). Repräsentationen können extern (z. B. auf Papier gedruckt) oder intern bzw. mental (im Geiste) vorliegen und werden in depiktive (bildliche) und deskriptive (beschreibende, Schnotz 2002) „Repräsentationsformen“ eingeteilt. Depiktive Repräsentationsformen nutzen Bildzeichen (vgl. Peirce 1958 [1931]) und stehen mit dem Referenten in einer Beziehung realistischer oder abstrakter struktureller Ähnlichkeit (Schnotz 2002). Dazu gehören z. B. die Repräsentationsform „realistisches Bild“, das einen Experimentalaufbau zeigt, die Repräsentationsform „schematisches Bild“, welches den Verlauf von Strahlen beschreibt (s. Abb. 1) und als Beispiele abstrakterer struktureller Ähnlichkeit die Repräsentationsformen „Graph“ oder „Diagramm“. Sie haben Vorteile bei Bereitstellung und Ablesbarkeit von konkreten und spezifischen Informationen (Schnotz 2002). Für deskriptive Repräsentationsformen besteht eine solche Ähnlichkeitsbeziehung nicht, sie nutzen auf Konventionen beruhende Symbolzeichen (vgl. Peirce 1958 [1931]), die ein Objekt beschreiben (Buchstaben, Zahlen etc.). Ihnen sind Repräsentationsformen wie gesprochene oder geschriebene Texte, Formeln, Tabellen und logische Ausdrücke zugeordnet (Schnotz 2002). Sie eignen sich gut für logische Schlussfolgerungen (Schnotz 2002). Verschiedene Repräsentationsformen des gleichen Gegenstandes können unterschiedliche Informationen beinhalten und unterschiedlich nützlich zur Lösung von Aufgaben sein (Schnotz 1994; Herrmann 1993; Schnotz et al. 2011). Als

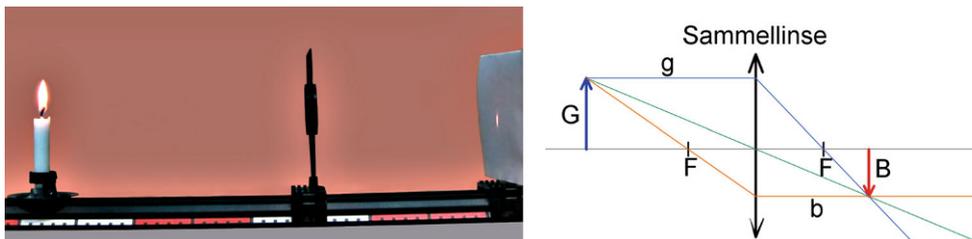


Abb. 1 Beispiele von bildhaften (depiktionalen) Repräsentationen: *links* ein realistisches Bild eines Versuchsaufbaus zur Bildentstehung und -konstruktion bei der Sammellinse. *Rechts* eine schematische Zeichnung, die den Verlauf von drei ausgewählten Strahlen angibt (G Gegenstandsgröße, B Bildgröße, g Gegenstandsweite, b Bildweite, F Brennpunkt)

Erweiterung der Arbeit mit einzelnen Repräsentationsformen wird insbesondere dem gleichzeitigen Arbeiten mit mehreren extern zum gleichen Sachverhalt vorliegenden oder intern zu konstruierenden, separaten Repräsentationsformen („multiple Repräsentation“, vgl. Schnotz 2002, 2005) eine Lernförderlichkeit zugeschrieben (vgl. Mayer 2005, „Multimediaprinzip“: verbessertes Lernen bei Verwendung von Text-Bild Kombinationen im Vergleich zur Verwendung von Text alleine). Der Begriff der „Kohärenz zwischen multiplen Repräsentationen“ (und der diesbezüglichen Fähigkeit dazu) wird weiter unten eingeführt. Die wichtige Rolle (multipler) Repräsentationen für naturwissenschaftliches Denken und Lernen ist wohl bekannt für die Naturwissenschaften im Allgemeinen (Tytler et al. 2013), für Biologie, Chemie, und Physik (Tsui und Treagust 2013; Gilbert und Treagust 2009; Docktor und Mestre 2014), die Geowissenschaften (Sell et al. 2006), und auch die Mathematik (Lesh et al. 1987; Even 1998). Auch eine umfangreiche Forschung in der kognitiven und pädagogischen Psychologie (Mandl und Levin 1989; Van Someren et al. 1998; Held et al. 2006; Verschaffel et al. 2010) belegt ihre hohe Bedeutung.

Wie Repräsentationsformen kognitiv verarbeitet werden, kann mit der *cognitive theory of multimedia-learning* (Mayer 2005) oder dem ihr verwandten integrierten Modell des Text- und Bildverstehens (Schnotz und Bannert 2003; Schnotz 2002, 2005) erklärt werden. Letzteres bietet eine alternative Sicht auf *multimedia learning*, in der repräsentationale Prinzipien stärker betont werden (Schnotz und Kürschner 2007). Es wird die Idee eines nach verbalem und des bildlichem Kanal getrennten kognitiven Untersystems (*dual coding*, Paivio 1986) mit zugehörigen beiden jeweils begrenzten Teilen eines Arbeitsgedächtnisses aufgegriffen (Baddeley 1986; Chandler und Sweller 1992). Dabei werden aus wahrgenommenen äußeren verbalen oder visuellen Reizen durch Herstellung von „*representational connections*“ mentale Repräsentationen der gleichen Form gebildet (Paivio 1986; Mayer und Anderson 1991). Verbindungen gleicher Repräsentationsformen können mental durch „*associative within system interconnections*“ gebildet werden (Paivio 1986), welche wir zusammengefasst für mentalen und externen Fall als „*intra-representational connections (ICs¹)*“ bezeichnen. Die verbalen und visuellen Repräsentationen können durch „*referential connections*“ (RCs) miteinander verbunden werden (Paivio 1986; Mayer

und Anderson 1991). Durch die Verteilung der Informationen auf die beiden Kanäle begrenzter Kapazität und der damit verbundenen besseren Kapazitätsausnutzung kann schon eine Vereinfachung des Lernens von Text durch Verwendung von Bildern erklärt werden. Das Verarbeiten der gleichen Informationsmenge in nur einem der beiden begrenzten Kanäle würde zu einer hohen Belastung des Arbeitsgedächtnisses führen, bei einer Überlastung ist das dem Lernen abträglich (zu hoher „*cognitive load*“, vgl. Sweller 1999; Chandler und Sweller 1992; Van Merriënboer und Sweller 2005; Schnotz und Kürschner 2007). Mit *cognitive load* ist die Belastung des Arbeitsgedächtnisses durch die Anzahl der gleichzeitig vorliegenden Informationselemente (Van Merriënboer und Sweller 2005), und durch deren kognitive Verarbeitung (Schnotz 2010) gemeint. Zu bemerken ist, dass das *dual coding* (im Sinne von Paivio 1986; Mayer 2005; Schnotz 2005) nach Ausschöpfung der Kapazitäten beider Kanäle auch begrenzt ist, aber auf höherem Niveau als bei Nutzung nur eines Kanals. Das integrierte Modell des Text- und Bildverstehens von Schnotz erklärt das kognitive Verarbeiten von Repräsentationsformen detailliert (die o. g. Begriffe „IC“ und „RC“ sind von großer Bedeutung für unsere Arbeit und werden daher in die folgenden Abschnitte zur Verdeutlichung integriert).

Das Modell nach Schnotz (2005) enthält zwei Kanäle begrenzter Kapazität, in dem die entsprechenden Repräsentationsformen verarbeitet werden: einen deskriptiven Kanal, in dem die Informationen durch Analyse von Symbolstrukturen verarbeitet werden, und einen depiktiven Kanal, in dem Informationen durch Prozesse der analogen Strukturabbildung verarbeitet werden. Nach Wahrnehmung einer externen deskriptiven Repräsentationsform wird auf der deskriptiven Seite des Modells eine interne mentale Oberflächenrepräsentation des Inhaltes gebildet. Daraus wird dann nach Selektion von Informationen eine interne propositionale Repräsentation des semantischen Inhalts als Beschreibung mittels Symbolen gebildet (vgl. IC). Dabei werden, je nach Fähigkeit der Person zu mehr oder weniger konsistenten mentalen Repräsentationen führend, kognitive Schemata aus dem Langzeitgedächtnis aktiviert (bzw. müssen erst im Arbeitsgedächtnis erzeugt werden). Diese beschreiben, wie Selektion und Organisation von Informationen auszuführen sind. Aus der propositionalen Repräsentation entsteht durch Modellkonstruktion (vgl. RC) ein mentales Modell (mentale Struktur, die in Struktur- oder Funktionsanalogie mit dem dargestellten Gegenstand steht). Wird eine externe depiktive Repräsentationsform wahrgenommen, entsteht eine interne visuelle Wahrnehmung der depiktiven Repräsentationsform. Daraus wird auf Basis von kognitiven Schemata nach Selektion von Informationen zuerst ein mentales Modell des Gegenstands gebildet (vgl. IC) und dann durch Modellinspektion eine interne propositionale Repräsentation erzeugt (vgl. RC). Durch Modellinspektion

¹ Wegen des häufigen Gebrauchs (und z. T. der späteren Verwendung als Indizes) war es sinnvoll, mehrere Abkürzungen einzuführen, für die hier als Lesehilfe ein Überblick gegeben wird: RKF-Test = Test auf repräsentationale Kohärenzfähigkeit; IC/RC/RFC = *intra-representational/referential/representation form connection*, RF = Repräsentationsform (treten alle als Indizes der entsprechenden Anzahlen auf); DU = *construct deeper understanding* (charakterisiert einen entsprechenden Itemtyp des Tests).

und -konstruktion werden bei der simultanen Verwendung von deskriptiven und depiktiven externen Repräsentationsformen Informationen zusammengeführt und können von einer in die andere übersetzt und integriert werden. Propositionale Repräsentationen und mentale Modelle der Referenten von externen Repräsentationsformen, können auf Basis von kognitiven Schemata miteinander abgeglichen werden (vgl. RC; und danach in äußere Repräsentationsformen überführt werden). Diese konzeptuelle Organisation führt, je nach Verfügbarkeit von kognitiven Schemata mehr oder weniger erfolgreich, zu Konsistenz von internen multiplen Repräsentationen. Wird nur eine Repräsentationsform verwendet (beispielsweise ein Strahlendiagramm als depiktive Repräsentationsform), können Informationen daraus nach Bildung des mentalen Modells durch Modellinspektion über eine Proposition in eine externe deskriptive Repräsentationsform (z. B. einen Text oder eine Formel) übersetzt werden (vgl. RC). Beim Üben werden die genannten spezifischen kognitiven Schemata gebildet und im Langzeitgedächtnis gespeichert. Ob diese Schemata aus dem Langzeitgedächtnis abgerufen werden können oder erst im Arbeitsgedächtnis gebildet werden müssen, hat einen Einfluss auf die Anzahl der zur Verfügung stehenden Schemata für Bildung und Organisation von interner propositionaler Repräsentation, mentalem Modell, Oberflächenrepräsentationen sowie die Selektionen durch die Filter (vgl. Schnotz und Bannert 2003). Sind wenig kognitive Schemata abrufbar und müssen deshalb erst bei Verarbeitung einer Aufgabe gebildet werden, sinkt dadurch die Speicherkapazität der Kanäle für miteinander zusammenhängende Informationen, das Arbeitsgedächtnis wird überlastet (zu hoher *cognitive load*, s. oben). Anspruchsvolle repräsentationale Aufgaben sind dann weniger gut lösbar.

Ein Anspruch an die Lernenden ist, zu erkennen, welche Informationen aus Repräsentationsformen abgelesen werden müssen. Damit zusammenhängend ist der Realitätsgrad ein wichtiger Parameter von Repräsentationsformen, der ein Spektrum von fotografischen über realistische bis zu schematischen Repräsentationsformen umfasst (Höffler und Leutner 2007; Niegemann et al. 2008), dabei variiert regelmäßig auch die Detailgenauigkeit (Niegemann et al. 2008). Bei schematischen Zeichnungen können für das Verstehen wesentliche Aspekte hervorgehoben werden. Realistische Bilder sind hingegen zwar vollständiger im Informationsgehalt, sie können jedoch dadurch eine Vielzahl auch irrelevanter Informationen enthalten. Dadurch werden an den Lerner höhere Ansprüche bezüglich der Identifizierung und Auswahl relevanter Aspekte gestellt (ebenda). Plötzner und Lowe (2012) verwenden in ihrem systematischen Forschungsüberblick den Abstraktionsgrad als etwas weiter gefassten Begriff, der von depiktionalen Repräsentationsformen unterschiedlichen Realitätsgrades bis zu mathema-

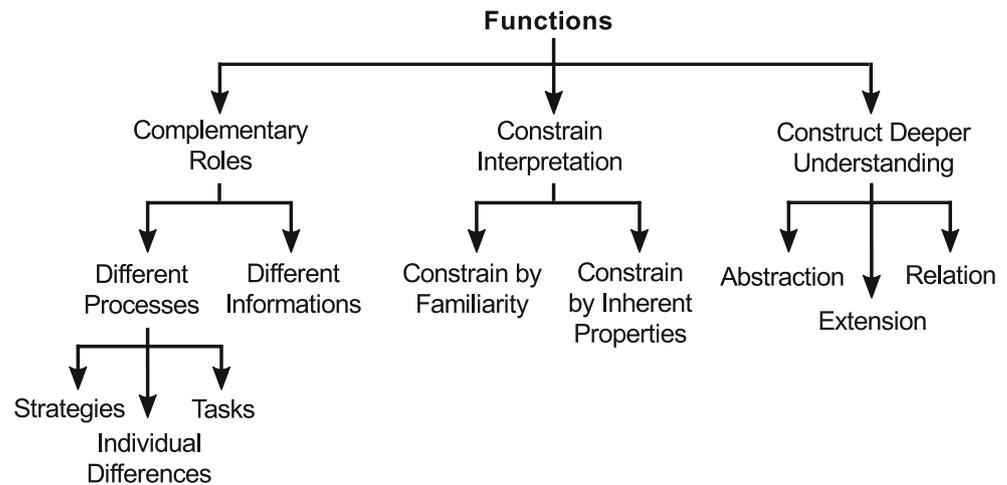
tischen Formeln als abstrakteste Form deskriptiver Repräsentationsformen reicht.

Ein weiteres zentrales Merkmal von multiplen Repräsentationen ist die Anzahl der vorkommenden Repräsentationsformen (Ainsworth 2006a) und damit auch von deren Verbindungen. Beispielsweise wachsen mit der Hinzunahme der Repräsentationsform „Formel“ (etwa in der Abbildungsgleichung) zu den beiden in Abb. 1 dargestellten bildhaften Repräsentationsformen der Informationsgehalt, aber auch die kognitiven Anforderungen. Dabei sei darauf hingewiesen, dass die *Multiplizität* im Begriff *multiple* Repräsentationen nicht ausschließlich an verschiedene Repräsentationsformen gebunden ist: z. B. könnte in einer Situation wie der von Abb. 1 auch ein zusätzlicher Vergleich zu einem weiteren Strahlendiagramm (gleiche Repräsentationsform) die Funktionalität erhöhen. Zugleich erhöht sich aber auch die Beanspruchung des begrenzten Arbeitsgedächtnisses (vgl. Baddeley 1986; Chandler und Sweller 1992; Pajvio 1986), und dies ist ein ganz zentraler Gesichtspunkt für das Lernen mit multiplen Repräsentationen. Im nächsten Abschnitt wenden wir uns einer genaueren theoretischen Beschreibung von kognitiven Funktionen und Effekten von multiplen Repräsentationen zu.

Kognitive Funktionen und Effekte von multiplen Repräsentationen

Die Kognitionsforschung, insbesondere das erwähnte integrierte Modell des Text- und Bild-Verstehens (Schnotz 2002, 2005; Schnotz und Bannert 2003) bietet einen theoretischen Rahmen, die verschiedenen kognitiven Funktionen verschiedener Repräsentationen alleine und in Wechselwirkung zu verstehen: Eine externe Repräsentation kann die kognitive Anstrengung reduzieren, welche beim Lösen einer gegebenen Aufgabe aufgebracht werden muss, da die enthaltenen Informationen nicht alle mental im Gedächtnis gehalten werden müssen (s. a. 2006a). Und sie kann, durch die Bereitstellung zeitlich stabiler und strukturell wesentlicher Information, domänenspezifische Denk- und Lernprozesse unterstützen. Je nach Repräsentationsform können sich dabei Art und Menge der externalisierten Information unterscheiden, ebenso, die Art der ausgelösten kognitiven Prozesse. Eine Darstellung von Informationen in einer anderen Repräsentationsform, welche die gleiche abstrakte Struktur hat (Re-Repräsentation) beeinflusst das Aufgabenlösen, z. B. wenn mehr Information externalisiert wird (Zhang und Norman 1994), oder die Informationen anders gruppiert (2006a) oder komplementär sind (Schnotz 2002, 2005). Das heißt, es ist regelmäßig nicht ausreichend, nur mit einer einzelnen Repräsentationsform zu arbeiten, und der Gedanke spezifischer, sich ergänzender Funktionen für verschiedene Repräsentationsformen ist grundlegend für

Abb. 2 Hauptfunktionen von multiplen Repräsentationen. (Ainsworth 2006a)



die o. g. *dual coding*-Modelle (Paivio 1986; Schnotz 2002, 2005; s. a. „Multimedia-Prinzip“, Mayer 2005).

Wie oben erwähnt, ist ein wesentlicher kognitiver Gesichtspunkt, der in dem integrierten Modell des Text- und Bild-Verstehens sowie bei verwandten genannten theoretischen Ansätzen (Schnotz 2002, 2005; Mayer 2005) und der Naturwissenschaftsdidaktik (Wu und Shah 2004; Cook 2006; Tsui und Treagust 2013, Kap. 1) vielfach thematisiert wird, *cognitive load*. Im Zusammenhang mit multiplen Repräsentationen ist dabei basierend auf der o. g. Definition die Be- und mögliche Überlastung des Arbeitsgedächtnisses durch die Anzahl und Verarbeitung von mehreren simultan zu bearbeitenden Repräsentationen und von deren Verbindungen gemeint (Tsui und Treagust 2013). Multiple Repräsentationen sind für Denken und Lernen in den Naturwissenschaften essentiell (s. oben), bedingen aber für den Lernenden zugleich die Gefahr einer kognitiven Überlastung.

Unter anderem mit Blick auf die möglichen positiven und negativen Funktionen von multiplen Repräsentationen und auf der Basis dieses kognitionspsychologischen Hintergrundes entwickelte Ainsworth (1999, 2006a, 2008) einen umfassenden Rahmen zu den lehr-lern-theoretischen Dimensionen von Design, Funktionen und Aufgaben von multiplen Repräsentationen (DeFT: *Design, Functions, Tasks*). Das Design von *externen* multiplen Repräsentationen wird durch Parameter wie Anzahl, Art und Reihenfolge der verwendeten Repräsentationsformen sowie enthaltene Informationen bestimmt (vgl. auch Merkmale von Repräsentationen weiter oben im Text). Ein weiteres Design-Merkmal (für Lernaufgaben) ist die Art, wie das Aufzeigen der Beziehungen zwischen Repräsentationsformen als Hilfe zur Übersetzung geschieht. Da es in dem vorliegenden Beitrag um Testaufgaben geht, in denen multiple Repräsentationen erst während der Lösung durch das Hinzukommen interner Repräsentationen entstehen, werden die Design-Merkma-

le von multiplen externen Repräsentationen hier nur kurz aufgeführt.

Der DeFT-Theorierahmen enthält eine Taxonomie der auf Lehren und Lernen bezogenen Funktionen von multiplen Repräsentationen. Als hauptsächliche Funktionen sind „*complementary roles*“, „*constrain interpretation*“ und „*construct deeper understanding*“ (siehe Abb. 2) aufgeführt. Sie sind für diese Arbeit besonders wesentlich und werden daher hier ausführlicher beschrieben (Ainsworth, 1999, 2006a, 2008).

Die komplementäre Funktion (*complementary roles*) besteht darin, dass multiple Repräsentationsformen sich in den Informationen unterscheiden können, die sie enthalten. So werden Informationen über verschiedene Repräsentationsformen verteilt, wenn die Darstellung in einer zu kompliziert wäre (z. B. aufgrund sehr verschiedener Größenordnungen von Information). Durch multiple Repräsentationsformen können komplementäre Prozesse unterstützt werden, da jede Repräsentationsform spezielle Vorteile bei der Lösung von Aufgaben hat. Zum Beispiel wird die Lösung erleichtert, wenn eine Repräsentationsform in einer Aufgabe angeboten wird (oder ausgewählt werden muss), bei der die verwendete Informationsstruktur zur durch die Aufgabe geforderten Informationsstruktur passt. Die Lösungsstrategien für die Aufgabe werden ebenso von den verwendeten Repräsentationsformen beeinflusst, Verwendung mehrerer Repräsentationsformen führt zu einem Wechsel von Lösungsstrategien (welche mehr oder weniger effektiv sein können). Weiterhin haben individuelle Unterschiede der Lerner wie Expertise und zum Teil auch individuelle Vorlieben, z. B. zu visuellen Repräsentationsformen, einen Einfluss auf das Arbeiten mit multiplen Repräsentationsformen.

Die einschränkende Funktion (*constrain interpretation*) besteht darin, dass die in einer Repräsentationsform möglicherweise bestehende Uneindeutigkeit durch die inhärenten Eigenschaften einer anderen eingeschränkt wird (z. B.

geben Abbildungen i. A. die räumliche Lage besser wieder als ein Text). Die einschränkende Funktion kann auch darin bestehen, dass die Ungewissheit über die Bedeutung einer unvertrauten Repräsentationsform durch eine andere, besser vertraute, vermindert wird (z. B. eine schematische Darstellung eines Schaltkreises oder Strahlendiagramms wird durch eine gesprochene oder geschriebene Erläuterung erklärt).

Die Verständnis-konstruierende Funktion (*construct deeper understanding*) schließlich besteht darin, dass Informationen aus verschiedenen Repräsentationsformen integriert werden und durch Abstrahierung, Ausweitung, und In-Beziehung-Setzen zu einem vertieften Verständnis führen (für die Strahlenoptik einschlägige Beispiele werden in dem Abschnitt „Material und Methoden“ angeführt). Ein wenig tiefes Verständnis ist demzufolge durch einen Mangel an solchen Relationen gekennzeichnet.

Der Teil „Tasks“ des DeFT-Theorierahmens beschreibt die kognitiven Aufgaben beim Arbeiten mit multiplen Repräsentationen: Lerner sollen verstehen, wie Repräsentationen aufeinander bezogen, ineinander übersetzt und integriert werden können. Dies fällt vielen Lernenden jedoch schwer (siehe z. B. Ainsworth et al. 2002; Schoenfeld et al. 1993). Die Schwierigkeiten dieser kognitiven Aufgaben hängen von Eigenschaften der Repräsentation und ihrer kognitiven Verarbeitung wie Verarbeitungskanal, Möglichkeit von *dual coding*, Art der Repräsentationsform, Abstraktionsgrad, Spezifität sowie Dimensionalität ab, und ob Repräsentationen separat oder integriert angeordnet sind (im theoretischen Hintergrund z. T. bereits aufgeführt). Individuelle Eigenschaften wie Alter, persönliche und kognitive Faktoren sowie Bekanntheit von Domäne und Repräsentation sind ebenso schwierigkeitsbestimmend (Ainsworth 2006a).

Im Folgenden wird der dargestellte Theorierahmen auf das Inhaltsfeld Strahlenoptik übertragen und an Beispielen erläutert. In der hier vorgestellten Studie wird der ursprünglich für Lernaufgaben formulierte DeFT-Theorierahmen als eine Basis für die Erhebung von Fähigkeiten, Repräsentationen aufeinander beziehen zu können, verwendet. Dafür sind die folgenden Kernpunkte dieses theoretischen Rahmens von zentraler Bedeutung (sie werden nicht in der Reihenfolge des DeFT-Akronyms behandelt; insbesondere kann *Design* hier erst sinnvoll erläutert werden, wenn klar ist, auf welche *Functions* und in den Testitems geforderten *cognitive Tasks* es sich bezieht):

Funktionen (*functions*): Beim naturwissenschaftlichen Arbeiten im Allgemeinen und der Strahlenoptik im Besonderen lassen sich die drei o. g. Grundfunktionen gut identifizieren:

a. Der Informationsgehalt von multiplen Repräsentationen ist wesensmäßig komplementär (*complementary roles*,

z. B. zwischen dem Text, dem Strahlendiagramm und ggf. der Formel in einer Aufgabenstellung).

b. Es gibt typische Fälle, in denen dabei eine Spezifikation der einen durch eine andere Repräsentationsform erfolgt, z. B. wenn man die geometrischen Gegebenheiten im Strahlengang an die einer Realsituation oder deren Fotografie anpassen muss (*constrain interpretation*). Man kann sogar sagen, dass diese besondere Form von *constrain* besonders wichtig im Sinne einer adäquaten, kohärenten Verbindung von Repräsentation und Referent und damit für die Rolle von multiplen Repräsentationen für das Experiment ist.

c. Fast durchwegs wird erst durch das Zusammenwirken mehrerer Repräsentationsformen, und insbesondere deren Konsistenz ein tieferes und vollständigeres Verständnis aufgebaut (*construct deeper understanding*). Hier ist die zentrale Rolle hervorzuheben, die Ainsworth (2006a, 2008) ebenso wie sich hieran anschließende Arbeiten anderer Autoren (z. B. Seufert 2003; Bodemer et al. 2004; Berthold und Renkl 2009), der *Integration* von Informationen aus verschiedenen multiplen Repräsentationen einräumen, und zwar sowohl für die Konstruktion und das Verständnis von Schlüsselkonzepten eines Fachgebiets als auch für das Verständnis der Beziehung dieser Konzepte untereinander.

Aufgaben (*Tasks*): Betreffs dieser Dimension von DeFT sind für diese Arbeit folgende zwei Typen kognitiver Aufgaben bzw. deren Anforderungen besonders wichtig (Ainsworth 2006a, 2008): Erstens die Anforderung zu verstehen, wie Repräsentationsformen aufeinander zu beziehen sind („*understand how to relate representations*“) und zweitens die Anforderung zu verstehen, wie korrekte Beziehungen zwischen Repräsentationen und der inhaltlichen Domäne herzustellen sind („*understand the relation between the representation and the domain*“). Durch Letzteres bekommen Repräsentationen erst Sinn. Dies stellt einen essentiellen, aber auch schwierigen Schritt beim Umgang mit Repräsentationen dar (McDermott 2001; Kozma und Russell 2005). Die Fähigkeiten, diese beiden Arten von Beziehungen herzustellen, gehören zum Kern von repräsentationaler Kohärenzfähigkeit im Sinne dieser Arbeit (s. unten).

Design: Mehrere Design-Merkmale (als dritte Dimension des DeFT-Bezugsrahmens) sind für die Operationalisierung von repräsentationaler Kohärenzfähigkeit wesentlich und werden mit Blick auf die Konstruktion der Items unten im Detail diskutiert.

Repräsentationale Kohärenzfähigkeit

Die Fähigkeit, multiple Repräsentationen als domänenspezifische Werkzeuge des Denkens und Lernens nutzen zu können, beruht zunächst auf der Fähigkeit, eine gegebene

Repräsentationsform sinngebend mit ihren Referenten in Beziehung setzen zu können (s. oben): Dies stellt grundlegend den Sinnzusammenhang für die Verwendung einzelner oder mehrfacher Repräsentationen dar (Ainsworth 2006a, 2008). Insbesondere gilt dies natürlich für Situationen in der Physik, in der Referenten von multiplen Repräsentationen oft Experimente oder Prozesse in der Natur sind. Ebenso ist die Fähigkeit, korrekte, konsistente, sinnvolle und damit kohärente Beziehungen zwischen verschiedenen externen Repräsentationsformen herzustellen, ganz wesentlich: Der o. g. Begriff „RC“ wird auf die Verbindung zwischen allen unterschiedlichen – auch externen – Repräsentationsformen erweitert². Schließlich spielt auch die Fähigkeit, kohärente Beziehungen innerhalb gleicher Repräsentationsformen herstellen zu können, eine große Rolle: Mit dem o. g. Begriff „IC“ („I“ wie „intra“) meinen wir in Bezug zu einer externen Repräsentationsform primär die Verbindungen innerhalb dieser (nicht die Verbindungen zwischen mehreren gleichen Repräsentationsformen)³. Mayer (2002) betont, dass Verbindungen und Kohärenz innerhalb einer Repräsentationsform eine wesentliche Voraussetzung von sinnhaftem Lernen (Ausubel 1963) sind. Ein Beispiel für ICs im gegenwärtigen Kontext ist die Herstellung von Bezügen innerhalb eines Strahlendiagramms. Wenn jedoch zwei oder mehr Varianten von Strahlendiagrammen für verschiedene physikalische Situationen miteinander verglichen werden sollen, müssen auch hier Verbindungen zwischen getrennten Repräsentationen hergestellt werden, d. h. es muss *information integration* geleistet werden. Wegen der Ähnlichkeit der Eigenschaften und Funktionen von ICs zwischen mehreren gleichen Repräsentationsformen und RCs zwischen verschiedenen Repräsentationsformen (Komplementarität, *information integration*, *coherence*) wird an mehreren Stellen die Verwendung des Oberbegriffs „Repräsentationsform-Verbindungen“ (RFC, *representation form connection*)⁴ sinnvoll sein. Die wesentliche Bedeutung dieser Verbindungen für naturwissenschaftliches Denken und Lernen steht natürlich in Einklang mit der zentralen Rolle, die auch in dem DeFT-Modell der Integration von

Informationen aus verschiedenen Repräsentationsformen zugewiesen wird (Ainsworth 2006a, 2008; s. oben).

Vor dem Hintergrund dieser umfangreichen empirischen und theoretischen Arbeiten ergibt sich zunächst ein Verständnis von (multiplen) Repräsentationen als domänenspezifisches „Denkwerkzeug“ („a powerful tool“, Ainsworth 2008; „tools for thinking“, Uttal und O’Doherty 2008; Singer et al. 2012), das sowohl für den wissenschaftlichen Fortschritt wie für das individuelle Lernen in dem betreffenden Gebiet unabdingbar ist (Tytler et al. 2013). Die Fähigkeit, Repräsentationsformen als Denkwerkzeugadäquat einzeln oder miteinander verbunden nutzen zu können, wird Repräsentationskompetenz genannt (Guthrie 2002; Kozma und Russell 2005). In Dänemark ist sie bereits als Anforderung im Lehrplan aufgeführt (Dolin 2007). Repräsentationskompetenz bedeutet, dass die Regeln, die den jeweiligen Repräsentationsformen unterliegen, adäquat und konsistent umgesetzt werden und Informationen fehlerfrei innerhalb einer Repräsentationsform und über Repräsentationsformen hinweg verarbeitet und übertragen werden können. Auf diese Weise werden Informationen einzelner Repräsentationsformen verbunden und Verarbeitungsmöglichkeiten mehrerer Repräsentationsformen genutzt. Auch wenn einzelne Repräsentationsformen verändert werden sollen, sind interne Vergleichsprozesse zwischen der ursprünglichen und der veränderten Repräsentationsform nötig. All diese Denkprozesse basieren wesentlich auf der Voraussetzung, die o. g. *Verbindungen zum Referenten*, *referential* und *intra-representational connections* bzw. „*representational coherence*“ oder „*integration*“ (Seufert 2003; Mayer 1997) herstellen zu können, d. h. Informationen aus einer oder mehreren Repräsentationsformen inhaltlich korrekt und zielführend auf einander beziehen, vergleichen, ändern oder ineinander übersetzen zu können. Dies sind innerhalb von Repräsentationskompetenz also speziell diejenigen Fähigkeiten, die zur Herstellung von Verbindungen zwischen Informationselementen sowohl innerhalb einer als auch über Repräsentationsformen hinweg nötig sind. Diese Fähigkeiten wurden in der pädagogisch-psychologischen Literatur als „*representational skills*“ bezeichnet („ability to translate and switch between representations“, Nistal et al. 2010), eng verwandte Überlegungen und Begrifflichkeiten finden sich in der Physikdidaktik („*representational consistency*“, Nieminen et al. 2010), Chemiedidaktik (Kozma et al. 2000) und Mathematikdidaktik (Even 1998). Auf der Basis dieser Arbeiten und im Sinne der Grundfunktionen des DeFT-Modells (s. oben) stellt in diesem Beitrag repräsentationale Kohärenzfähigkeit einen Teil der allgemeineren Repräsentationskompetenz dar.

Repräsentationale Kohärenzfähigkeit wird als die Fähigkeit verstanden, korrekte, sinnvolle Verbindungen von Informationen innerhalb einzelner und zwischen

² Ursprünglich sind damit Verbindungen von verbalen/deskriptiven und visuellen/depiktiven mentalen Repräsentationsformen gemeint, s. oben und Paivio (1986); Mayer und Anderson (1991); siehe desweiteren „*representational coherence*“, Seufert (2003); „*information integration*“, Glenberg und Langston (1992); Mayer (1997); „*conceptual organisation*“ Schnotz und Bannert (2003); Schnotz (2005).

³ Kognitiv entspricht dies entweder einer deskriptiven Verarbeitung durch „Analyse von Symbolstrukturen“ oder einer depiktiven Verarbeitung durch „analoge Strukturabbildung“ s. oben und Schnotz und Bannert (2003) und Schnotz (2005); bezüglich Kohärenz siehe Mayer (1989); Seufert (2003); Seufert und Brünken (2006).

⁴ Wir verwenden auch bei der Abkürzung „RFC“ das englische Wort „*connections*“, um mit den in der Literatur gängigen Begriffen für „*referential*“ und „*intra-representational connections*“ konsistent zu sein.

mehreren Repräsentationsformen, sowie mit ihren Referenten herstellen zu können.

Damit folgt diese Arbeit einer Sicht von Repräsentationskompetenz und repräsentationaler Kohärenzfähigkeit nicht als *isolierte Fähigkeiten*, die von domänen-spezifischer Expertise unterschieden werden können, sondern sie ist eine *wesentliche Komponente* davon (siehe z. B. Anzai 1991; van Heuvelen 1991). Domänenspezifisch ist die Expertise, weil Repräsentationsformen die Domäne mit ihren Anforderungen repräsentieren und dadurch selbst domänenspezifisch sind (vgl. Konstruktion von Strahlendiagrammen oder die Verwendung der Abbildungsgleichung als bestimmte Formel). Weitere Argumente folgen nach den Beispielen. Man sieht dies schon an konventionellen Aufgabenstellungen wie z. B. der Aufgabe, einen Strahlengang mit gegebenen Werten von G , g und f zu zeichnen (z. B. Item 1a im Test, s. Onlinematerial), aber auch an repräsentational fokussierten Aufgabenstellungen wie z. B. die in Abb. 1 gezeigten verschiedenen Bildelemente (z. B. Gegenstands-, Bild- und Brennweite etc.), die in den beiden Repräsentationsformen miteinander zu vergleichen sind. Im ersten Fall handelt es sich um das Übersetzen von einer Repräsentationsform in eine andere (Text nach Strahlendiagramm), im zweiten Falle um das Vergleichen gegebener Informationselemente (Prüfung auf ein Grundverständnis der Begriffe und der Fähigkeit sie aus verschiedenen Repräsentationsformen herauslesen und sie über Repräsentationsformen hinweg vergleichen zu können). Vor allem Aufgaben des ersten Typs sind in der Strahlenoptik und in vielen anderen physikalischen Teilgebieten absolut gängig (z. B. Zeichnen eines Geschwindigkeitsvektors mit gegebenen Eigenschaften, d. h. Repräsentationsform „Text“ – Repräsentationsform „schematische Darstellung“ wie eben, oder Berechnung der Widerstandes in einem Schaltkreis mit gegebenen Eigenschaften, Repräsentationsform „Text/schematische Darstellung“ – Repräsentationsform „Formel“). Solche Aufgabentypen zur sachgerechten und kohärenten Verwendung von multiplen Repräsentationen sind in der Disziplin Physik üblich und gehören quasi implizit zu den Grundanforderungen. Darüber hinaus sind Repräsentationskompetenz und repräsentationale Kohärenzfähigkeit aber auch notwendige Voraussetzungen für den weitergehenden Gebrauch von multiplen Repräsentationen als domänenspezifische Denkwerkzeuge, mit hoher Bedeutung für andere Fähigkeiten, siehe die eingangs genannte Literatur (Abschnitt „multiple Repräsentationen“) für Überblicksdarstellungen, und insbesondere für konzeptuelles Verständnis (Van Heuvelen und Zou 2001; Hubber et al. 2010), schlussfolgerndes Denken („reasoning“, van Heuvelen 1991; Plötzner und Spada 1998; Verschaffel et al. 2010), Problemlösen (Kohl und Finkelstein 2006; Verschaffel et al. 2010, Kap. 4), und Kreativität (Schnotz et al. 2010). Daher ist zu erwarten, dass mit der

Verbesserung von Repräsentationskompetenz und repräsentationaler Kohärenzfähigkeit auch insgesamt eine Verbesserung der Leistungsfähigkeit in einer gegebenen Domäne einhergeht und sich repräsentationale Kohärenzfähigkeit und die Fähigkeit, mit Repräsentationsformen lernen zu können beim Arbeiten mit Repräsentationen sukzessive entwickeln und somit wechselseitig in einer Kausalbeziehung zueinander stehen. Vor diesem Hintergrund muss betont werden, dass multiple Repräsentationen nicht nur ein didaktischer Ansatz sind, den man (im besten Fall evidenzbasiert) nutzen könnte, sondern sie sind durchgängig ganz wesentlicher Bestandteil der naturwissenschaftlichen Denk- und Arbeitsweise, den man – ähnlich wie Experimente – nicht einfach aus didaktischen Gründen weglassen könnte, ohne etwas Wesentliches zu verfehlen. Das heißt pointiert gesagt, multiple Repräsentationen sind, ähnlich wie Experimente, nicht nur ein *Verfahren*, sondern grundlegender *Gegenstand* naturwissenschaftlichen Unterrichts (van Heuvelen 1991).

Wie oben erwähnt, ist *cognitive load* ein wesentlicher Gesichtspunkt für das Lernen mit multiplen Repräsentationen. Dabei ist nachrangig, welche Repräsentationsform und welche Art von Verbindungen (RC, IC) involviert sind. Entscheidend ist die Anzahl der Informationselemente, d. h. der involvierten Repräsentationsformen und deren Verbindungen (Van Merriënboer und Sweller 2005, Schnotz und Kürschner 2007). Auf dieser theoretischen Basis, und in Anlehnung an das DeFT Modell, sowie an Neumann et al. (2007) und Kauertz (2008) werden wir bei der Operationalisierung der Items (Abschn. 2) daher auch die Anzahl der Elemente und ihrer Verbindungen als Design-Merkmal aufnehmen.

Wir gehen an dieser Stelle noch auf die Frage ein⁵, ob Repräsentationskompetenz und repräsentationale Kohärenzfähigkeit in kognitionspsychologischen Begriffen ein Teil des prozeduralen Wissens einer Domäne ist, des „*Knowing How*“ in Form von domänenspezifischen Fähigkeiten, Techniken, Algorithmen und Methoden sowie deren Anwendungsregeln (Krathwohl 2002; Koziol und Budding 2012). Die vorliegende Literatur geht nur sehr vereinzelt explizit auf die Beziehung von multiplen Repräsentationen und prozeduralem Wissen ein (uns sind nur Beispiele aus der Mathematikdidaktik bekannt: z. B. Haapsalo und Kadjevich 2000), jedoch ist der Gedanke von multiplen Repräsentationen als Denkwerkzeuge sehr ausgeprägt (s. oben). Nimmt man diesen ernst, so ergibt sich die Antwort fast zwingend: Der Gebrauch von Werkzeugen setzt *Knowing*

⁵ Wir danken den Referees einer früheren Version der Arbeit für diesen Hinweis. Es sei darauf hingewiesen, dass es neben der o. g. Auffassung von prozeduralem Wissen im breiteren Sinne (Krathwohl 2002; Koziol und Budding 2012) eine – historisch frühere – Auffassung in einem engeren Sinne von weitgehend unbewussten „production rules“ gibt, unter die Repräsentationskompetenz nur zum Teil fallen würde.

How voraus. Das *Knowing how* von multiplen Repräsentationen besteht darin, domänenspezifische multiple Repräsentationen und deren Verbindungen zu verstehen, sowie sie einzeln oder miteinander verbunden als Denkwerkzeuge nutzen zu können. In diesem Sinne sind Repräsentationskompetenz und repräsentationale Kohärenzfähigkeit in der Tat Bestandteile des prozeduralen Wissens einer Domäne, z. B. wie Information aus einem Diagramm erschlossen werden kann (Plötzner und Lowe 2012). Laut Ainsworth (2006) beeinflusst die Domänenkenntnis auch die Schwierigkeit der Arbeit mit multiplen Repräsentationen. Um mit den in Abb. 1 dargestellten Repräsentationsformen arbeiten zu können, ist repräsentationale Kohärenzfähigkeit in Form der Bildung von ICs nötig, um das Strahlendiagramm zu verstehen und die Bildung von RCs, um Informationselemente wie z. B. die Bildweite zwischen beiden Bildern vergleichen zu können. Grundvoraussetzung ist immer, die abgebildeten Gegenstände mit ihrem Referenten in Beziehung setzen zu können.

Notwendigkeit eines Instruments für repräsentationale Kohärenzfähigkeit

Man steht nach den o. g. angeführten Überlegungen vor dem didaktischen Dilemma, dass multiple Repräsentationen einen wesentlichen Bestandteil physikalischen Denkens und Lernens darstellen, zugleich aber (z. T. sehr) hohe Anforderungen mit sich bringen. Instrumente zur Bestimmung von Fähigkeiten sind in diesem Spannungsfeld offensichtlich von Bedeutung (s. weitere Begründung unten). Es gibt viele Hinweise, dass Aneignung und Gebrauch des kognitiven Werkzeugs multiple Repräsentation alles andere als leicht für Lernende sind. Empirische Belege hierfür gibt es von der Primarstufe (Ainsworth et al. 2002) bis zum universitären Niveau (Elektrodynamik, Saniter 2003; Mechanik, Nieminen et al. 2010). Als ein Hauptproblem wird auf Lernerebene die Schwierigkeit genannt, RCs herzustellen (Schwonke et al. 2009), d. h. es liegt eine unzureichende repräsentationale Kohärenzfähigkeit im Sinne dieses Beitrages vor. Eine weitere, für Verständnis und Gebrauch von Repräsentationen zentrale Schwierigkeit für Lernende liegt darin, die Beziehung von Repräsentationen zu wirklichen Objekten und Prozessen herzustellen (McDermott 2001). Damit kann der eigentlich erst sinnstiftende Bezug zum Referenten als eine der zentralen Funktionen des DeFT-Modells nur unzulänglich oder gar nicht ausgeführt werden (s. oben). Auf Unterrichtsebene wird eine Hauptschwierigkeit durch eine Analyse von 26 Unterrichtsstunden über das Thema Strahlenoptik in der Sekundarstufe I belegt (Lee 2009), nach der im Unterricht überwiegend eine nur implizite, rezeptive und kurze Verwendung (<3min) von Repräsentationen umgesetzt ist. Weniger als 10 % der Repräsentationen werden durch die Schülerinnen und Schüler selbst er-

zeugt. Ein ähnliches Bild zeigte sich bei einer Analyse von über 800 Aufgaben aus gymnasialen Physikschulbüchern (Scheid 2013): ganz überwiegend enthalten die Aufgabenstellungen nur eine Repräsentationsform (mittlere Anzahl „ \bar{N} “, Standardabweichung davon in Klammern: $\bar{N}_{RF} = 1,24 (0,44)$), und sie erfordern für die Lösung i. d. R. erst recht wenige RCs ($\bar{N}_{RC} = 0,66 (0,54)$).

Empfehlungen für Lehr-Lern-Ansätze, die auf der o. g. Forschungsbasis beruhen, sehen in dem Fehlen einer gezielten und systematischen kognitiven Aktivierung zu multiplen Repräsentationen (Cox und Brna 1995; Verschaffel et al. 2010, Kap. 6) eine wesentliche Quelle von Fehlern und Lernschwierigkeiten, und fordern als Gegenmaßnahme einen „*representational focus*“ (Tytler et al. 2013; Hubber et al. 2010), im Sinn einer aktiven und expliziten Auseinandersetzung mit multiplen Repräsentationen (van Someren et al. 1998, Kap. 16; Tsui und Treagust 2013, Kap. 7). Ausdrücklich werden dabei auch auf multiple Repräsentationen bezogene Lernaktivitäten gefordert, die über vereinfachte Schulbeispiele hinausgehen und komplexere Formen von RCs enthalten. Insbesondere sind solche mit höherer Anzahl von verwendeten Repräsentationsformen und von deren Verbindungen gemeint (van Someren et al. 1998, Kap. 15; van Heuvelen und Zou 2001; Tsui und Treagust 2013, Kap. 1). Dabei soll für diese Aktivitäten auch deren Charakter als authentische Tätigkeit im Sinne der disziplinären Praxis der Bezugsdomäne sichergestellt werden (Tytler et al. 2013, Kap. 5). Ein besonders wichtiger Fall für eine solche höhere Anzahl ($N_{RF} > 2$, $N_{RC} > 1$) ist die Kombination von zwei verschiedenen Repräsentationsformen und dem Referenten, z. B. einer Text-Bild Kombination und einem Experiment oder empirischen Phänomen (ein für die Naturwissenschaften quasi „natürlicher“ Fall). Beispiele für explizite Lernaktivitäten und -ansätze dieser Art finden sich in der Literatur etwa bei Tytler et al. (2013), Kap. 1 (biologisches Thema), Kap. 7 und 8 (astronomisches Thema), sowie bei van Heuvelen und Zou (2001, physikalisches Thema). Zusammengefasst lässt sich sagen, dass repräsentationale Kohärenzfähigkeit in dem Sinne korrekte, sinnvolle Verbindungen von multiplen Repräsentationen untereinander und mit ihrem Referenten herstellen zu können, zum einen eine ganz wesentliche Komponente von domänen-spezifischer Expertise ist. Zum anderen aber stellen diese Verbindungen eine erhebliche Lernschwierigkeit dar: Die Fähigkeit dazu ist häufig auch nach Unterricht und sogar noch auf universitärem Niveau gering ausgeprägt. Des Weiteren fehlt es vielfach an Gelegenheiten für auf Repräsentationen bezogene kognitive Aktivierung und selbstständige repräsentationale Arbeit der Lernenden, mit der die genannten Lernschwierigkeiten überwunden und das kognitive Potential ausgeschöpft werden könnten.

Um den Stand an repräsentationaler Kohärenzfähigkeit feststellen zu können und einschlägige fördernde Maßnah-

men bewerten zu können, gibt es daher einen gut begründeten Bedarf an validen und reliablen Messinstrumenten (van Someren et al. 1998, Kap. 2; Tytler et al. 2013, Kap. 2 und 9), der auch im Rahmen des DeFT-Modells unterstrichen wird (Ainsworth 2006a, 2008). Seitens der Literatur stehen jedoch sehr wenige Instrumente für repräsentationale Kohärenzfähigkeit in der Physik zur Verfügung (siehe Nieminen et al. 2010 für eine der wenigen Ausnahmen im Bereich Newtonsche Mechanik), und keines im Bereich Strahlenoptik, als einem an multiplen Repräsentationen besonders reichen und in vielen Lehrplänen in der Sekundarstufe I früh angesiedelten Thema. Sollte sich hier eine niedrige repräsentationale Kohärenzfähigkeit bestätigen, wäre es als Konsequenz nötig, durch geeignete Fördermaßnahmen deren Entwicklung zu unterstützen sowie den Erfolg der Fördermaßnahmen zu evaluieren. Aus diesen Gründen war das Ziel der hier vorgestellten Arbeit die Entwicklung und psychometrische Charakterisierung eines Testinstrumentes zur Erhebung von repräsentationaler Kohärenzfähigkeit (RKF-Test) von Lernenden im Bereich Strahlenoptik.

Material und Methoden

Design-Merkmale der Items und Operationalisierung von repräsentationaler Kohärenzfähigkeit

Eine Kompetenz ist nicht direkt, sondern indirekt an der Performanz, Aufgaben lösen zu können, diagnostizierbar (Köller 2008). In dieser Arbeit beruht die Operationalisierung von repräsentationaler Kohärenzfähigkeit auf der oben eingeführten Konzeptualisierung als Fähigkeit, korrekte, kohärente und sinnvolle Verbindungen von Informationselementen innerhalb einzelner und zwischen verschiedenen Repräsentationsformen herstellen zu können (ICs und RCs). Wesentliche Eigenschaft von repräsentationalen Kohärenzfähigkeit-Items ist also, dass die korrekte Lösung ausschließlich mittels solcher kohärenter Verbindungen (RCs, ICs und als Voraussetzung die Verbindung zum Referenten, s. Abschn. 1) möglich ist, die hier wie folgt operationalisiert sind: Aufeinander beziehen, Übersetzen, Vergleichen und Ändern/Anpassen von Informationselementen innerhalb und zwischen Repräsentationsformen.

Verwendet werden dabei Text und Formeln als deskriptive und Fotos und Strahlendiagramme/Zeichnungen als depiktionale Repräsentationsformen (siehe Beispiele unten

Tab. 1 Übersicht über die im RKF-Test verwendeten Repräsentationsformen, der für die Lösung der Items notwendigen Verbindungen auf der Ebene von Repräsentationsformen und auf der Ebene von Einzelelementen

Item	Typ	Verwendete Repräsentationsformen							Zahl und Art der Verbindungen von Repräsentationsformen		Zahl und Art der Element-Verbindungen	
		N_{RF}	N_{RFC}	Text	Tabelle	Formel	Strahlengang	Real. Abbildung/Foto	RFn beziehen, übersetzen	RFn vergleichen, ändern/anpassen	N_{IC}	N_{RC}
1a	Standard	2	1	1 ^a	–	–	1	–	1	0 (per Konstruktion)	10	4
2		2	1	1	1	–	–	–	1 (wdh.)		2	2
3		2	1		1 ^b	1	–	–	1 (wdh.)		5	4
4a		2	1	1	–	1	–	–	1		3	3
7		2	1	1	–	–	1	–	1 (wdh.)		4	8
\bar{N}		2,0	1,0	–	–	–	–	–	1,0		4,8	4,2
1b		DU	4	3	2	–	–	2	–	2	1 (wdh.)	4
1c	4		3	2	–	–	2	–	2	1	3	5
4b	3		2	2	–	–	1	–	1	1	0	4
5a	4		2	2	–	–	–	2	1	1	0	2
5b	4		2	2	–	–	–	2	1	1	0	2
5ca	4		3	2	–	–	–	2	2	1	0	2
5cb	4		3	2	–	–	–	2	2	1	0	2
6	5		5	2	–	1	2	–	2	3	4	2
8	4	4	2	–	–	–	2	3	1	4	4	
\bar{N}	4,0	3,0	–	–	–	–	–	1,8	1,2	1,7	3,0	

N_{RF} Anzahl von verschiedenen Repräsentationsformen, N_{RFC} Anzahl von Verbindungen zwischen verschiedenen Repräsentationsformen, *DU deeper understanding*, N Anzahl, \bar{N} mittlere Anzahl, *wdh.* wiederholt: eine bereits erzeugte Repräsentationsform-Verbindung wird nochmals mit einer anderen Größe verwendet und insgesamt nur einmal gezählt

^aEine in den Text integrierte Größenangabe (z. B. $f = 3 \text{ cm}$) wird nicht als eigenständige Repräsentationsform gezählt

^bTabelle und kurzer Text (Tabellenüberschrift) bilden eine Einheit

und Tab. 1 für einen Überblick). Mit Blick auf curriculare Validität in Rheinland-Pfalz und anderen Bundesländern mit ähnlichem Lehrinhalt (siehe Diskussion) sollen die Items des Tests die in der schulischen Zielstufe typischen Themengebiete im Bereich Strahlenoptik zur Bildentstehung abdecken (MBWW 1998). Formeln als Repräsentationsform sind dabei wenig vertreten (4 von 14 Items), weil laut Lehrplan auf einen starken Anteil der Mathematisierung verzichtet werden soll (MBWW 1998). So kann die Studie mit schulrelevanten Themen durchgeführt werden und der effektive Zeitverlust für den Unterricht ist gering. Um Testzeit gewinnbringend für den Unterricht nutzen zu können, sollen Lehrer nach Lerngelegenheit der Schülerinnen und Schüler die Möglichkeit haben, die Testergebnisse zur Bewertung der Lernenden verwenden zu können, sofern die curriculare Validität inklusive Eignung der Items für einen Leistungstest positiv evaluiert wird.

Wie im Theorieteil erläutert, müssen die verwendeten Repräsentationsformen dabei nicht alle extern vorliegen oder generiert werden, auch eine interne Verwendung von Repräsentationsformen im Arbeitsgedächtnis z. B. als mentales Modell oder Proposition (Schnotz 2005) ist möglich. Als Lösung des jeweiligen Items muss natürlich immer (wie bei jedem Test) eine externe Repräsentationsform erzeugt werden, anhand welcher der Erfolg der durch die Lernenden vollzogenen Verknüpfungsprozesse von Informationselementen, d. h. die Kohärenzfähigkeit im Sinne dieser Arbeit, erkennbar wird.

In Begriffen des o. g. Theorierahmens (DeFT, s. Abschn. 1) gilt für die Design-Parameter der repräsentationalen Kohärenzfähigkeit-Items:

- a. Es sind mindestens zwei verschiedene Repräsentationsformen beteiligt ($N_{RF} \geq 2$). Außerdem ist mindestens eine Verbindung zwischen den verwendeten Repräsentationsformen erforderlich ($N_{RFC} \geq 1$). Man beachte, dass solche Verbindungen im Sinne von Abschn. 1 („Repräsentationale Kohärenzfähigkeit“) zwischen verschiedenen Repräsentationsformen bestehen können (RC-Typ), oder zwischen zwei verschiedenen Versionen der gleichen Repräsentationsform (IC-Typ).
- b. Die Verteilung der Informationen über die beteiligten Repräsentationsformen ist komplementärer Art, mit teilweise gemeinsamen Inhalten (partielle Redundanz); das ist bei multiplen Repräsentationen in den Naturwissenschaft-

- ten der Regelfall (z. B. tritt die Bildweite in einem Strahlengang und in einer Formel auf).
- c. Das Repräsentationssystem ist das eines Papier- und Bleistift-Dokumentes, d. h. statisch und monomodal (Sehen).
- d. Die Sequenzierung hat die Grundform:
 1. Die Aufgabenstellung besteht primär aus Text und erläutert ggf. eine weitere gegebene depiktionale Repräsentationsform.
 2. Die Aufgabenlösung enthält verschiedene Kombinationen von Repräsentationsformen und Verbindungen zwischen solchen, die Voraussetzung für eine erfolgreiche Bearbeitung der Aufgabenstellung sind. Eine Besonderheit besteht dabei darin, dass die Sequenzierung im Sinne der DeFT-Taxonomie innerhalb des eigentlichen Lösungsschrittes von dem Lernenden selbst festgelegt wird.
 3. Das Ergebnis enthält eine der Repräsentationsformen aus Schritt 2, z. T. zusätzlich Text in Form eines Lösungssatzes.
- e. Die Übersetzungsunterstützung im Sinne von DeFT (Ainsworth 2006a) hat hier insofern einen besonderen Status, als sie durch den Aufgabencharakter seitens des Lernenden selbst festgelegt wird bzw. Gegenstand des Items ist (Beziehungen zwischen Repräsentationsformen werden nicht gegeben, sondern erfragt).

Ein weiteres Unterscheidungsmerkmal von Aufgaben ist, ob die Art der Verbindung von Repräsentationsformen zum Lösen eines Items einen Bezug bzw. eine Übersetzung darstellt, und damit der konventioneller Standardaufgaben entspricht, oder ob die Verbindung von der Art her ein Vergleichen von Repräsentationsformen darstellt und über ein bloßes Beziehen bzw. Übersetzen hinausgeht. Wie in dem in Abschn. 1 dargestellten theoretischen Hintergrund erläutert, ist repräsentationale Kohärenzfähigkeit eine wichtige Teilkomponente von physikalischem Verständnis überhaupt, oder anders gesagt, es gibt domänenspezifische Fragestellungen, die man gar nicht ohne den impliziten Gebrauch von multiplen Repräsentationen und repräsentationaler Kohärenzfähigkeit beantworten kann. Es gibt also auch unter gängigen Aufgaben schon solche, die implizit auf repräsentationale Kohärenzfähigkeit testen, z. B. muss für jede Formel, oder insbesondere für Strahlengänge in der Optik, eine kohärente Verbindung zu der im Aufgabentext beschriebenen Situation hergestellt werden. In solchen Items werden die Verbindungsarten „Beziehen“ (Bsp. Informationselemente in den beiden Repräsentationsformen von Abb. 1, oder Einsetzen aus Text in Formel in Item 4a, s. Abb. 3) und „Übersetzen“ (Übersetzen von Strecken als Zahlenwert in Strecken eines Strahlendiagramms, s. Item 1a) vorausgesetzt. Zweck dieses Itemtyps ist es, denjenigen Teil von repräsentationaler Kohärenzfä-

Ines möchte gerne einen 5 mm großen Marienkäfer auf 20 mm vergrößert zeichnen. Um ihn abpausen zu können will sie zunächst ein Bild des Käfers auf einem Schirm erzeugen. Der Marienkäfer befindet sich 10 mm vor einer passenden Linse. In welcher Entfernung muss sich der Schirm befinden? Löse die Aufgabe rechnerisch!

Abb. 3 Item, welches als RC das Beziehen zwischen den Repräsentationsformen „Text“ und „Formel“ erfordert sowie als IC die Weiterverarbeitung innerhalb der Rechnung (Item 4a des Tests)

higkeit abzubilden, der in gängigen Unterrichtszielen und Leistungstests implizit vorausgesetzt wird. Um diesen Aufgabentyp innerhalb des RKF-Tests so ähnlich wie möglich zu üblichen Aufgaben zu machen, wurden dabei für die Anzahl der nötigen multiplen Repräsentationen und Verbindungen von Repräsentationsformen deren Minimalwerte genommen ($N_{RF} = 2$; $N_{RFC} = 1$). Wir werden diese Form der Anforderung an repräsentationale Kohärenzfähigkeit und die zugehörigen Items als „Standard“ bezeichnen, weil sie Standard-Aufgaben in einer gegebenen Domäne entsprechen.

Das DeFT-Modell sieht jedoch auch eine darüber hinaus gehende Funktion vor, nämlich „construct deeper understanding“ (Ainsworth 2006a), s. Abschn. 1. Für Testaufgaben ist „construct“ schon erfolgt, da tieferes Verständnis schon konstruiert sein muss, um die Aufgabe lösen zu können. Items, welche diese Anforderung an repräsentationale Kohärenzfähigkeit haben, bezeichnen wir als „*deeper understanding*-Items“ („DU-Items“). Die Anforderungsart wird hier durch Items mit den Verbindungsarten „Vergleichen“ und „Ändern/Anpassen“ realisiert: Beim Vergleichen geht es beispielsweise darum, bei einer Linsenabbildung den Abbildungsfall⁶ zu identifizieren, wenn die Größe und der Abstand des Gegenstands nicht gezeigt sind (Vergleich Foto-Tabelle; Item 5a). Beim Ändern/Anpassen geht es zum Beispiel um die Änderungen der Bildeigenschaften nach Änderung der Gegenstandsweite (Item 1b). Man beachte, dass mit „Vergleich“ hier nicht der einzelner Elemente gemeint ist, sondern der einer (informationstragenden) *Beziehung* von Elementen, und dass zielgerechtes Ändern immer einen Vergleich voraussetzt (nämlich den der ursprünglichen und der geänderten Situation). Eine für die Operationalisierung interessante und nützliche Verbindung zu einer weiteren theoretischen Perspektive besteht zu dem Komplexitätsmodell für Lerninhalte der Essener Gruppe (Kauertz und Fischer 2006; Neumann et al. 2007; Kauertz 2008), welches Komplexität als Anzahl von Informationselementen (Fakten) und von deren Zusammenhängen fasst. Danach entsprechen die Standard-Items den Komplexitätsstufen „Ein Zusammenhang“, evtl. durch bloßes Abarbeiten mehrerer paarweiser Beziehungen auch der Stufe „Mehrere unverbundene Zusammenhänge“ (nämlich durch Beziehen oder Übersetzen). Die DU-Items hingegen fallen unter die höheren Komplexitätsstufen mit *verbundenen* Zusammenhängen, für die o. g. Beispiele ereignet sich dies in folgender Form: Bei der Identifikation des Abbildungsfalles in einem Abbildungsexperiment (als Foto vorliegend; Aufgabe 5) muss die Relation mehrerer Zusammenhänge betrachtet werden: z. B.

- die Bildweite im Abbildungsexperiment muss mittels der Bildweite in der Tabelle der Bildfälle gefunden werden, um den Zusammenhang zur Gegenstandsweite dort ablesen zu können und auf das auf dem Foto gezeigte Abbildungsexperiment übertragen zu können.
- Alternativ kann das gegebene Foto des Abbildungsexperimentes auch nach Erinnerung an durchgeführte Abbildungsexperimente und der Auswahl des passenden Experimentes durch eine realistische Abbildung ergänzt werden. Dasselbe würde gelten, wenn man den Abbildungsfall mit einem Strahlendiagramm rekonstruiert.

Bei der Frage zu den Änderungen der Bildweite und Bildgröße nach Änderung der Gegenstandsweite (Item 1b) müssen im ursprünglichen und geänderten Strahlengang z. B. jeweils der Winkel zwischen Brennpunktstrahl und optischer Achse auf der einen Seite der Linse und der Abstand der Parallelstrahlen auf der anderen betrachtet werden (mehrere Zusammenhänge zwischen Informationselementen). Dazu muss man schon weitgehend über das mentale Modell der Bildkonstruktion verfügen.

Zweck des DU-Itemtyps ist es, höhere Ausprägungen von repräsentationaler Kohärenzfähigkeit zu erfassen und die Wirkung von didaktischen Ansätzen zu testen, bei denen – im Unterschied zum üblichen Unterricht, aber im Einklang mit Vorschlägen aus der Literatur, s. Abschn. 1 – ein Fokus auf kognitive Aktivierung und eine i. a. *höhere Anzahl* von multiplen Repräsentationen und Verbindungen von Repräsentationsformen Lern- und Übungsgegenstand gelegt wird. Dementsprechend beinhalten DU-Items mindestens drei Repräsentationsformen und mindestens zwei Verbindungen zwischen ihnen ($N_{RF} \geq 3$; $N_{RFC} \geq 2$). Es sind also jeweils mindestens eine Repräsentationsform und eine Verbindung von Repräsentationsformen mehr als bei Standard-Items (im Durchschnitt sogar je zwei mehr: $\bar{N}_{RF} = 4,0$; $\bar{N}_{RFC} = 3,0$).

Es sei darauf hingewiesen, dass schon die Standard-Items auf einer mittleren Komplexitätsstufe des Essener Modells (a. a. O.) stehen. Es ist also davon auszugehen, dass sie keineswegs einfach für Lernende sind. Für die DU-Items gilt dies in verstärktem Maße. Man kann also durch die Verwendung der Komplexitätsstufen nach diesem Modell gut noch einmal das eingangs genannte didaktische Dilemma von multiplen Repräsentationen verdeutlichen: Wesentlicher Bestandteil von naturwissenschaftlichem Denken und Lernen einerseits, von hoher intrinsischer Komplexität und Schwierigkeit andererseits.

Übersicht über das Instrument für repräsentationale Kohärenzfähigkeit

Inhaltlich verbundene Items sind in acht „Aufgaben“ (im Sinne gängiger Unterrichtspraxis) gruppiert. Diese enthal-

⁶ Abbildungsfall: Bildgröße im Vergleich zur Gegenstandsgröße, Bildweite im Vergleich zur Gegenstandsweite, reelles oder virtuelles Bild.

Ein gewöhnliches Zimmerfenster wird durch eine Sammellinse auf einer Wand in einem Raum abgebildet.



- a) Um welchen Bildfall handelt es sich? _____
- b) Kreuze an, wie groß die Gegenstandsweite g etwa sein muss!
- g ist hier ungefähr so groß wie der Abstand zwischen Wand und Sammellinse
 - g ist hier kleiner als der Abstand zwischen Wand und Sammellinse
 - g ist sehr viel größer als der Abstand zwischen Wand und Sammellinse
 - g ist doppelt so groß als der Abstand zwischen Wand und Sammellinse
- c) Erkläre, wie du auf die Lösung des Aufgabenteils a) und b) gekommen bist!

Abb. 4 Item-Beispiele für den Aufgabentyp „Vergleich von Repräsentationsformen“ (Items 5a und 5b des Tests; die Aufgabe enthält außerdem einen qualitativen Teil 5c, der Aufschluss über den Lösungsweg zu Item 5a und 5b gibt, s. Text)

ten in der Mehrzahl auch Items („Teilaufgaben“) im offenen Antwortformat, um auch Informationen zu erhalten, auf welchem Weg die Lernenden zur Lösung gelangt sind. Insgesamt enthält das Instrument für repräsentationale Kohärenzfähigkeit fünf Standard-Items ($N_{RF} = 2$; $N_{RFC} = 1$ per Konstruktion) und neun DU-Items ($\bar{N}_{RF} = 4,0$; $\bar{N}_{RFC} = 3,0$).

Tab. 1 fasst die Variation der wesentlichen Parameter über die Items des RKF-Tests zusammen, und die Operationalisierung wird im Folgenden anhand mehrerer Item-Beispiele erläutert (s. Tab. 3 im Onlinematerial 2 für weitere Angaben zu den Items, u. a. zu den o. g. Anzahlen von representation form connections und der Verbindungsarten sowie diesbezüglichen Erläuterungen).

Um Standard-Items zu lösen genügt es, zwei Repräsentationsformen aufeinander zu beziehen. Dies entspricht der auch bei traditionellen Aufgaben implizit erforderlichen repräsentationalen Kohärenzfähigkeit im Sinne einer einfachen Verbindung bzw. Übersetzung von Informationen aus zwei Repräsentationsformen. Abb. 3 zeigt ein Beispiel dieser Art (Item 4a). Es handelt sich um eine traditionelle Textaufgabe zur Anwendung der Abbildungsgleichung. In der Lösung müssen die Repräsentationsformen Text und Formel aufeinander bezogen werden ($N_{RFC} = 1$), dazu müssen dem Text Informationselemente über die Größen entnommen werden und ein fachlich korrekter Bezug zu den Größen in der Formel hergestellt werden. Die Ausprägung des Erfolgs dieses Verbindungsaufbaus gibt Aufschluss auf die

interrepräsentationale Komponente der repräsentationalen Kohärenzfähigkeit der Lernenden, also in diesem Fall die Qualität der von Lernenden erzeugten Kohärenz zwischen der Textrepräsentation und der mathematischen Darstellung. Die (Weiter-)Verarbeitung von Informationselementen innerhalb einer Repräsentationsform gibt Aufschluss auf die intrarepräsentationale Komponente der repräsentationalen Kohärenzfähigkeit der Lernenden.

Ein zweites Beispiel dieser Art ist die zeichnerische Ermittlung der Brennweite einer Linse auf der Basis der im Aufgabentext enthaltenen Angaben zu Bildweite, Gegenstandsweite und Gegenstandsgröße (ebenfalls eine gängige Aufgabenstellung). Dabei müssen die Repräsentationsformen Text und Strahlengang aufeinander bezogen werden ($N_{RFC} = 1$) und dazu Informationselemente (Strecken, Größen) aus der Repräsentationsform „Text“ in die Repräsentationsform „Strahlengang“ übersetzt und dort weiterverwendet werden (s. Item 7, Onlinematerial 1 und 2). Um die Lösung anzugeben („Brennweite“) wird die oben erwähnte Verbindung zwischen den Repräsentationsformen „Text“ und „Strahlengang“ in umgekehrter Richtung wiederholt (keine neue Kohärenzleistung).

Zu den DU-Items gehört der Aufgaben-Typ „Vergleichen von Repräsentationsformen“. Die Aufgabenstellung verlangt neben dem bei Standard-Items nötigen Beziehen von zwei Repräsentationsformen eine Prüfung auf Abweichungen bzw. Übereinstimmungen von Sachverhalten in (mindestens) zwei Repräsentationsformen. Dafür müssen mehrere verbundene Informationselemente aus zwei Repräsentationsformen entnommen und miteinander in Beziehung gesetzt werden; eine dritte bzw. evtl. vierte Repräsentationsform kommt durch den Aufgabentext und die Antwortrepräsentation ins Spiel. Abb. 4 zeigt ein Beispiel dieser Art (Aufgabe 5). Dabei muss die Aufgabenstellung auf ein gegebenes Foto bezogen werden (1. RFC) und das Foto zur Lösung der Teile a) und b) zusätzlich mit einer Abbildungsfalltabelle oder bereits gemachten Experimenten mental verglichen werden (2. RFC, alternativ können auch Strahlendiagramme verwendet werden, die mental erzeugt werden). Man beachte, dass Fotos von Repräsentationsformen zwar einen hohen Grad an Realismus haben, aber dennoch einen hohen Anspruch bei der Erstellung von Verbindungen zu anderen Repräsentationen mit sich bringen, weil sie viele irrelevante Informationen enthalten, und vom Betrachter verlangen, diese Informationen herauszufiltern. In jedem Fall müssen auf Elementebene relevante Informationen wie Bildweite bzw. Bild- und Gegenstandsgröße der beiden Repräsentationsformen von den Lernenden verglichen und angepasst werden, bis sie mit den im Experimentausschnitt gezeigten Parametern übereinstimmen. Dann erfolgt direkt die Angabe der in der Tabelle abgelesenen Lösung, bzw. bei Verwendung einer bildhaften Repräsentationsform anstatt der Tabelle, die Übersetzung der

gefundenen Lösung in den Antworttext (in letzterem Fall gibt es eine zusätzliche representation form connection). Ein weiterer Aufgabenteil (Items 5ca/b) fragt jeweils nach dem Lösungsweg der Lernenden (z. B. mit welcher RF) im offenen Antwortformat. Hierzu muss die Bearbeitung des Aufgabenteils 5a bzw. 5b (jeweils 1. und 2. RFC) in eine Textrepräsentation übersetzt werden (3. RFC).

Als weiteren Typ enthalten die DU-Items Aufgabenstellungen, in denen das Ändern oder Anpassen von Repräsentationsformen gefordert ist. Dabei muss eine Repräsentationsform anhand von Informationen geändert werden, welche aus einer oder mehreren anderen Repräsentationsformen entstammen. Dies setzt zunächst einen Vergleich zwischen Repräsentationsformen im Sinne des vorigen Abschnittes voraus (nämlich zw. Ausgangs- und Ziel-RF), und darüber hinaus das eigenständige Generieren der nötigen Änderung; damit geht eine erhöhte Anzahl von Repräsentationsformen und Verbindungen zwischen ihnen einher. Beispielsweise fragt Item 1b (siehe Onlinematerial 1 und 2) in Bezug auf eine gegebene Strahlenkonstruktion einer optischen Abbildung durch eine Sammellinse nach der Auswirkung einer vergrößerten Gegenstandsweite auf Bildgröße und Bildweite. Zur Lösung wird in einem ersten Schritt das Beziehen des Aufgabentextes auf den Strahlengang (1. RFC) erwartet. In einem zweiten Schritt soll der gegebene Strahlengang zur Lösungsgenerierung nach den Vorgaben aus dem Text verändert werden, wobei eine selbst neu zu generierende (interne oder externe) schematische Zeichnung mit der vorliegenden zu vergleichen ist, um Unterschiede erkennen und Anpassungen vornehmen zu können (2. RFC). Die Erkenntnisse aus dem Strahlengang werden in eine Textrepräsentation als Antwort überführt (3. RFC). Auf Ebene der Informationselemente müssen Informationen zwischen Repräsentationsformen und innerhalb von Repräsentationsformen verbunden werden (RCs bzw. ICs).

Zur Auswertung des RKF-Tests werden für alle Items Punkte nach der Anzahl der für die Lösung nötigen Verbindungen von Informationselementen vergeben. Das heißt, ein Item trägt umso mehr zum Gesamtscore bei, je mehr ICs oder RCs auf Elementebene seine Lösung voraussetzt, im Sinne des Komplexitätsbegriffs des Essener Modells. Wir treffen dabei die Annahme, dass alle zur Antwortgenerierung verwendeten Verbindungen als gleichwertig anzusehen sind. Dabei können bei manchen Items (vor allem bei dem DU-Typ) RCs und ICs in den Antworten nicht getrennt erfasst werden, sondern nur das Ergebnis einer Kombination von mehreren solchen Verbindungen, deren Punktzahl dann für die Beantwortung des Items vergeben wird. Die Komplexitätsstufe eines Items steigt mit der Anzahl aller für die Lösung nötigen Einzel-Verbindungen. Um je nach Komplexität (Anzahl der RCs und ICs) Punkte feinstufig zuordnen zu können, wurde das oben erwähnte Essener

Modell auf repräsentationale Kohärenzfähigkeit wie folgt angepasst: Die Komplexitätsniveaus 1 und 2 („Einzelfaktum“, „mehrere (unverbundene) Einzelfakten“) finden bei Aufgaben zu *Verbindungen* von multiplen Repräsentationen per Konstruktion keine Anwendung. Das Komplexitätsniveau 3 („Ein Zusammenhang“) entspricht bei Verbindungen von multiplen Repräsentationen per definitionem der niedrigsten Komplexitätsstufe. Bei den Komplexitätsniveaus 4 und 5 („Mehrere Zusammenhänge“) findet folgende Anpassung statt: Die Komplexitätseinstufung wird proportional zu deren Anzahl feiner aufgelöst; dabei kommen unverbundene Zusammenhänge (Niveau 4) bei mehrfachen RCs per Konstruktion nicht vor (ein mehrfacher Zusammenhang z. B. zwischen Text, Strahlendiagramm und Formel bezieht sich ja auf den denselben Sachverhalt). Insgesamt ergibt sich folgende Einstufung: ein Zusammenhang: 0,5 P („gering komplex“); zwei Zusammenhänge: 1 P („mittelkomplex“); vier Zusammenhänge: 2 P („hochkomplex“); sechs Zusammenhänge: 3 P („höchstkomplex“).

Um Missverständnisse zu vermeiden, sei darauf hingewiesen, dass es in der Operationalisierung von Komplexität (nach dem Essener Vorschlag, a. a. O.) hier um Verbindungen zwischen einzelnen *Informationselementen* geht, die von Verbindungen zwischen verschiedenen *Repräsentationsformen* zu unterscheiden sind. Ersteres meint herzustellende Verbindungen von einzelnen Informationselementen (die in der gleichen Repräsentationsform oder in verschiedenen Repräsentationsformen liegen können) und mit denen die Punktzahl des Tests operationalisiert wird (s. Beispiele oben), letzteres meint Kombinationen von ganzen Repräsentationsformen, die zur Lösung eines Items nötig sind (gezählt als N_{RFC}).

Wir fügen noch zwei Beispiele für Punktzweisungen bei höherkomplexen Antworten hinzu, bei denen mehrere Verbindungen zwischen einzelnen Informationselementen (IC und RC) in der Antwort integriert sind:

- mittelkomplex (1P, zwei Zusammenhänge). Aus DU-Item 2: „Ein Gegenstand befindet sich vor einer Linse. Die Gegenstandsweite liegt zwischen einfacher und doppelter Brennweite der Linse. Wie groß ist die Bildweite verglichen mit der Brennweite?“
 - RC: Gegenstandsweite in Abbildungsfalltabelle übertragen und Bildweite suchen
 - IC: $f < g < 2f \rightarrow b$ größer als $2f$ (zur anschließenden Angabe der Antwort wird der Tabellentext einfach übernommen, was keine Kohärenzleistung erfordert)
- hochkomplex (2P, vier Zusammenhänge). Aus DU-Item 1b: „Wie würde sich die Bildgröße und Bildweite verändern, wenn man den Gegenstand weiter von der Linse entfernt (gegeben: Strahlendiagramm eines Abbildungsexperiments mit Sammellinsen)?“

- RC: Gegenstand im Strahlendiagramm finden und nach links rücken
- IC: Feststellung: Mittelpunktstrahl wird flacher
- IC: Feststellung: Schnittpunkt Mittelpunktstrahl-Brennpunktstrahl ist näher an Achse
- RC: den im Strahlengang erkannten Sachverhalt als Text ausdrücken

Eine detaillierte Erläuterung des Punkteschlüssels für die Items findet sich im Onlinematerial 2.

Auf diese Weise entsteht ein Auswerteverfahren, das die Komplexität der repräsentationalen Verbindungen als schwierigkeiterzeugendes Merkmal der Items in Form der Anzahl der zu vergebenden Punkte berücksichtigt. Diese Herangehensweise wurde gewählt, weil sie eine konsistente Verbindung von theoretischen und unterrichtspraktischen Überlegungen darstellt, zum einen nämlich die theoretische Begründung durch das Essener Modell (s. oben), zum anderen ist eine Berücksichtigung der nötigen Anzahl von Lösungsschritten (hier: Verbindungen) auch gängige Bewertungspraxis im Unterricht.

Empirische Prüfung des RKF-Testinstruments

Die Untersuchung wurde an rheinland-pfälzischen Gymnasien durchgeführt, in denen der Bereich Strahlenoptik (Wirkung von Sammellinsen und Bildentstehung) überwiegend in der 8. Klassenstufe unterrichtet wird und fand im Rahmen des DFG-Graduiertenkollegs „Unterrichtsprozesse“ (DFG-Nr. GK 1561, Graduiertenkolleg Unterrichtsprozesse, Universität Koblenz-Landau, Campus Landau) statt.

Der Ablauf der Gesamtstudie gliedert sich in Pilotstudie und Hauptstudie, deren methodische Grundlagen im Folgenden beschrieben werden. In der Pilotstudie wurden zur Validierung zunächst Items auf Verständlichkeit geprüft, anschließend wurde eine erste psychometrische Itemanalyse durchgeführt. Anschließend folgte die Hauptstudie zur psychometrischen Prüfung des Testinstruments mit höherer Teilnehmerzahl. Als zusätzliche Fragestellung in der Hauptstudie wurde untersucht, ob die Items des RKF-Tests bezüglich des Bundeslandes Rheinland-Pfalz curricular valide sind oder sich stark von üblichen Schulaufgaben unterscheiden.

Pilotstudie

Eine Vorversion eines RKF-Tests mit 16 Items, welche zu sechs Aufgaben zusammengesetzt sind, wurde in zwei Schritten einer Prüfung auf folgende Kriterien unterzogen: Im ersten Schritt ($N = 30$, 7. Klasse einer Realschule) wurden die Aufgaben qualitativ auf Verständlichkeit der Instruktion geprüft, indem die Lernenden schriftlich Kommentare zu den Aufgaben geben konnten. Nach der Überarbeitung von unverständlichen Formulierungen wurden die

Aufgaben in einem zweiten Schritt ($N = 59$, 7. Klasse einer Realschule) nach Behandlung des Themas im Unterricht eingesetzt und qualitativ und quantitativ analysiert. Im Test wurden bei 6 Items Begründungen und Erklärungen im offenen Antwortformat gefordert, um qualitative Auskunft darüber zu erhalten, ob die Lösung in intendierter Weise aufgrund von RCs zustande kam, also im Sinne der o. g. Operationalisierung das Konstrukt repräsentationale Kohärenzfähigkeit misst, oder die Lernenden anhand von Oberflächenmerkmalen bzw. zufällig zur richtigen Lösung kamen. Bei keiner der fachlich richtig gelösten Aufgaben ergaben sich Hinweise auf nicht repräsentationale Kohärenzfähigkeit-basierte Lösungswege. Anschließend wurde zur Feststellung der Dimensionalität des Tests eine exploratorische Faktorenanalyse durchgeführt, bei der sich bis auf eine Aufgabe ein Generalfaktor zeigte. Bei einer Bestimmung der Kenngrößen des Tests zeigte diese Aufgabe außerdem nicht ausreichende psychometrische Kenngrößen (zu geringe Trennschärfe, zu hohe Schwierigkeit) und wurde daher aus dem Test entfernt. Die bereinigte Version des Tests enthielt 14 Items und ergab in der Pilotstudie akzeptable Werte für die Aufgabenschwierigkeiten ($0,32 < P_i < 0,92$; $\bar{P}_i = 0,61$) und Trennschärfen ($0,34 < r_{ii} < 0,56$; $\bar{r}_{ii} = 0,43$), sowie einen noch ausreichenden Wert für die interne Konsistenz ($\alpha_c = 0,65$). Sieben Items der Pilotstudie konnten nahezu unverändert in die Hauptstudie übernommen werden. Acht Items wurden in Hinblick auf eine weiter optimierte Umsetzung der Operationalisierung überarbeitet, fünf davon wurden stark überarbeitet. Die daraus entstandene Testversion enthielt 15 Items, die für den Einsatz in der Hauptstudie vorgesehen waren.

Hauptstudie

In der Hauptstudie wurde im Rahmen einer im o. g. DFG-Graduiertenkolleg durchgeführten Interventionsstudie zu drei Zeitpunkten Daten erhoben (s. Scheid 2013 für Details): Ein Prätest (t_1) direkt vor Beginn einer sechswöchigen Unterrichtsreihe über Bildentstehung an Sammellinsen mit bzw. ohne besonderen Fokus auf Repräsentationen, ein Posttest direkt danach (t_2) und ein Follow up-Test sechs Wochen nach deren Abschluss (t_3). An der Hauptstudie nahmen sechs rheinland-pfälzische Schulen mit 17 Klassen und insgesamt 488 Schülerinnen und Schülern teil. Die Schüler waren zwischen 12 und 14 Jahre alt ($M = 13,3$; $SD = 0,54$), davon waren 54 % weiblich und 46 % männlich.

Der folgende Hauptabschnitt des Artikels befasst sich mit der Bestimmung der üblichen psychometrischen Gütekriterien (Bühner 2011; Bortz und Döring 2002), die auch für normative Kompetenztests Anwendung finden (Schecker und Parchmann 2006). Die aus der Pilotstudie weiterentwickelte Testversion enthielt wie oben beschrieben ur-

sprünglich 15 Items, die in der Hauptstudie zum Einsatz kommen sollten. Davon wurde ein Item bei einer Voranalyse der Hauptstudienmaterialien wegen zu geringer Trennschärfe ausgeschlossen. Daher besteht der RKF-Test in der hier vorgestellten Version aus 14 Items, die wegen thematischer Verwandtschaft in acht Aufgaben zusammengefasst wurden. Wortlaut und Charakteristika des Tests finden sich in den Onlinematerialien. Für den Test standen 20 min zur Verfügung. Die Vergabe der Punkte für die einzelnen Items in der Auswertungsvorschrift (siehe Onlinematerial 2) geschah unter Berücksichtigung der Anzahl erfolgreicher Elementverbindungen sowie des oben erläuterten Komplexitätsmodells. Die Auswertung des Summenscores der repräsentationalen Kohärenzfähigkeit-Skala als intervallskalierte Größe und mit parametrischen Tests folgt dabei gängiger Praxis (Bortz und Schuster 2010; Bühner 2011), die aber weit über eine bloße „per fiat“-Setzung hinaus auf einer breiten Literatur-Basis und mehreren forschungsbasierten Argumenten beruht: dem zentralen Grenzwertsatz (Borgatta und Bohrnstedt 1980), Simulationen zum Vergleich mit und ohne Intervallskalen-Annahme (Brockett und Golden 1992), sowie der Robustheit parametrischer Verfahren gegenüber Verletzung der Annahmen in einem breiten Spektrum von Anwendungsbereichen (Dowling und Midgley 1991; Carifio und Perla 2007; Norman 2010).

Weil bei einem neuentwickelten Testinstrument nicht bekannt ist, wie viele Formen von repräsentationaler Kohärenzfähigkeit den Items ggf. zu Grunde liegen, sollte dies durch eine exploratorische Faktorenanalyse herausgefunden werden (Eid et al. 2011). Dabei wurde die quartimax-Rotation als Verfahren gewählt, die auch einen Generalfaktor zulässt, auf dem alle Items laden (Eid et al. 2011, S. 903). Die Anzahl der Faktoren wurde nach dem Scree-Test bestimmt (Faktoren sind relevant, wenn sie deutlich mehr Varianz erklären als die anderen Faktoren; Cattell 1966); dabei sollen sie auch nicht weniger Varianz aufweisen als einzeln beobachtbare Variablen (Kaiser-Kriterium: Eigenwert ≥ 1 ; Kaiser und Dickman 1959, Eid et al. 2011). Faktoren können unabhängig von der Stichprobengröße erst ab einer Faktorladung von $F_L \geq 0,4$ bei mindestens 10 Variablen bzw. ab einer Faktorladung von $F_L \geq 0,6$ bei mindestens 4 Variablen als einer Dimension zugehörig interpretiert werden (nach Guadagnoli und Velicer 1988).

Zur weiteren Klärung der Eignung von Items zur Testkonstruktion wird die interne Konsistenz α_C berechnet, welche den Zusammenhang der einzelnen Items untereinander angibt (Eid et al. 2011) und einen Schätzer für die Reliabilität darstellt. Folgende Bereiche werden angegeben:

- Werte von $\alpha_C \geq 0,60$ gelten als „zur Gruppendiagnostik geeignet“ (Evers 2001, EFPA 2013)

- Werte von $\alpha_C \geq 0,70$ gelten als akzeptabel (Bühner 2011, S. 81) und „zur Individualdiagnostik mit weniger hoher Tragweite geeignet“ (EFPA 2013).
- der Bereich $0,80 \geq \alpha_C \geq 0,90$ gilt als gut und „zur Individualdiagnostik mit hoher Tragweite geeignet“ (EFPA 2013).

Das Maß der korrigierten Item-Skala-Korrelation (Trennschärfe) ist hilfreich, wenn der Zusammenhang eines Items mit der Summe der übrigen Items ermittelt werden soll. Werte ab $r_i \geq 0,3$ gelten als akzeptabel (Weise 1975). Des Weiteren wird die Lösungswahrscheinlichkeit berechnet, da Werte, die außerhalb des erwünschten Bereichs von $0,20 \leq P_i \leq 0,80$ liegen, aufgrund von Boden- oder Deckeneffekten Unterschiede nicht gut erfassen können und so zu einer geringen Trennschärfe führen (Bortz und Döring 2002).

Als zusätzliche Fragestellung der Hauptstudie wurde untersucht, ob sich die Items des RKF-Tests stark von gewöhnlichen Schulaufgaben unterscheiden oder ob eine gewisse Konformität besteht. So wurde die curriculare Validität bezüglich des Bundeslandes Rheinland-Pfalz geprüft. Dies geschah durch ein Expertenrating mit 11 Gymnasiallehrern (mittlere Unterrichtserfahrung 14,5 (SD = 9,18) Jahre). Dabei sind die Einschätzungen zu folgenden Fragen erhoben worden (6-stufige Likertskala mit folgenden Antwortmöglichkeiten: „trifft überhaupt nicht zu“, „trifft nicht zu“, „trifft eher nicht zu“, „trifft eher zu“, „trifft zu“, „trifft voll und ganz zu“):

- Ich halte die Aufgabe inhaltlich für einen Leistungstest im Fach Physik geeignet.
- Diese Aufgabe ist lehrplankonform.

Ergebnisse

Faktoren- und Item-Analyse

Als Ergebnis der Faktorenanalyse zum ersten Messzeitpunkt (Hauptkomponentenanalyse, quartimax-Rotation konvergiert über drei Iterationen mit Kaiser-Normalisierung) ergab sich eine uneindeutige Dimensionsstruktur: Die Eigenwerte der Faktoren fielen zwar nach dem sechsten bzw. dritten Faktor mit gleichmäßiger Steigung ab, was auf fünf bzw. zwei Dimensionen hindeutet, jedoch ließen sich die Faktoren im fünf- bzw. zweifaktoriellen Modell inhaltlich nur widersprüchlich zuordnen und in anderen Modellen war die Itemanzahl pro Faktor selbst im Falle einer inhaltlichen Interpretierbarkeit unzureichend (vgl. Guadagnoli und Velicer 1988). Zum ersten Messzeitpunkt ließ sich also keine Dimensionsstruktur erkennen. Die Ergebnisse des zweiten und dritten Messzeitpunkts ergaben die im Folgenden beschriebene Dimensionsstruktur. Im Scree-Plot erkennt man, dass Faktor 1 mit Abstand den

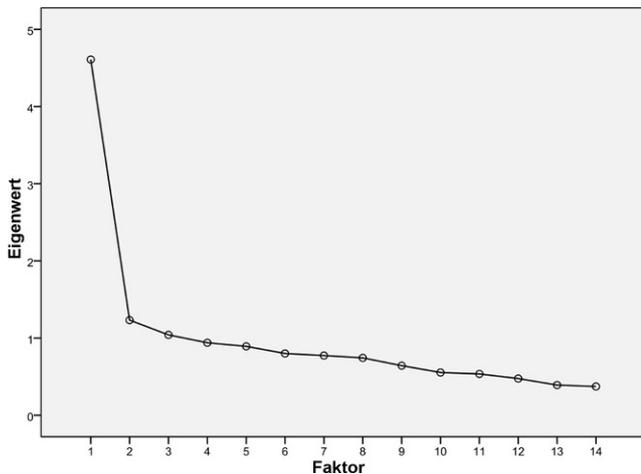


Abb. 5 Faktorenanalyse des RKF-Tests (Postzeitpunkt): Scree-Plot (Hauptkomponentenanalyse, quartimax-Rotation mit Kaiser-Normalisierung konvergiert über 8 Iterationen, Scree-Plot im dritten Messzeitpunkt nicht als abweichend erkennbar)

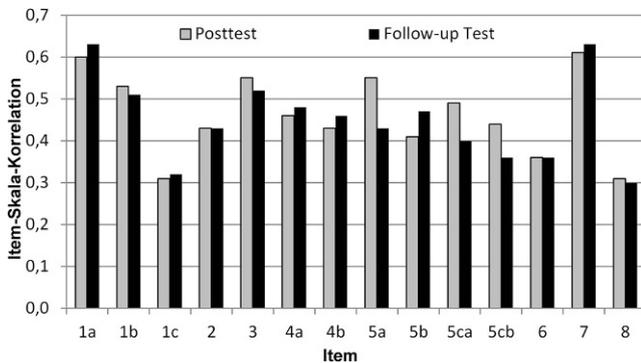


Abb. 6 Korrigierte Item-Skala-Korrelationen der Items des RKF-Tests

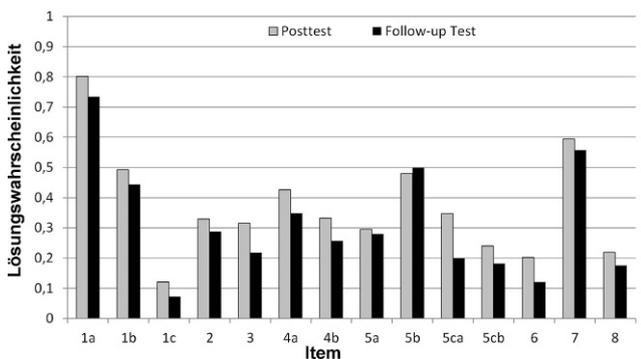


Abb. 7 Lösungswahrscheinlichkeiten der Items des RKF-Tests ($N(t_1) = 490$, $N(t_2) = 488$, $N(t_3) = 484$)

höchsten Eigenwert von 4,65 (post) bzw. 4,52 (follow-up) besitzt (Abb. 5). Die anderen Faktoren besitzen Eigenwerte im Bereich von Einzelvariablen oder darunter. Alle Items dieses Faktors außer Item 8 ($F_{L(t_3)} = 0,36$) weisen in der Komponentenmatrix der Hauptkomponentenanalyse Ladungen von $F_L = 0,4$ oder mehr auf $\bar{F}_{L(t_2)} = 0,56$ (0,10); $\bar{F}_{L(t_3)} = 0,56$ (0,09).

Die interne Konsistenz des RKF-Tests liegt zu den einzelnen Messzeitpunkten bei $\alpha_C(t_1) = 0,53$ ($N = 464$), $\alpha_C(t_2) = 0,79$ ($N = 488$) bzw. bei $\alpha_C(t_3) = 0,78$ ($N = 484$, Cronbach und Snow 1977). Durch das Ausschließen einzelner Items wurde keine Verbesserung der internen Konsistenz erreicht.

Abb. 6 zeigt die korrigierte Item-Skala-Korrelation, welche den Zusammenhang zwischen einzelnen Items mit der Summe der übrigen Items des RKF-Tests angibt. Im ersten Messzeitpunkt ($N(t_1) = 464$) korrelieren nur die Items 1a, 5b und 7 mit $r_{it} \geq 0,3$ mit der Summe der übrigen Items, im zweiten ($N(t_2) = 488$) und dritten Messzeitpunkt ($N(t_3) = 484$) korrelieren alle Items mit $r_{it} \geq 0,3$ mit der Summe der übrigen Items.

Im ersten Messzeitpunkt (direkt vor der Unterrichtssequenz) lagen die Lösungswahrscheinlichkeiten groÙtenteils unter dem Wert $P_i < 0,2$ (Abb. 7). Zu den übrigen Messzeitpunkten (direkt und 6 Wochen nach der Unterrichtssequenz) lagen die Lösungswahrscheinlichkeiten hingegen überwiegend im Bereich $0,2 < P_i < 0,8$ (bis auf Item 1c und Item 6 nur im Follow-up Test, die kleinere Werte aufwiesen). Die mittlere Lösungswahrscheinlichkeit (Bortz und Döring 2002) für den gesamten Test lag vor dem Unterricht über Bildentstehung an Sammellinsen bei $\bar{P}(t_1) = 0,16$ (0,09), direkt danach bei $\bar{P}(t_2) = 0,37$ (0,17) und sechs Wochen nach Abschluss bei $\bar{P}(t_3) = 0,31$ (0,18).

Die Auswertung der qualitativen Fragen bezüglich des eingeschlagenen Lösungswegs, welche in Item 5a und 5b gestellt wurden ergab, dass sich viele Lerner an Experimentiererfahrungen erinnerten und auf diese Weise das Item lösen konnten. Nur in selteneren Fällen wurden mental Strahlendiagramme konstruiert und verändert oder Tabellen genutzt. Alle Varianten erfordern jedoch wie intendiert repräsentationale Kohärenzfähigkeit, es werden lediglich andere Repräsentationsformen dazu verwendet.

Curriculare Validität

Für die Auswertung des Expertenratings stellt bei Annahme einer höchstens schwachen Verletzung der Gleichabständigkeit der Ratingskala (Baker 1966; Wirtz und Caspar 2002) eine unjustierte Intra-Class-Correlation (ICC) ein geeignetes Verfahren dar. Die Interrater-Übereinstimmung war jeweils höchstsignifikant und führte zu folgenden Ergebnissen:

- Elf Items wurden inhaltlich für einen Leistungstest im Fach Physik als „geeignet“ bzw. „voll und ganz geeignet“ eingeschätzt, zwei als eher geeignet (Items 5ca, 5cb) und eines als eher ungeeignet (Item 6). Der (unjustierte) *ICC*-Wert ist 0,61, was nach Cicchetti und Sparrow (1981) einer guten Übereinstimmung entspricht.
- Zehn Items wurden als „lehrplankonform“ bzw. „voll und ganz lehrplankonform“ eingeschätzt, drei als eher geeignet (Items 4a, 5ca, 5cb), eines als eher ungeeignet (Item 6). Der (unjustierte) *ICC*-Wert ist 0,52, was einer ausreichenden Übereinstimmung entspricht (nach Cicchetti und Sparrow 1981).

Interpretation und Diskussion

Die Item- und Testkennwerte des RKF-Tests lagen in den Post- und Follow-up-Messzeitpunkten durchweg in den psychometrisch erwünschten Bereichen. Betreffs der Dimensionalität des Tests hat sich faktorenanalytisch gezeigt, dass ein Faktor das bei weitem höchste Gewicht hat, drei andere Faktoren liegen jeweils in der Nähe des Eigenwertes von 1, also an der Grenze der Bedeutsamkeit als getrennte Faktoren. Weil repräsentationale Kohärenzfähigkeit auch innerhalb eines Themengebietes wie Strahlenoptik eine Kompetenz mit mehreren Anforderungsbereichen ist, enthält das Instrument mehrere Inhalte und Itemtypen. Diese haben jedoch einen starken inhaltlichen Zusammenhang, und die Ergebnisse der Faktorenanalyse sprechen für eine Eindimensionalität des Tests. Dies bedeutet bezogen auf die Theorie, dass das Konstrukt repräsentationale Kohärenzfähigkeit nicht in Komponenten zerlegt werden muss und es sowohl Standard- als auch DU-Items zu Grunde liegt. Cronbach's Alpha und die Item-Skala-Korrelationen als Indikatoren der Reliabilität und der Trennschärfe erreichten als akzeptabel bis gut angesehene Werte ($\alpha_C \approx 0,8$; $r_{it} \geq 0,3$; Nunnally und Bernstein 1994; Ding und Beichner 2009). Der Wert für Cronbach's Alpha ist gemäß gängiger Richtlinien (COTAN-System, Evers 2001; darauf beruhend: EFPA 2013) für Einzelfalldiagnostik mit weniger hoher Tragweite und Gruppenuntersuchungen zulässig. Einschränkend ist festzustellen, dass die interne Konsistenz aber keinen sehr guten Wert erreicht, und hinsichtlich Individualdiagnostik mit hoher Tragweite die Schwelle des empfohlenen Reliabilitätsbereiches nur annähernd erreicht. Diese Begrenzung wird bei der Belastbarkeit der gefundenen Ergebnisse mit berücksichtigt (Moosbrugger und Kelava 2012). Einige weitere Einschränkungen werden noch diskutiert.

Nach Unterricht liegen die Lösungswahrscheinlichkeiten der Items zwar bis auf einzelne Ausnahmen im Zielintervall (Bortz und Döring 2002, s. Diskussion unten), das Ergebnis der repräsentationalen Kohärenzfähigkeit liegt aber auch nach Unterricht und für curricular valide eingeschätz-

te Testitems (s. unten) im unteren Teil des Zielintervalls für die Lösungswahrscheinlichkeiten ($\bar{P}(t_2) = 0,37$, $\bar{P}(t_3) = 0,31$). Wir sehen unser Resultat im Einklang mit den Ergebnissen vorheriger Studien (vgl. Saniter 2003; Nieminen et al. 2010), wonach repräsentationale Kohärenzfähigkeit (hier im Bereich Strahlenoptik) für Lernende eine schwierige Anforderung darstellt.

Die Item- und Testkennwerte des RKF-Erhebungsinstruments lagen zum ersten Messzeitpunkt hingegen nicht durchweg in den psychometrisch erwünschten Bereichen. Vor Unterricht lag der Wert der Reliabilität ca. 10 % unterhalb des für Gruppendiagnostik benötigten Wertes, die Item-Skala-Korrelation ist gering und es ergeben sich sehr niedrige Lösungswahrscheinlichkeiten ($\bar{P}(t_1) = 0,16$). Ein theoretischer Bezug zur geringen Lösungswahrscheinlichkeit kann z. B. in DeFT (Ainsworth 2006) gesehen werden, hier wird geringe Kenntnis der Domäne als schwierigkeits-erzeugendes Merkmal beim Arbeiten mit multiplen Repräsentationen aufgeführt. Wie schon beschrieben wurde, ist es selbst nach Unterricht in der Domäne noch schwierig, repräsentationale Kohärenzfähigkeit herzustellen. Vor Unterricht ist es dem Ergebnis nach zu schwierig. Damit zusammenhängend ergibt sich ebenfalls eine Erklärung für geringe interne Konsistenz (und auch geringe Item-Skala-Korrelationen) zu diesem Zeitpunkt. Interne Konsistenz setzt eine Vernetzung des Wissens innerhalb eines Fachgebietes voraus und ist bei Vorliegen eines überwiegend geringen Wissensstands im Fachgebiet auch nur in geringem Ausmaß zu erwarten (Zeilik et al. 1997; Ramlo 2008, Lasry et al. 2011). Dies ist beispielsweise bei Konzepttests wohlbekannt (ebenda). Dass dies insbesondere auch bei Tests zutage tritt, die auf Kohärenz abzielen, dürfte wenig überraschen und wurde auch in früheren einschlägigen Studien gefunden (Nieminen et al. 2010). Die Ergebnisse des Instrumentes sind vor domänenspezifischem Unterricht deshalb nur sehr eingeschränkt interpretierbar. Vor Unterricht sind eine Individualdiagnose und damit auch eine differenzierte Analyse auf der Ebene von Änderungen zwischen Prä- und Post-Ergebnissen nicht möglich. Die Ergebnisse für die Lösungswahrscheinlichkeiten können aber als Indikatoren für den Lernfortschritt im Rahmen eines messwiederholten Designs zur Gruppendiagnostik und auch zur Kontrolle vergleichbarer Anfangsbedingungen im Rahmen von Vergleichsstudien sinnvoll sein, obwohl die Reliabilität des RKF-Tests vor Unterricht um ca. 10 % unter dem für Gruppendiagnostik notwendigen Wert liegt (Evers 2001; EFPA 2013). Das ist ähnlich wie beispielsweise bei Konzepttests, die trotz Eindimensionalität ein etwas breiter gefasstes Konstrukt enthalten können. Dies kann sich u. U. auf die Reliabilitätsinterpretation auswirken, denn das Maß der inneren Konsistenz ist eigentlich für eindimensionale Tests mit sehr eng gefasstem Konstrukt ausgelegt (vgl. Adams und Wieman 2010). Darüber hinaus

gehende Aussagen zu der Situation vor Unterricht sind nicht möglich.

Grenzen des vorgestellten Instruments liegen des Weiteren in einzelnen nicht ganz zufrieden stellenden Werten psychometrischer Indizes von Einzelitems. Die Lösungswahrscheinlichkeiten zu den Messzeitpunkten t_2 und t_3 sind bis auf Item 1c und Item 6 (nur Follow-up Messzeitpunkt) im gewünschten Bereich. Item 1c erfragt eine Begründung über die Konstruktion ausgezeichneter Strahlen, wie die Gegenstandsweite die Bildgröße beeinflusst, Item 6 erfragt den Grund für die Unmöglichkeit der Herleitung der Abbildungsgleichung aus zwei angegebenen Dreiecken eines Strahlengangs. Beide Items erfordern jedoch komplexe Vorgänge zur Lösungsgenerierung ($N_{RF} = 4$ bzw. 5), und damit ist eine hohe Schwierigkeit plausibel. Die Items leisten für die Erhebung einer hohen Ausprägung von repräsentationaler Kohärenzfähigkeit einen Erkenntnisgewinn. Dies spricht für ihre Beibehaltung; ebenso, dass sich durch ihren Ausschluss die Reliabilität α_C des Instruments nicht verbessert (s. Abschn. 3).

Die hohe curriculare Validität (inkl. Einschätzung der Eignung der Items für einen Leistungstest) wurde für 10 von 14 Items in dem Expertenrating mit guter Übereinstimmung bestätigt, für drei Items wurde curriculare Validität bestätigt. Dies bedeutet, dass es trotz deutlich von Standard-Anforderungen abweichender Operationalisierung bis auf lediglich eine Ausnahme gelungen ist, curriculärer Validität zu entsprechen. Nach einer Lerngelegenheit für repräsentationale Kohärenzfähigkeit (siehe Scheid et al. 2014, 2015) könnte ein Großteil des RKF-Tests möglicherweise in der Schule zur Leistungsbeurteilung verwendet werden (das „eher nicht“ geeignete Item 6 und je nach Lerngelegenheit auch „die eher geeigneten“ Items 4a, 5ca, 5cb sollten dazu aus der Bewertung genommen werden).

Auch die Validitäts-Einschätzung ist punktuell noch verbesserbar: Die Items 5ca/5cb wurden nur als „eher curricular valide“ eingeschätzt, sie fragen nach einem Lösungsweg, nicht nach der Lösung selbst. Dies ist in der Schule eher unüblich und könnte ein Grund für die eher verhaltene Zustimmung zur curricularen Validität sein. Item 6 (Einschätzung „curricular eher nicht valide“) enthält ein Thema (s. oben), das im Lehrplan von Rheinland-Pfalz nicht ausdrücklich gefordert wird. Somit ist die Einschätzung der Experten nachvollziehbar. Dieses Item sollte daher ggf. ausgeschlossen werden. Der Test ist zwar nur für das Land Rheinland-Pfalz curricular validiert worden, mit teilweiser Übereinstimmung sind seine Inhalte jedoch auch in den Ländern Berlin (Stufe 7/8) und Brandenburg (Stufen 9/10) vorgesehen. Mit weitgehender Übereinstimmung zum Lehrplan von Rheinland-Pfalz sind die Inhalte des Tests von den 8 Ländern Bremen und Hamburg (jeweils Stufe 8), Hessen (Stufe 7 G8, Stufe 8 G9), Nordrhein-Westfalen (Stufe 8), Thüringen (Stufe 8), Schleswig-Holstein (Stufe 8,

mit Abbildungsgleichung), Saarland (Stufe 9, im naturwissenschaftlicher Zweig mit Herleitung der Abbildungsgleichung) und Sachsen Anhalt (Stufe 6, Stufe 10 mit Herleitung der Abbildungsgleichung) vorgesehen, so dass eine Übertragbarkeit der Ergebnisse auf die genannten Bundesländer wahrscheinlich gegeben ist (Quellen der Lehrpläne in der angegebenen Reihenfolge: LISUM 2006, 2008; SBW 2006; LLS 2011; HK o.J.a, o.J.b; Qualis-NRW 2008; TMBWK 2012; MBWFK o.J.; MBKS 2014a und 2014b; MBSA 2016).

Hinsichtlich der Validität des Instrumentes als Ganzes bleibt kritisch betrachtet aufgrund einer theoretisch möglichen Verletztheit der Gleichabständigkeitsannahme der sechsstufigen Likertskala eine geringe Restwahrscheinlichkeit, dass die ICC die Rater-Übereinstimmung leicht überschätzt (Wirtz und Caspar 2002). Die gefundenen Intraklassen-Korrelationen sind jedoch genügend hoch, um das Expertenrating als reliabel einschätzen zu können. Das Instrument zur Erhebung der repräsentationalen Kohärenzfähigkeit kann daher insgesamt als „curricular valide“ eingeschätzt werden. Damit ist der Einsatz auch als Bewertungskomponente im Unterricht möglich, sofern zumindest die fachliche Thematik des Instrumentes schon im Unterricht behandelt wurde oder es explizit Lerngelegenheiten zur Entwicklung von repräsentationaler Kohärenzfähigkeit gab (Beispiele zu Aufgaben zur Förderung von repräsentationaler Kohärenzfähigkeit siehe Scheid et al. 2014, 2015). Ein solcher Einsatz ist wünschenswert, denn wie häufig gefordert (Black und Wiliam 1998; Bell 2007) sollen als didaktisch und lernpsychologisch wichtig erkannte Faktoren – wie hier repräsentationale Kohärenzfähigkeit – auch ihren Niederschlag in der Bewertungspraxis finden.

Schließlich ist noch folgende Limitierung der Studie einzuräumen: Das vorgestellte Instrument erlaubt es zwar, repräsentationale Kohärenzfähigkeit in praktikabler Weise und mit psychometrisch annehmbarer Güte zu erheben, und es werden auch bei allen Items entweder wichtige Schritte des Lösungswegs erfasst oder anderweitig (offene Items) Einblicke in die Vorgehensweise der Lernenden gegeben. Das Instrument ist aber nicht geeignet, Aufschluss über die kognitiven Prozesse beim gelingenden und misslingenden Entstehen von repräsentationaler Kohärenzfähigkeit zu liefern, wozu ergänzend weitere qualitative Daten und Methoden nötig wären. Diese Fragestellung ist von offensichtlichem Interesse für das Lernen mit multiplen Repräsentationen, geht aber über den hier gesteckten Rahmen hinaus.

Ausblick

Ein sinnvoller möglicher Einsatzzweck des Instrumentes ist die Diagnostik von repräsentationaler Kohärenzfähig-

keit auf Gruppen- oder Individualebene (für letzteres unter den o. g. Einschränkungen). Im Falle einer niedrigen Ausprägung können so Maßnahmen zur Förderung von repräsentationaler Kohärenzfähigkeit eingeleitet werden. Ein Beispiel für eine wirksame Maßnahme ist der Einsatz von „*Representational Analysis Tasks*“ (siehe z. B. Scheid et al. 2014). Auch für die Prüfung solcher Förderstrategien mittels Interventionsstudien kann der vorgestellte Test verwendet werden. Perspektivisch sind Untersuchungen von repräsentationaler Kohärenzfähigkeit in Verbindung mit anderen Kompetenzen wie z. B. Optik-Konzeptverständnis (siehe Hettmannsperger 2015) interessant, um mögliche Verbindungen zwischen verschiedenen Kompetenzbereichen besser zu verstehen – eine für Forschung und Unterrichtspraxis gleichermaßen fruchtbare Fragestellung. Schließlich erscheint es für die Zukunft sinnvoll, RKF-Tests für andere repräsentationsreiche Themengebiete im naturwissenschaftlichen Unterricht zu entwickeln (z. B. Elektrizitätslehre (Cheng 2002), elektromagnetische Felder (Törnkvist et al. 1993), die Verbindung von Mikro- und Makro-Ebene in der Chemie (Gilbert und Treagust 2009, Kap. 2, 8, 14), Genetik (Tsui und Treagust 2013, Kap. 7, 9) u. a. m.) und darauf zielende Lehr-Lern-Maßnahmen damit zu untersuchen.

Förderung Die Autoren danken der Deutschen Forschungsgemeinschaft für die Ermöglichung dieser Forschungsarbeit im Rahmen des Graduiertenkollegs „Unterrichtsprozesse“ (DFG-Nr. GK 1561, Graduiertenkolleg Unterrichtsprozesse, Universität Koblenz-Landau, Campus Landau).

Literatur

- Adams, W. A., & Wieman, C. E. (2010). Development and validation of Instruments to measure learning of expert-like thinking. *International Journal of Science Education*, 33(9), 1–24. doi:10.1080/09500693.2010.512369.
- Ainsworth, S. (1999). The functions of multiple representations. *Computers & Education*, 33, 131–152.
- Ainsworth, S. (2006a). DeFT: A conceptual framework for considering learning with multiple representations. *Learning and Instruction*, 16(3), 183–198.
- Ainsworth, S. (2006b). Reprinted from *Learning and Instruction*, 16(3), Ainsworth, S., DeFT: A conceptual framework for considering learning with multiple representations, 183–198, Copyright (2006), with permission from Elsevier.
- Ainsworth, S. (2008). The educational value of multiple representations when learning complex scientific concepts. In J. Gilbert, M. Reiner & M. Nakhleh (Hrsg.), *Visualization: theory and practice in science education* (S. 191–208). Dordrecht: Springer.
- Ainsworth, S. E., Bibby, P. A., & Wood, D. J. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences*, 11, 25–61.
- Anzai, Y. (1991). Learning and use of representations for physics expertise. In K. A. Ericsson & J. Smith (Hrsg.), *Towards a general theory of expertise: prospects and limits* (S. 64–92). Cambridge: Cambridge University Press.
- Ausubel, D. (1963). *The Psychology of Meaningful Verbal Learning*. New York: Grune & Stratton.
- Baddeley, A. D. (1986). *Working memory*. Oxford: Clarendon Press.
- Baker, B. O. (1966). Weak measurements vs. strong statistics: an empirical critique of S.S. Stevens Statistics. *Educational and Psychological Measurement*, 26, 291–309.
- Bell, B. (2007). Classroom assessment of science learning. In S. K. Abell & N. G. Lederman (Hrsg.), *Handbook of research in science education* (S. 965–1006). Mahwah: Lawrence Erlbaum.
- Berthold, K., & Renkl, A. (2009). Instructional aids to support a conceptual understanding of multiple representations. *Journal of Educational Psychology*, 101, 70–87.
- Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy and Practice*, 5(1), 7–74.
- Bodemer, D., Plötzner, R., Feuerlein, I., & Spada, H. (2004). The active integration of information during learning with dynamic and interactive visualisations. *Learning and Instruction*, 14(3), 325–341.
- Borgatta, E. F., & Bohrnstedt, G. W. (1980). Level of measurement. Once over again. *Sociological Methods and Research*, 9(2), 147–160.
- Bortz, J., & Döring, N. (2002). *Forschungsmethoden und Evaluation. Für Human- und Sozialwissenschaftler* (3. Aufl.). Berlin: Springer.
- Bortz, J., & Schuster, C. (2010). *Statistik für Human- und Sozialwissenschaftler*. Berlin: Springer.
- Brockett, P. L., & Golden, L. L. (1992). A comment on „using rank values as an interval scale“ by Dowling and Midgley. *Psychology & Marketing*, 9(3), 255–261.
- Bühner, M. (2011). *Einführung in die Test- und Fragebogenkonstruktion* (3. Aufl.). München: Pearson Studium.
- Carifio, J., & Perla, R. J. (2007). Ten common misunderstandings, misconceptions, persistent myths and Urban legends about Likert scales and Likert response formats and their antidotes. *Journal of Social Sciences*, 3(3), 106–116.
- Cattell, R. B. (1966). The scree test for the number of factors. *Multivariate Behavioral Research*, 1, 245–276.
- Chandler, P., & Sweller, J. (1992). The split-attention effect as a factor in the design of instruction. *British Journal of Educational Psychology*, 62, 233–246.
- Cheng, P. C. H. (2002). Electrifying diagrams for learning: principles for complex representational systems. *Cognitive Science*, 26(6), 685–736.
- Cicchetti, D. V., & Sparrow, S. S. (1981). Developing criteria for establishing interrater reliability of specific items: applications to assessment of adaptive behavior. *American Journal of Mental Deficiency*, 86, 127–137.
- Cook, M. P. (2006). Visual representations in science education: The influence of prior knowledge and cognitive load theory on instructional design principles. *Science education*, 90(6), 1073–1091.
- Cox, R., & Brna, P. (1995). Supporting the use of external representations in problem solving: the need for flexible learning environments. *Journal of Artificial Intelligence in Education*, 6, 239–302.
- Cronbach, L., & Snow, R. (1977). *Aptitudes and instructional methods: a handbook for research on interactions*. New York: Irvington.
- Ding, L., & Beichner, R. (2009). Approaches to data analysis of multiple-choice questions. *Physical Review Physics Education Research*, 5(2), 020103.
- Docktor, J. L., & Mestre, J. P. (2014). Synthesis of discipline-based education research in physics. *Physical Review Special Topics-Physics Education Research*, 10(2), 020119.
- Dolin, J. (2007). Science education standards and science assessment in Denmark. In D. Waddington, P. Nentwig & S. Schanze (Hrsg.), *Making it comparable. Standards in science education* (S. 71–82). Münster: Waxmann.
- Dowling, G. R., & Midgley, D. F. (1991). Using rank values as an interval scale. *Psychology & Marketing*, 8(1), 37.
- Eid, M., Gollwitzer, M., & Schmitt, M. (2011). *Statistik und Forschungsmethoden* (2. Aufl.). Weinheim: Beltz.
- EFPA (2013). *EFPA review model for the description and evaluation of psychological and educational tests*. Brussels: European Federation of Psychology Associations.

- Even, R. (1998). Factors involved in linking representations of functions. *Journal of Mathematical Behaviour*, 17, 105–121.
- Evers, A. (2001). Improving test quality in the Netherlands: results of 18 years of test ratings. *International Journal of Testing*, 1, 137–153.
- Gilbert, J. K., & Treagust, D. (Hrsg.). (2009). *Multiple representations in chemical education*. The Netherlands: Springer.
- Glenberg, A. M., & Langston, W. E. (1992). Comprehension of illustrated text: pictures help to build mental models. *Journal of memory and language*, 31(2), 129–151.
- Guadagnoli, E., & Velicer, W. F. (1988). Relation of sample size to the stability of component patterns. *Psychological Bulletin*, 103, 265–275.
- Guthrie, J. W. (Hrsg.). (2002). *Encyclopedia of education*. New York: Macmillan.
- Haapsalo, L., & Kadjevich, D. (2000). Two types of mathematical knowledge and their relation. *Journal für Mathematik-Didaktik*, 21(2), 139–157.
- Held, C., Knauff, M., & Vosgerau, G. (Hrsg.). (2006). *Mental models and the mind: current developments in cognitive psychology, neuroscience, and philosophy of mind* (1. Aufl.). Advances in Psychology, Bd. 138. Amsterdam: Elsevier.
- Herrmann, T. (1993). Mentale Repräsentation ein erläuterungsbedürftiger Begriff. In J. Engelkamp & T. Pechmann (Hrsg.), *Mentale Repräsentation* (S. 17–30). Bern: Huber.
- Hettmannsperger, R. (2015). *Lernen mit multiplen Repräsentationen aus Experimenten: Ein Beitrag zum Verstehen physikalischer Konzepte*. Wiesbaden: Springer.
- HK: Hessisches Kultusministerium (o. J. a). *Lehrplan Physik. Gymnasialer Bildungsgang, Jahrgangsstufen 7 bis 13*. Hessen. <https://kultusministerium.hessen.de/sites/default/files/media/g9-physik.pdf>. Zugegriffen: 23. November 2016.
- HK: Hessisches Kultusministerium (o. J. b). *Lehrplan Physik. Gymnasialer Bildungsgang, Jahrgangsstufen 6G bis 9G*. Hessen. <https://kultusministerium.hessen.de/sites/default/files/media/g8-physik.pdf>. Zugegriffen: 23. November 2016.
- Höfler, T. N., & Leutner, D. (2007). Instructional animation versus static pictures: a meta-analysis. *Learning and Instruction*, 17(6), 722–738.
- Hubber, P., Tytler, R., & Haslam, F. (2010). Teaching and learning about force with a representational focus: pedagogy and teacher change. *Research in Science Education*, 40(1), 5–28.
- Kaiser, H. F., & Dickman, K. W. (1959). Analytic determination of common factors. *American Psychologist*, 14, 425.
- Kauertz, A. (2008). Schwierigkeitserzeugende Merkmale physikalischer Leistungstestaufgaben. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* Bd. 79. Berlin: Logos.
- Kauertz, A., & Fischer, H. E. (2006). Assessing students' level of knowledge and Analysing the reasons for learning difficulties in physics by Rasch analysis. In X. Liu & J. Boone (Hrsg.), *Applications of Rasch measurement in science education* (S. 212–246). Minnesota: Maple Grove.
- Kohl, P. B., & Finkelstein, N. D. (2006). Effect of instructional environment on physics students' representational skills. *Physical Review Special Topics: Physics Education Research*, 2(1), 010102.
- Köller, O. (2008). Bildungsstandards – Verfahren und Kriterien bei der Entwicklung von Messinstrumenten. *Zeitschrift für Pädagogik*, 54(2), 163–173.
- Koziol, L. F., & Budding, D. L. (2012). Procedural learning. In N. M. Seel (Hrsg.), *Encyclopedia of the sciences of learning* (S. 2694–2696). New York: Springer.
- Kozma, R. B., & Russell, J. (2005). Students becoming chemists: developing representational competence. In J. K. Gilbert (Hrsg.), *Visualization in science education* (S. 121–146). Dordrecht: Springer.
- Kozma, R., Chin, E., Russell, J., & Marx, N. (2000). The roles of representations and tools in the chemistry laboratory and their implications for chemistry learning. *Journal of the Learning Sciences*, 9(2), 105–143.
- Krathwohl, D. R. (2002). A revision of bloom's taxonomy: an overview. *Theory Into Practice*, 41(4), 212–218.
- Lasry, N., Rosenfield, S., Dedic, H., Dahan, A., & Reshef, O. (2011). The puzzling reliability of the force concept inventory. *American Journal of Physics*, 79(9), 909–912.
- Lee, V. (2009). *Examining patterns of visual representation use in middle school science classrooms*. National Association of Research in Science Teaching (NARST) Annual Meeting Compact Disc. Garden Grove: Omnipress. conference paper
- Lesh, R., Post, T., & Behr, M. (1987). Representations and translations among representations in mathematics learning and problem solving. In C. Janvier (Hrsg.), *Problems of representation in the teaching and learning of mathematics* (S. 33–40). Hillsdale: Erlbaum.
- LISUM (2006). *Rahmenlehrplan für die Sekundarstufe I, Jahrgangsstufe 7–10, Hauptschule, Realschule, Gesamtschule, Gymnasium, Physik. Senatsverwaltung für Bildung, Jugend und Sport Berlin* (Hrsg.). Berlin: Oktoberdruck AG.
- LISUM (2008). *Rahmenlehrplan für die Sekundarstufe I, Jahrgangsstufen 7–10, Physik. Ministerium für Bildung, Jugend und Sport des Landes Brandenburg* (Hrsg.). Bielefeld: Hans Gieselmann Druck- und Medienhaus GmbH & Co KG.
- LLS: Landesinstitut für Lehrerbildung und Schulentwicklung (2011). *Bildungsplan Gymnasium, Sekundarstufe I, Physik*. Freie und Hansestadt Hamburg, Behörde für Schule und Berufsbildung. <http://www.hamburg.de/contentblob/2373266/f190388289579ec2aa1e7071b38aeb10/data/physik-gym-seki.pdf>. Zugegriffen: 23. Nov 2016.
- Mandl, H., & Levin, I. R. (Hrsg.). (1989). *Knowledge acquisition from text and pictures*. Amsterdam: North Holland.
- Mayer, R. E. (1989). Systematic thinking fostered by illustrations in scientific text. *Journal of educational psychology*, 81(2), 240.
- Mayer, R. E. (1997). Multimedia learning: are we asking the right questions? *Educational psychologist*, 32(1), 1–19.
- Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice*, 41(4), 226–232.
- Mayer, R. E. (2005). Cognitive theory of multimedia learning. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (S. 31–48). New York: Cambridge University Press.
- Mayer, R. E., & Anderson, R. B. (1991). Animations need narrations: an experimental test of a dual-coding hypothesis. *Journal of Educational Psychology*, 83(4), 484–490.
- MBKS: Ministerium für Bildung und Kultur Saarland (2014a). *Lehrplan Physik Gymnasium Klassenstufe 9 Sprachenzweig*. http://www.saarland.de/dokumente/thema_bildung/LP_Ph_Gym_9_sp.pdf. Zugegriffen: 23. Nov 2016.
- MBKS: Ministerium für Bildung und Kultur Saarland (2014b). *Lehrplan Physik Gymnasium Klassenstufe 9 Naturwissenschaftlicher Zweig*. http://www.saarland.de/dokumente/thema_bildung/LP_Ph_Gym_9_nw.pdf. Zugegriffen: 23. Nov 2016.
- MBSA: Ministerium für Bildung Sachsen Anhalt (2016). *Fachlehrplan Gymnasium*. http://www.bildung-lsa.de/pool/RRL_Lehrplaene/Erprobung/Gymnasium/FLP_Gym_Physik_LTn.pdf. Zugegriffen: 23. Nov 2016.
- MBWFK: Ministerium für Bildung, Wissenschaft, Forschung und Kultur des Landes Schleswig-Holstein (Hrsg.) (o. J.). *Lehrplan für die Sekundarstufe I der weiterführenden allgemeinbildenden Schulen Hauptschule, Realschule, Gymnasium, Physik*. <http://lehrplan.lernnetz.de/intranet1/index.php?DownloadID=36>. Gesehen 23. November 2016.
- MBWW – Ministerium für Bildung, Wissenschaft und Weiterbildung Rheinland-Pfalz (Hrsg.). (1998). *Lehrplanentwurf Physik – Realschule, Gymnasium (Klassen 7/8 – 10)*. Grünstadt: Sommer Druck und Verlag.

- McDermott, L. C. (2001). Oersted medal lecture 2001: physics education research – the key to student learning. *American Journal of Physics*, 69(11), 1127–1137.
- Moosbrugger, H., & Kelava, A. (2012). *Testtheorie und Fragebogenkonstruktion*. Berlin: Springer.
- 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, 125–143.
- Niegemann, H.M., Domagk, S., Hessel, S., Hein, A., Hupfer, M., & Zobel, A. (2008). *Kompendium multimediales Lernen*. Heidelberg: Springer.
- Nieminen, P., Savinainen, A., & Viiri, J. (2010). Force Concept Inventory-based multiple-choice test for investigating students' representational consistency. *Physical Review Special Topics – Physics Education Research*, 6(2), 020109.
- Nistal, A., Van Dooren, W., Clarebout, G., Elen, J., & Verschaffel, L. (2010). Representational flexibility in linear-function problems: a choice/no-choice study. In L. Verschaffel, E. De Corte, T. de Jong & J. Elen (Hrsg.), *Use of representations in reasoning and problem solving: analysis and improvement* (S. 74–93). Oxon: Routledge.
- Norman, G. (2010). Likert scales, levels of measurement and the „laws“ of statistics. *Advances in Health Science Education*, 15(5), 625–632.
- Nunnally, J., & Bernstein, I. (1994). *Psychometric theory* (3. Aufl.). New York: McGraw-Hill.
- Paivio, A. (1986). *Mental Representations: a dual coding approach*. New York: Oxford University Press.
- Peirce, C.S. (1958). *Collected Writings (Vols. 1–8)*. Cambridge: Harvard University Press. Erstausgabe 1931.
- Peterson, D. (Hrsg.). (1996). *Forms of representation: an interdisciplinary theme for cognitive science*. Exeter: Intellect Books.
- Plötzner, R., & Lowe, R. (2012). A systematic characterisation of expository animations. *Computers in Human Behavior*, 28(3), 781–794.
- Plötzner, R., & Spada, H. (1998). Constructing quantitative problem representations on the basis of qualitative reasoning. *Interactive Learning Environments*, 5, 95–107.
- Qualis-NRW (2008). Kernlehrplan für das Gymnasium – Sekundarstufe I, in Nordrhein-Westfalen, Physik. In Ministerium für Schule und Weiterbildung des Landes Nordrhein-Westfalen (Hrsg.), *In Schule in NRW, Heft 3411*. Frechen: Ritterbach.
- Ramlo, S. (2008). Validity and reliability of the force and motion conceptual evaluation. *American Journal of Physics*, 76(9), 882–886.
- Saniter, A. (2003). Spezifika der Verhaltensmuster fortgeschrittener Studierender der Physik. In H. Niedderer & H. Fischler (Hrsg.), *tudien zum Physiklernen* Bd. 28 Berlin: Logos.
- SBW: Senator für Bildung und Wissenschaft (2006). Naturwissenschaften, Biologie – Chemie – Physik, Bildungsplan für das Gymnasium, Jahrgangsstufe 5–10. Bremen. http://www.lis.bremen.de/sixcms/media.php/13/06-12-06_nat_gy.pdf. Zugegriffen: 23. Nov 2016.
- Scheid, J. (2013). Multiple Repräsentationen, Verständnis physikalischer Experimente und kognitive Aktivierung: Ein Beitrag zur Entwicklung der Aufgabenkultur. In H. Niedderer, H. Fischler & E. Sumfleth (Hrsg.), *Studien zum Physik- und Chemielernen* Bd. 151 Berlin: Logos.
- Scheid, J., Müller, A., Hettmannsperger, R., & Schnotz, W. (2014). Fostering the Understanding of Scientific Experiments and Phenomena through Representational Analysis Tasks. In C.P. Constantinou, N. Papadouris & A. Hadjigeorgiou (Hrsg.), *E-Book Proceedings of the ESERA 2013 Conference: Science Education Research For Evidence-based Teaching and Coherence in Learning. Part 1 (Ko-Hrsg. R. P. Casulleras, & M. Papaevripidou)* (S. 102–108). Nicosia: European Science Education Research Association.
- Scheid, J., Müller, A., Hettmannsperger, R., & Schnotz, W. (2015). Scientific experiments, multiple representations, and their coherence. A task based elaboration strategy for ray optics. In W. Schnotz, A. Kauertz, H. Ludwig, A. Müller & J. Pretsch (Hrsg.), *Multidisciplinary research on teaching and learning* (S. 239–252). Basingstoke: Palgrave Macmillan.
- Schecker, H., & Parchmann, I. (2006). Modellierung naturwissenschaftlicher Kompetenz. *Zeitschrift für Didaktik der Naturwissenschaften*, 12, 45–66.
- Schnotz, W. (1994). *Aufbau von Wissensstrukturen. Untersuchungen zur Kohärenzbildung bei Wissenserwerb mit Texten*. Weinheim: Beltz, Psychologie-Verlags Union.
- Schnotz, W. (2002). Towards an integrated view of learning from text and visual displays. *Educational Psychology Review*, 14(1), 101–119.
- Schnotz, W. (2005). An integrated model of text and picture comprehension. In R. E. Mayer (Hrsg.), *The Cambridge handbook of multimedia learning* (S. 49–70). New York: Cambridge University Press.
- Schnotz, W. (2010). Reanalyzing the expertise reversal effect. *Instructional Science*, 38(3), 315–323.
- Schnotz, W., & Bannert, M. (2003). Construction and interference in learning from multiple representation. *Learning and Instruction*, 13(2), 141–156.
- Schnotz, W., & Kürschner, C. (2007). A reconsideration of cognitive load theory. *Educational Psychology Review*, 19(4), 469–508.
- Schnotz, W., Baadte, C., Müller, A., & Rasch, R. (2010). Creative thinking and problem solving with depictive and descriptive representations. In L. Verschaffel, E. De Corte, T. de Jong & J. J. Elen (Hrsg.), *Use of representations in reasoning and problem solving. Analysis and improvement* (S. 11–35). Abingdon: Routledge.
- Schnotz, W., Baadte, C., Müller, A., & Rasch, R. (2011). Kreatives Denken und Problemlösen mit bildlichen und beschreibenden Repräsentationen. In R. Sachs-Hombach & R. Totzke (Hrsg.), *Bilder Sehen Denken* (S. 204–252). Köln: Halem.
- Schoenfeld, A.R., Smith, J.P., & Arcavi, A. (1993). Learning: the microgenetic analysis of one student's evolving understanding of a complex subject matter domain. In R. Glaser (Hrsg.), *Advances in instructional psychology* Bd. 4 Hillsdale: LEA.
- Schwonke, R., Berthold, K., & Renkl, A. (2009). How multiple external representations are used and how they can be made more useful. *Applied Cognitive Psychology*, 23, 1227–1243.
- Sell, K.S., Herbert, B.E., Stuessy, C.L., & Schielack, J. (2006). Supporting student conceptual model development of complex earth system through the use of multiple representations and inquiry. *Journal of Geoscience Education*, 54(3), 396–407.
- Seufert, T. (2003). Supporting coherence formation in learning from multiple representations. *Learning and Instruction*, 13, 227–237.
- Seufert, T., & Brünken, R. (2006). Cognitive load and the format of instructional aids for coherence formation. *Applied Cognitive Psychology*, 20, 321–331.
- Singer, S.R., Nielsen, N.R., & Schweingruber, H.A. (2012). *Discipline-based education research: understanding and improving learning in undergraduate science and engineering*. Washington, DC: National Academies Press.
- Sweller, J. (1999). *Instructional design in technical areas*. Melbourne: ACER Press.
- TMBWK: Thüringer Ministerium für Bildung, Wissenschaft und Kultur (2012). Lehrplan für den Erwerb der allgemeinen Hochschulreife Physik. https://www.schulportal-thueringen.de/tip/resources/medien/14477?dateiname=Physik_GY_Nov_2012_7_12.pdf. Zugegriffen: 23. Nov 2016.
- Törnkvist, S., Pettersson, K.A., & Tranströmer, G. (1993). Confusion by representation: on student's comprehension of the electric field concept. *American Journal of Physics*, 61(4), 335–338.
- Tsui, C., & Treagust, D. (2013). *Multiple Representations in Biological Education*. Dordrecht: Springer.

- Tytler, R., Prain, V., Hubber, P., & Waldrup, B. (Hrsg.). (2013). *Constructing representations to learn in science*. Rotterdam: Sense.
- Uttal, D.H., & O'Doherty, K. (2008). Comprehending and learning from „visualization“: a developmental perspective. In J.K. Gilbert, M. Reiner & M. Nakhleh (Hrsg.), *Visualization in science education* (S. 53–72). Dordrecht: Springer.
- Van Heuvelen, A. (1991). Learning to think like a physicist: a review of research-based instructional strategies. *American Journal of Physics*, 59, 891–897.
- Van Heuvelen, A., & Zou, X. (2001). Multiple representations of work-energy processes. *American Journal of Physics*, 69(184), 2001.
- Van Merriënboer, J.J.G., & Sweller, J. (2005). Cognitive load theory and complex learning: Recent development and future directions. *Educational Psychology Review*, 17(2), 147–176.
- Van Someren, M., Reimann, P., Boshuizen, H.P.A., & de Jong, T. (1998). *Learning with multiple representations*. Amsterdam: Elsevier.
- Verschaffel, L., De Corte, E., de Jong, T., & Elen, J. (2010). *Use of external representations in reasoning and problem solving*. New York: Routledge.
- Weise, G. (1975). *Psychologische Leistungstests*. Göttingen: Hogrefe.
- Wirtz, M., & Caspar, F. (2002). *Beurteilerübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.
- Wu, H.K., & Shah, P. (2004). Exploring visuospatial thinking in chemistry learning. *Science education*, 88(3), 465–492.
- Zeilik, M., Schau, C., Mattern, N., Hall, S., Teague, K.W., & Bisard, W. (1997). Conceptual astronomy: a novel model for teaching postsecondary science courses. *American Journal of Physics*, 65(10), 987–996.
- Zhang, J.J., & Norman, D.A. (1994). Representations in distributed cognitive tasks. *Cognitive Science*, 18(1), 87–122.