

Physik erklären – Messung der Erklärens-fähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation

Christoph Kulgemeyer · Elisabeth Tomczyszyn

Eingegangen: 25. September 2014 / Angenommen: 9. April 2015 / Online publiziert: 26. April 2015

© Gesellschaft für Didaktik der Physik und Chemie (GDPC); Fachsektion Didaktik der Biologie im VBIO (FDdB im VBIO) and Springer-Verlag GmbH Berlin Heidelberg 2015

Zusammenfassung Gut erklären zu können wird als wichtige Eigenschaft von Lehrkräften oft genannt, aber nur selten erforscht. In diesem Beitrag wird zunächst der theoretische Hintergrund von Unterrichtserklärungen herausgearbeitet, um diese von wissenschaftlichen Erklärungen abzugrenzen und spezifisch zu modellieren. Anschließend wird eine Methode vorgestellt, mit der es auf Basis dieses Modells gelingen kann, aus Videos von Erklärungen auf die Ausprägung der Erklärens-fähigkeit der Erklärenden zurückzuschließen. Dazu wird eine standardisierte Erhebungsmethode verwendet, die möglichst real eine Erklärens-situation nachstellt: die Experten-Novizen-Dialogmethode. Dabei erklärt eine angehende Physiklehrkraft einer Schülerin ein physikalisches Phänomen. Die Schülerin darauf wurde trainiert, in jeder Erklärens-situation vergleichbare Fragen zu stellen bzw. Prompts zu geben. Die Situationen werden gefilmt und kategorienbasiert ausgewertet. Die Kategorien, aus denen sich das Maß für Erklärens-fähigkeit zusammensetzt, werden präsentiert (z. B. Umschreibung von Fachbegriffen in Alltagssprache). Aus dem Auftreten der Kategorien in den Erklärungen wird ein quantitatives Maß entwickelt, das die Ausprägung von Erklärens-fähigkeit repräsentieren soll. Dazu wurde eine Expertenbefragung durchgeführt, um die Validität des Testverfahrens überprüfen zu können. Die Ergebnisse zeigen, dass das Maß in der Lage ist, die Expertenentscheidung bei

Paarvergleichen über den besseren von zwei Erklärern vorherzusagen ($\kappa=0,750$), es ist zudem ausreichend reliabel ($\alpha=0,772$). Auch bei einem zweiten Erhebungszeitpunkt derselben Erklärer mit anderen Themen und Adressaten der Erklärung zeigen sich weitgehend stabile Urteile über den besseren Erklärer (Übereinstimmung des Urteils in zwei Messzeitpunkten: $\kappa=0,600$).

Schlüsselwörter Erklären · Unterrichtserklärung · professionelle Handlungskompetenz · Performanztest

Explaining physics: assessing physics explaining competence with a teaching performance test

Abstract Explaining is often regarded as an integral part of science teachers' professional competence. In this paper we present a way to measure explaining competence. As a theoretical framework we use a model on science teaching explanations. We adopt an assessment method for science communication competence that is close to real teaching situations. This method is based on an expert-novice dialogue: A teacher trainee (expert) explains a physics phenomenon to a high-school student (novice) in a controlled test setting which is standardized by topic, duration (10 min) and especially by standardized prompts given by the high-school student (e.g., "Can you give me another example?"). We videotape this dialogue. To measure explaining competence we firstly analyse the videos qualitatively. Categories and indicators for the quality of explaining were derived inductively from the reactions to these prompts, and deductively from the theory of explaining. These categories were used to develop an initial quantitative measure for explaining skills. In a successive process of improving the reliability of the measure and its validity

C. Kulgemeyer (✉) · E. Tomczyszyn
Institut für Didaktik der Naturwissenschaften,
Abt. Physikdidaktik, Universität Bremen,
Otto-Hahn-Allee 1,
28359 Bremen, Deutschland
E-Mail: kulgemeyer@physik.uni-bremen.de

E. Tomczyszyn
E-Mail: tomczyszyn@physik.uni-bremen.de

by comparing it with what the experts regard as good explaining, we reduced the number of categories needed to identify the level of explaining skills.

Keywords Explaining · Science teaching explanation · Professional competence · Performance test

Einleitung

Physik erklären zu können, wird als eine wesentliche Fähigkeit von Physiklehrkräften gesehen – und manchmal sogar als Kern des Lehrens von Naturwissenschaft aufgefasst: „Providing explanations ist the bread and butter of the science teacher’s existence“ (Osborne und Patterson 2011, S. 632) oder „Explaining may come close to being the essence of instruction“ (Gage 1968, S. 3). Zudem ist Erklären nicht einfach. Merzyn (2005) kam bei seiner Befragung von Physikreferendaren zu der aus seiner Sicht „etwas überraschenden“ Erkenntnis, dass Erklären als die schwierigste Herausforderung beim Unterrichten von Physik wahrgenommen werde. Von Schülerinnen Schülern wird es aber hoch geschätzt: Wilson und Mant (2011a) finden bei einer Befragung von 5044 Schülerinnen und Schülern in England, dass die Gruppe der besonders gut beurteilten Lehrkräfte vor allem durch ihre Fähigkeit, erklären zu können, hervorsticht – während die Lehrkräfte selbst das Erklären nicht als Qualitätskriterium ihres Unterrichts erwähnen (Wilson und Mant 2011b).

Das ist ein Hinweis darauf, dass Erklären in der naturwissenschaftsdidaktischen Ausbildung unterrepräsentiert ist. Geelan (2012) zeigte zudem in einem Review der Forschungsliteratur zum Erklären, dass die naturwissenschaftsdidaktische Forschung sich mit gutem Erklären höchstens am Rande beschäftigt. Angesichts der Stimmen, die eine hohe Relevanz des Erklärenkönnens betonen, ist dies zunächst überraschend. Der Grund liegt möglicherweise darin, dass Erklären mit dozierendem Frontalunterricht und lehrerzentriertem Handeln assoziiert wird, also einer Unterrichtsform, die geradezu den Grundsätzen einer konstruktivistischen Auffassung des Lernens zu widersprechen scheint (vgl. Kulgemeyer 2013). In diesem Beitrag wird eine konstruktivistische Auffassung des Erklärensprozesses vertreten und durch ein theoretisch begründetes Modell beschrieben. Zudem wird ein Ansatz gezeigt, wie durch videogestützte Analysen von Erklärungen die Ausprägung adressatengemäßer Erklärensfähigkeit bestimmt werden kann. Der Ansatz baut auf einer simulierten unterrichtlichen Erklärungssituation auf, die das Potential hat, einen Aspekt professioneller Handlungskompetenz von Physiklehrkräften zu erheben (ein sog. „Performanztest“, vgl. Miller (1990)) ohne dabei auf schriftliche Testformate und deren bekannte Probleme zurückzugreifen (vgl. Vogelsang 2014).

Der Fokus liegt dabei auf mündlichen Erklärungen sowie den hierbei auftretenden Interaktionen zwischen Erklärenden und Adressaten der Erklärung. Schriftliche Erklärungstexte werden hier nicht betrachtet, deren Verständlichkeit folgt zudem zumindest teilweise anderen Kriterien (Kulgemeyer und Staraschek 2014).

Theoretischer Hintergrund

Um Erklären als didaktische Handlung zu verstehen und die adressatengemäße Komponente des Erklärens zu beschreiben, müssen drei wesentliche theoretische Abgrenzungen getroffen werden. Das sind vor allem die Unterscheidungen zwischen wissenschaftlichen Erklärungen und Unterrichtserklärungen (Abschn. 2.1) sowie zwischen Erklärungen und Argumentationen (Abschn. 2.2). Zudem relevant ist die hier in Kap. 3 vorgenommene Unterscheidung zwischen dem Prozess des Erklärens und seinem Produkt, der Erklärung. Da in diesem Beitrag der Prozess des mündlichen Erklärens mit Vermittlungsabsicht beleuchtet werden soll, ist es schließlich notwendig, die kommunikative Natur des Erklärens zu operationalisieren und insbesondere ihre adressatengemäße Komponente hervorzuheben.

Wissenschaftliche Erklärungen und Unterrichtserklärungen

Begriffliche Unterscheidungen

Der Begriff des Erklärens bzw. der Erklärung wird in der naturwissenschaftsdidaktischen Literatur nicht einheitlich verwendet. Man findet im Wesentlichen zwei verschiedene Bedeutungsfelder. Zum einen wird die Erklärung als etwas typisch Naturwissenschaftliches verstanden (z. B. nach Dilthey, vgl. Kiel (1999)), zum anderen als didaktische Handlung. Im ersten Falle wird sich auf die Wissenschaftstheorie bzw. die Natur der Naturwissenschaften bezogen. Wissenschaftliche Erklärungen sind in der Literatur insbesondere in der Philosophie ein viel behandelter Gegenstand (z. B. Bartelborth 2007). Erklären in diesem Sinne meint vereinfachend das Zurückführen eines Phänomens auf ein zugrundeliegendes Prinzip. Diese Prinzipien können unterschiedlicher Natur sein. Es gibt in der Literatur eine Vielzahl an Unterscheidungen zwischen verschiedenen Erklärungsformen und eine große Diskussion über deren Wissenschaftlichkeit. Die wohl meist beachtete Form der Erklärung ist die *deduktiv-nomologische Erklärung* (Hempel und Oppenheim 1948). Aus Sicht von Hempel und Oppenheim (1948) muss beim Erklären eine logische Beziehung zwischen einem Phänomen, allgemeinen Prinzipien sowie Randbedingungen, unter denen das Phänomen auftritt, hergestellt werden. Das Phänomen (bei Hempel und Oppenheim „explanandum“) ist dann deduktiv aus dem

zugrundeliegende allgemeine Gesetz („explanans“) ableitbar. Bei voller Kenntnis aller Rahmenbedingungen und des allgemeinen Gesetzes kann das Auftreten des Phänomens also vorhergesagt werden. Oftmals ist dies die einzige Form, die als wissenschaftliche Erklärung anerkannt wird. In der Folge haben sich aber andere akzeptierte Erklärungsformen ausgebildet. Die (*induktiv-*)*statistische Erklärung* (Rescher 1962) ist letztlich eine Erweiterung der deduktiv-nomologischen Erklärung. Hier ist der Schluss vom allgemeinen Gesetz auf das Auftreten des Phänomens nicht mehr logisch eindeutig; das Auftreten des Phänomens ist lediglich wahrscheinlich. Die Erklärung beruht auf der Erkenntnis, dass ein Phänomen unter bestimmten Rahmenbedingungen häufig auftritt. Nagel (1961) erkennt (im Gegensatz zu Hempel und Oppenheim) zudem noch *funktionale bzw. teleologische Erklärungen* als wissenschaftlich an. Beispielsweise ist die Frage, warum ein Pfau ein Rad schlägt, nicht deduktiv-nomologisch zu erklären, da es kein allgemeines Gesetz gibt, das dieses vorhersagt. Das Rad ist nur über seine Funktion in der Balz zu erklären – aber andere Vögel haben anderen Formen der Balz entwickelt. Eine weitere Form der Erklärung ist die *narrative Erklärung*, die z. B. für Historiker aber auch Psychologen bedeutend ist. Hier wird ein zeitliches Ereignis erklärt, indem dargelegt wird, wie es zustande kam. Dabei ist das Ereignis oftmals sehr unwahrscheinlich, tritt aber dennoch ein.

Kitcher (1981) gibt eine allgemeinere Theorie der wissenschaftlichen Erklärung wieder. Eine Erklärung ist in diesem Sinne immer dann befriedigend, wenn sie die Anzahl an Fakten (Zusammenhängen, Gesetzen) reduziert, die man zur Ableitung eines Sachverhalts heranziehen muss. Eine wissenschaftliche Erklärung ist dann gut, wenn sie möglichst viele Fakten aus möglichst wenigen Zusammenhängen bzw. allgemeinen Gesetzen ableiten kann. Im Sinne von Hempel und Oppenheim muss jedoch mindestens hinzugefügt werden, dass diese Gesetze empirisch überprüfbar und wissenschaftlich als wahr akzeptiert sein müssen.

Treagust und Harrison (1999) nennen diese Art, von Erklären zu sprechen, „scientific explanations“ und unterscheiden sie von „science teaching explanations“. Im Deutschen kann man analog von „wissenschaftlichen Erklärungen“ und „Unterrichtserklärungen“ sprechen. Letztere haben immer eine Vermittlungsabsicht: Inhalte werden so aufbereitet, dass sie für eine bestimmte Gruppe von Adressatinnen und Adressaten bzw. für ein konkretes Individuum verstehbar werden. Erklären mit Vermittlungsabsicht hat also einen *kommunikativen Kern*. Beide Arten, von Erklären zu sprechen, können sich deshalb im Resultat deutlich unterscheiden. Wissenschaftliche Erklärungen bedürfen nicht notwendigerweise kommunikativer Überlegungen über etwaige Adressaten der Erklärung, hier geht es vor allem um sachstrukturelle Erwägungen: ist der Schluss auf das zugrundeliegende Prinzip nachvollziehbar bzw. genügend

wahrscheinlich und mit den richtigen Daten illustriert? Sind die Prinzipien akzeptiert und falsifizierbar? Bei Unterrichtserklärungen ist hingegen gerade die Überlegung, was das Gegenüber benötigt, um zu verstehen, entscheidend. Das kann z. B. bedeuten, dass Beispiele verwendet werden, um einen abstrakten Sachverhalt zu veranschaulichen. In diesem Fall verläuft Erklären dann gerade umgekehrt zu dem, was es bei wissenschaftlichen Erklärungen meint: hier wird das abstrakte allgemeine Gesetz durch etwas Konkretes veranschaulicht – anstatt etwas Konkretes (ein Phänomen) durch etwas Abstraktes (das zugrundeliegende Prinzip) zu beschreiben. Dennoch haben beide einen gemeinsamen Kern: eine Erklärung soll etwas verständlich machen (Brewer et al. 2000). Erklären hat also einen doppelten Wortsinn: zum einen meint es das logische Begründen und zum anderen das Veranschaulichen durch ergänzende Informationen.

Empirische Untersuchungen zu Unterrichtserklärungen

Geelan (2012) hat den Forschungsstand zu Unterrichtserklärungen aufgearbeitet und insgesamt lediglich 24 Studien gefunden, die sich damit auseinandergesetzt haben. Die meisten davon haben sich jedoch nicht mit dem Messen oder Modellieren von Erklärens-fähigkeit oder mit der Qualität von Erklärungen beschäftigt, sondern vielmehr die Rolle des Erklärens im Unterricht beleuchtet. Es finden sich lediglich zwei Studien, die die Beschreibung mündlicher Unterrichtserklärungen thematisieren. Norris et al. (2005) haben analysiert, welche Eigenschaften narrative Erklärungen im Naturwissenschaftsunterricht aufweisen (siehe auch Abschn. 2.3 zu Ogborn et al. (1996)), also Erklärungen, die ihre Konzepte in Form von Geschichten darlegen. Ihre Analyse führt zu Beschreibungskriterien für narrative Erklärungen, die angewendet werden können. Sie unterscheiden nicht zwischen mündlichen und schriftlichen Erklärungen und können auch keine Qualitätsmaßstäbe ableiten. Anders sind Sevian und Gonsalves (2008) vorgegangen. Sie haben tatsächliche Erklärungen von Wissenschaftlerinnen und Wissenschaftlern, die diese vor einem Publikum über ihre Forschung unternommen haben, analysiert. Ziel war, ein Beschreibungsraster zu finden, das die Qualität der Erklärungen geeignet abbildet. In ihrem theoretischen Konzept findet sich zudem ein neuer Ansatz, da sie die Verbindung zwischen der Fähigkeit, erklären zu können und dem Professionswissen von Lehrkräften herstellen. Sie bauen ihr Beschreibungsraster auf den Theorien im Anschluss an Shulman (1987) auf und unterteilen die notwendigen Handlungen beim guten Erklären in die drei Bereiche Fachwissen, fachdidaktisches Wissen und pädagogisches Wissen. Damit suchen sie den theoretischen Anschluss an eine didaktische bzw. pädagogische Theorie, wenn ansonsten über Erklärungen in den Naturwissenschaften vorrangig sachstrukturell nachgedacht wird. Ihr Beschreibungsraster

bleibt jedoch hochinferent und ist als diagnostisches Instrument problematisch (z. B. Kategorie „Der Erklärung mangelt es an Struktur, sie verwirrt dadurch“ (Sevian und Gonsalves 2008, S. 1449); Übersetzung des Autors). Auch Sevian und Gonsalves (2008) arbeiten zudem den kommunikativen Kern, der im mündlichen Erklären steckt, nicht heraus. Über die Wirkung der Erklärungen und eine Validierung ihres Kriteriensystems hinsichtlich der Verständlichkeit der Erklärungen liegen keine Ergebnisse vor.

Erklären und Argumentieren

Deutlich umfangreicher erforscht als das Erklären mit Vermittlungscharakter ist das Argumentieren im Naturwissenschaftsunterricht. Tatsächlich wird Argumentieren und Erklären oftmals als einander sehr ähnlich verstanden (Osborne und Patterson 2011; Berland und McNeill 2012). Dies betrifft aber vor allem wissenschaftliche Erklärungen, weniger Unterrichtserklärungen.

Osborne und Patterson (2011, S. 629) beispielsweise geben an, der Zweck des Argumentierens sei, eine Behauptung zu unterstützen, während Erklärungen vor allem darauf abzielen, zu verstehen, warum und wie ein Phänomen auftritt. In diesem Sinne verstehen Berland und McNeill (2012, S. 808) Erklären und Argumentieren als nicht identische, sondern sich gegenseitig ergänzende Praktiken. Beide haben tatsächlich strukturelle Eigenheiten gemein. Beim Argumentieren nach Toulmin (1958) wird mit Daten und einer Schlussregel, warum diese so verwendet werden dürfen, eine Behauptung („claim“) gerechtfertigt. Beim deduktiv-nomologischen Erklären wird ein Sachverhalt („explanandum“) durch eine zugrundeliegende Theorie („explanans“) erklärt, die Verbindung zwischen beiden kann auch durch Daten und Schlussregeln gerechtfertigt werden (vgl. Osborne und Patterson, 2011).

Eine wichtige Unterscheidung zwischen Erklärung und Argumentation liegt allerdings in der Gewissheit, mit der man von den begründenden allgemeinen Gesetzen – also dem explanans – ausgehen kann. Bei Argumentationen werden andere Maßstäbe akzeptiert als bei Erklärungen. Während bei Erklärungen nur solche allgemeinen Gesetze als Begründungen akzeptiert werden, die empirisch überprüfbar und allgemein wissenschaftlich als wahr anerkannt sind (s. Abschn. 2.1.1), werden bei Argumentationen (insbesondere im Alltag) auch Plausibilitäten oder ad hoc konstruierte Gesetze verwendet. Dies liegt daran, dass die wesentliche Unterscheidung zwischen Argumentation und Erklärung deren Intention ist. Fleischhauer (2013, S. 10) führt an, dass Argumentationen von einer Position überzeugen, während Erklärungen etwas durch Rückführung auf höhere Prinzipien verständlich machen sollen. Dies kann man weiter ausdifferenzieren und auf eine Unterscheidung zwischen Argumentationen, wissenschaftlichen Erklärun-

gen und Unterrichtserklärungen beziehen: Argumentationen sollen durch geeignete Daten ein Gegenüber von einer Behauptung überzeugen. Wissenschaftliche Erklärungen sollen die Beziehung zwischen einem explanandum und einem explanans darlegen, dann gilt das Phänomen als wissenschaftlich erklärt. Dazu ist kein gedachter Adressat notwendig. Die Natur der Unterrichtserklärungen hingegen ist eine andere: hier soll *jemand* etwas verstehen. Auch dazu ist die Rückführung auf ein Prinzip – eine Ursache – notwendig, aber der Weg, der beschritten wird, kann sich je nach Adressat deutlich unterscheiden. Unterrichtserklärungen wollen bezwecken, dass *ein Adressat das Phänomen versteht*, das ist etwas Anderes und möglicherweise komplexer, als den Adressaten von etwas zu überzeugen. Beim Erklären würde man sich nicht damit zufriedengeben, dass jemand eine Erklärung akzeptiert, weil er glaubt, sie verstanden zu haben. Selbst wenn Schülerinnen und Schülern bereits vorunterrichtlich gefestigt von falschen Erklärungen physikalischer Phänomene überzeugt sind, ist das Verstehen der anschlussfähigeren Erklärungen das Ziel – der Weg dahin könnte allerdings über das Überzeugen verlaufen. Der Aspekt der Adressatengerechtigkeit von Erklärungen – also was eine Erklärung gut verständlich macht – ist in der Naturwissenschaftsdidaktik bislang sehr selten beschrieben, aber in anderen Wissenschaften wie der Psychologie durchaus thematisiert worden.

Gütekriterien für verständliches Erklären im Unterricht

Ogborn et al. (1996) präsentieren eine der wenigen naturwissenschaftsdidaktischen Arbeiten zum verständlichen Erklären. Sie haben erste Schritte zur Beschreibung des Erklärens im Unterricht als kommunikative Praxis unternommen. Dies geschieht allerdings ohne Bezug zu kommunikationstheoretischen Modellen allein auf Basis eines Grundverständnisses von wissenschaftlichen Erklärungen. Ogborn et al. (1996) schließen an die Grundform der narrativen Erklärung an und gehen davon aus, dass letztlich alle wissenschaftlichen Erklärungen auch in Form einer Geschichte dargeboten werden können. Es gebe dort große Parallelen, so würden in Geschichten Protagonisten mit bestimmten Eigenschaften auftreten, im Sinne dieser Eigenschaften handeln und aus diesen Handlungen Konsequenzen folgen (vgl. Ogborn et al. 1996, S. 9). Diese Konsequenzen würden durch die Geschichte sinnvoll hergeleitet – dies sei ähnlich zu wissenschaftlichen Erklärungen und darin liege ein großes didaktisches Potential. Dies ist aus philosophischer Sicht eine diskussionswürdige Sicht auf wissenschaftliche Erklärungen, es wird aber eine pragmatische Brücke zwischen wissenschaftlichen Erklärungen und Unterrichtserklärungen geschlagen. Zudem wird von Ogborn et al. (1996) der Prozess des Erklärens näher beleuchtet und nicht lediglich das Produkt einer Erklärung analysiert. Dies ist

wesentlich für eine kommunikationstheoretische Sicht auf das Erklären; die Interaktion zwischen mehreren an der Kommunikation beteiligten Parteien wird dadurch betont.

Außerhalb der naturwissenschafts-didaktischen Literatur ist Erklären im Unterricht allerdings als kommunikativer Akt beschrieben worden. Der Psychologe Brown (2006) z. B. begreift das Erklären im Unterricht als besondere kommunikative Anforderung. Aus seiner Sicht lassen sich Elemente finden, die gute Unterrichtserklärungen aufweisen. Auch der Pädagoge Wellenreuther (2005) führt eine solche Aufzählung von Elementen guten Erklärens im Unterricht an. In Kulgemeyer und Schecker (2013) wurde zudem untersucht, welche Elemente in als besonders gut eingeschätzten Erklärungen von Schülerinnen und Schülern häufig auftauchen. Diese Elemente decken sich ganz wesentlich mit den von Brown (2006) und Wellenreuther (2005) angeführten (Kulgemeyer und Schecker 2013, S. 2252):

1. Vorbereitung auf die Erklärung
2. Gebrauch von graphischen Darstellungsformen, die die gesprochene Sprache veranschaulichen
3. Einbezug des bzw. der Adressaten in die Erklärung, z. B. durch direkte Ansprache oder Handlungsaufforderungen
4. Evaluation des Verständnisses nach einer Erklärungs-passage
5. Kurze und prägnante Antworten auf direkte Fragen
6. Beispiele oder Analogien, die das zu Erklärende auf einen bereits bekannten Sachverhalt beziehen
7. Berücksichtigung des Vorwissens des bzw. der Adressaten
8. Betonung besonders relevanter Punkte
9. Gelegenheiten geben, um Fragen zu stellen
10. Ein logischer Bezug zwischen inhaltlichen Teilen der Erklärung („roter Faden“)

Diese zehn Punkte können verwendet werden, um gutes Erklären zu identifizieren. Es gelingt damit jedoch nicht, zu analysieren, wie adäquat ein Erklärender auf den Adressaten der Erklärung reagiert – also wie genau eine Erklärung verändert wird, um verständlich zu werden. Dies zu beschreiben ist Zweck des im Folgenden vorgestellten Modells.

Erklärung und Erklären: Der Prozess des Erklärens im Modell

Ziel dieses Beitrags ist es, Unterrichtserklärungen so zu analysieren, dass auf die Erklärens-fähigkeit von (angehenden) Lehrkräften geschlossen werden kann. Dazu soll hier das Modell vorgestellt werden, nach dem die Analysen gerichtet sind.

Eine Unterrichtserklärung richtet sich immer an einen oder mehrere Adressaten. Sie kann nach ihrer Sachstruktur analysiert werden und damit den Anschluss an das Ver-

ständnis wissenschaftlicher Erklärungen schaffen. Sie kann aber auch nach der Adressatengemäßheit beleuchtet werden. Um zu operationalisieren, was Adressatengemäßheit meint, braucht es ein kommunikationstheoretisches Modell des Erklärens. Wir beziehen uns dazu auf das Modell von Kulgemeyer und Schecker (2009a), das als Modell naturwissenschaftlicher Kommunikationsfähigkeit von Schülerinnen und Schülern entworfen wurde, und modifizieren es der neuen Zielgruppe entsprechend leicht. Das Modell beruht auf einem konstruktivistischen Kommunikationsmodell nach Rusch (1999). Darin wird Kommunikation als Akt aufgefasst, den Sender und Empfänger gleichermaßen aktiv betreiben. Der Sender stellt dabei Angebote zur Kommunikation bereit, die vom Empfänger genutzt werden können oder nicht – selbstverständlich können sich die Rollen nach jedem Turn auch umdrehen. Kulgemeyer und Schecker (2009a) haben das Modell auf fachlich-physikalische Kommunikation bezogen und zeigen können, dass damit der physikbezogene Kommunikationsprozess von Schülerinnen und Schülern beschrieben (Kulgemeyer und Schecker 2013) sowie der Kompetenzbereich Kommunikation modelliert werden kann (Kulgemeyer und Schecker 2012). Tomas und Ritchie (2014) haben sich auf das Modell bezogen, um Schülerinnen und Schülern durch das Verfassen adressatengerechter Texte Grundlagen der Biologie zu vermitteln. Im Rahmen des Modells wird im adaptiven Erklären der Kern fachlicher Kommunikation gesehen. Es zeigte sich auch empirisch, dass das Modell die Vorgänge beim Erklären unter Schülerinnen und Schülern geeignet beschreiben kann (Kulgemeyer und Schecker 2013). Es bietet also gute Voraussetzungen, um auch Unterrichtserklärungen von Lehrkräften zu modellieren.

Dazu soll zunächst festgelegt werden, was unter dem Prozess des Erklärens verstanden werden soll:

Der Prozess des Erklärens mit Vermittlungsabsicht besteht in der adressatengemäßen und sachgerechten Erstellung und Modifikation von Unterrichtserklärungen.

Im Sinne von z. B. Clark und Brennan (1991) kann man bei kommunikativen Prozessen vom Suchen eines gemeinsamen geteilten Referenzrahmens, des sogenannten *common grounds*, sprechen. Adressatengemäße Kommunikation macht es aus, dass dieser common ground getroffen wird, dabei kann man sich sukzessive an den common ground annähern. Deshalb sind Modifikationen der Unterrichtserklärungen notwendig.

Dies soll am Beispiel eines Physiklehrers erläutert werden, der einer Schülerin etwas erklärt. Der genannten Auffassung vom Prozess des Erklärens folgend wird Erklären stets so ablaufen, dass der Lehrer zunächst eine initiale Erklärung anbietet. In die Art, wie er diese Erklärung gestaltet hat, sind Annahmen darüber eingegangen, was die

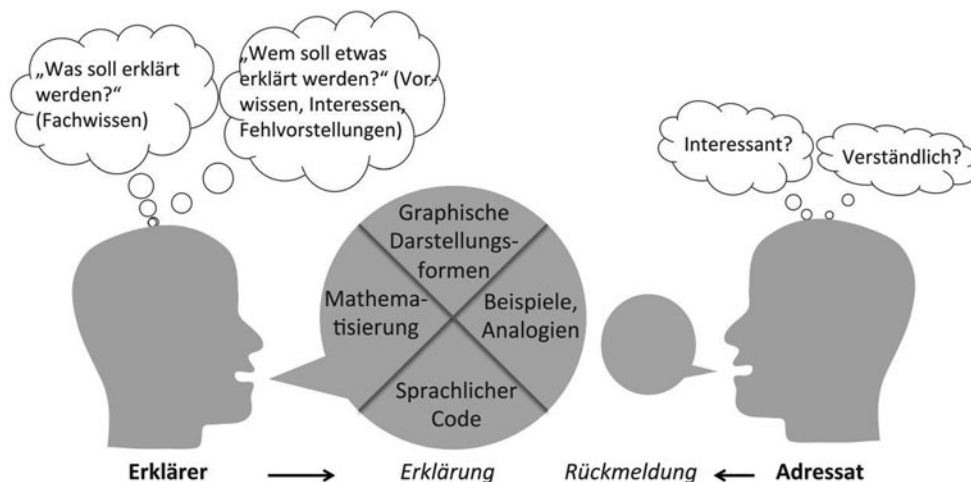
Schülerin bereits weiß und was sie noch benötigt, damit der Sachverhalt verständlich ist – aber natürlich sind auch Überlegungen über den physikalischen Sachverhalt eingeflossen. Der Lehrer muss dann durch aktive Evaluation (z. B. durch Fragen oder Aufgaben) – oder auch passiv (z. B. durch non-verbale Gesten der Schülerin oder Fragen, die sie stellt) erfahren, ob seine Erklärung verständlich war bzw. an welcher Stelle Unklarheiten oder Missverständnisse geblieben sind. Die Schülerin meldet also etwas darüber zurück, ob sie das Angebot angenommen hat und wie sie es nutzen konnte. Entscheidet der Lehrer aufgrund der Rückmeldung, dass das Beabsichtigte noch nicht vermittelt wurde, muss er in einem zweiten Schritt Modifikationen an seinem Angebot vornehmen und die Erklärung umgestalten. Auch daran schließt sich wieder eine Rückmeldung an und der Prozess beginnt erneut bis von beiden Seiten eine befriedigende Übereinkunft darüber getroffen wurde, dass der Inhalt verstanden wurde. Selbstverständlich kann es sein, dass auch dann – objektiv gesehen – noch kein Verständnis vorliegt, es ist aber zumindest eine für zwei Personen subjektiv befriedigende Basis gefunden worden, die ausgebaut werden kann.

Nachdem der Prozess makroskopisch beschrieben ist, muss nun mikroskopisch modelliert werden, wie adressatengemäße und sachgerechte Modifikationen einer Erklärung verstanden werden können. Dazu dient das Kommunikationsmodell von Kulgemeyer und Schecker (2009a). Kulgemeyer und Schecker (2009a) beschreiben jede sprachliche Äußerung mit Kommunikationsintention als ein sogenanntes *Kommunikat*, das bestimmte Eigenschaften aufweist: es hat einen bestimmten (fachlichen) *Sachinhalt*, der in einem *Kontext* dargestellt wird, dazu können verschiedene *Darstellungsformen* und verschiedene *sprachliche Codes* – z. B. Fach- oder Alltagssprache – verwendet werden. Darstellungsformen meinen dabei im engeren Sinne graphische Darstellungsformen, also logische Bilder, Bildanalogien oder realistische Bilder (Kulgemeyer und Schecker 2009b) sowie reale Gegenstände. Für die Physik ist zudem die

mathematische Darstellung wichtig (z. B. Krey 2012) – je abstrakter bzw. allgemeiner Sachverhalte betrachtet werden, desto wichtiger ist sie vermutlich. Zur Beschreibung von Unterrichtserklärungen wird deshalb hier an das Modell von Kulgemeyer und Schecker (2009a) angeschlossen, dieses aber um den Mathematisierungsgrad als Eigenschaft erweitert. Mathematisierungen in der Physik können z. B. durch mündlich beschriebene je-desto-Zusammenhänge vorgenommen werden, Krey (2012, 141 ff.) unterscheidet zudem zwischen graphischen und symbolischen, d. h. algebraischen, Repräsentationen. Zudem wird weiterhin zwischen sprachlichem Code und Darstellungsform unterschieden. Zwar zeigte sich bei Kulgemeyer und Schecker (2012), dass kompetentes Handeln in diesen beiden Aspekten vermutlich dieselbe Fähigkeit als Grundlage haben, doch bei mündlichen Erklärungen scheint es besonders wichtig zu sein, Gesagtes und Gesehenes aufeinander zu beziehen – dies wird nämlich in der Literatur als Qualitätskriterium gesehen (siehe Abschn. 2.3, Kriterium 2).

Abbildung 1 veranschaulicht den Prozess des Erklärens. Er kann so verstanden werden, dass jede Erklärung der oben genannten Begriffsfestlegung folgend einen sprachlichen Akt darstellt, dessen Adressatengemäßheit und Sachgerechtigkeit sich in vier „Variablen“ ausdrückt: Graphische Darstellungsform, Mathematisierung, Beispiele/Analogien sowie sprachlicher Code. Nach der Rückmeldung des Adressaten kann die Erklärung in diesen vier Variablen umgestaltet werden, um die Interessen des Adressaten besser zu bedienen und die Verständlichkeit der Erklärung zu verändern. Der Erklärende hat dabei bei jeder sprachlichen Äußerung die sachgerechte Perspektive („Was soll erklärt werden?“) und die adressatengemäße Perspektive („Wem soll etwas erklärt werden?“) zu bedenken und die vier Variablen entsprechend zu wählen. Die Rückmeldungen des Adressaten geben neue Informationen über die adressatengemäße Perspektive und führen deshalb zur Variation der Erklärung. Es ist eine wesentliche Anforderung an die Erklärens-fähigkeit,

Abb. 1 Der Prozess des Erklärens mit Vermittlungscharakter von Physik



zum einen die Erklärung in diesen vier Variablen zu gestalten und zum anderen die Rückmeldungen des Adressaten so zu interpretieren, dass Rückschlüsse gezogen werden, die zur zielgerichteten Modifikation der Erklärung führen.

Aus konstruktivistischer Sicht muss natürlich betont werden, dass Erklären nicht die Suche nach der idealen Darstellungsform ist, die dann in Verständnis resultiert. Eine adressatengemäße Darstellungsform macht allerdings die Konstruktion von Sinn auf Seiten des Adressaten leichter und führt somit wahrscheinlicher zu einem Verstehen. Aus diesem Grund wird auch nicht wie in klassischen Sender-Empfänger-Modellen Informationstransfer beschrieben. Die Erklärung ist vielmehr ein Angebot, wie dieses genutzt wird, kann der Erklärende nicht beeinflussen.

Forschungsfragen und Methodik

In der Literatur ist bislang kein Ansatz zu finden, der Unterrichtserklärungen mit Vermittlungsabsicht niedriginferent einschätzen oder Erklärens-fähigkeit von Lehrkräften in ihrer Ausprägung messen kann. Das hier vorgestellte Modell für Erklärens-fähigkeit und die Gütekriterien mündlicher Erklärungen sollen die Grundlage bilden, Erklärens-fähigkeit auch quantitativ messen zu können. Oftmals wird diskutiert, inwiefern mit schriftlichen Tests überhaupt die professionelle Handlungskompetenz von Lehrkräften erhoben werden kann (Aufschnaiter und Blömeke 2010; Vogelsang und Reinhold 2013). Hier soll explizit das Handeln in einer unterrichtsnahen Handlungssituation analysiert werden. Dabei soll auf die Ansätze in der Literatur aufgebaut werden, die sich auf die Analysen realer Erklärungshandlungen beziehen. Deswegen wird ein Ansatz verfolgt, der auf der Analyse möglichst realer Erklärungssituationen aufbaut und diese durch Standardisierungen vergleichbar macht (Experten-Novizen-Dialog, siehe Abschn. 4.1). Im Sinne von Millers (1990) für die Medizin getroffene Unterscheidung von Assessmentformanten handelt es sich um einen Performanztest, der die Umsetzung von berufsbezogener Kompetenz in einer simulierten Handlungssituation erfasst. Schriftliche Testformate können demnach zwar Kompetenztests im engeren Sinne darstellen, der Rückschluss auf Performanzqualität ist aber nicht selbstverständlich; die Ergebnisse zur Erforschung professioneller Handlungskompetenz im Physiklehramt zeigen dazu passend auch keine eindeutigen Zusammenhang zwischen professioneller Kompetenz und Unterrichtsqualität (z. B. Olszewski 2010; Ohle et al. 2011; Vogelsang 2014). Ein Performanztest einer ausgewählten nachgestellten Unterrichtssituation, dem Erklären, wäre ein Novum in der Diagnostik professioneller Handlungskompetenz und hinsichtlich der Validität besonders wertvoll.

Im Folgenden werden Argumente für die Validität präsentiert und das Setting detailliert beschrieben.

Die Forschungsfrage lautet:

F1: Wie kann aus den Gütekriterien für Erklärungen und dem Modell des Erklärens ein quantitatives Maß für adressatengemäße Erklärens-fähigkeit erstellt werden?

Erhebungsinstrument: der Experten-Novizen-Dialog

Die Erhebung der Daten erfolgt nach dem handlungsorientierten diagnostischen Verfahrens des Experten-Novizen-Dialogs (Kulgemeyer und Schecker 2013). Darin werden Erklärende dabei gefilmt, wie sie einem Adressaten mit geringerem Wissensstand ein Phänomen erklären – etwa vergleichbar einer Nachhilfesituation. Ihnen ist dabei jedoch nicht bekannt, dass die Adressatin bzw. der Adressat im Vorfeld geschult wurde, sich so zu verhalten, dass die Erklärenden ihre Erklärungsansätze verändern müssen – z. B. indem die Adressatin oder der Adressat gezielt Unverständnis zeigen und vorgegebene Nachfragen stellt. Jede dieser Nachfragen (Prompts) stellt eine Anforderung an die Erklärens-fähigkeit dar und dadurch, dass gleiche Fragen in allen Erklärungssituationen vorkommen, werden die einzelnen Situationen miteinander vergleichbar.

Gestaltung der Erhebungsumgebung¹

Die Rahmenbedingungen der Erklärungssituation sind genau festgelegt: ein Erklärender bekommt zehn Minuten Vorbereitungszeit für die Erklärung und wird danach zehn Minuten beim Erklären gefilmt. Dazu wird ihm ein Thema vorgegeben. In der Studie, auf die sich dieser Beitrag bezieht, wurden fünf Themen verwendet, deren physikalischen Schwerpunkte sowohl in der universitären Physikausbildung als auch der Schulphysik thematisiert werden:

1. Warum gleitet man bei nasser Straße leichter aus der Kurve als bei trockener Straße? (Physikalische Schwerpunkte: Reibung zwischen Oberflächen, Kräfte bei Kreisbewegungen)
2. Film „Armageddon“: warum gelingt es, einen Asteroiden, der auf die Erde zurast, durch eine Sprengung in zwei Hälften ausreichend von der Bahn abzulenken? (Physikalische Schwerpunkte: Impulserhaltung, Superpositionsprinzip)
3. Warum fühlt man sich beim Durchfahren mancher parabelförmiger Hügel in einer Achterbahn schwerelos? (Physikalische Schwerpunkte: Schwerelosigkeit, Wurfbewegungen)
4. Warum ist es besonders schwierig ein U-Boot, das auf Grund gelaufen ist, vom Meeresboden anzuheben? (Physikalische Schwerpunkte: Druck, Auftriebskraft)

¹ Die Materialien sind auf Anfrage bei den Autoren erhältlich.

5. Warum trägt eine Knautschzone zur Minderung des Verletzungsrisikos von PKW-Insassen bei? (Physikalische Schwerpunkte: Kraft, Energieumwandlung)

Es wurden fünf Themen gewählt, um eine hohe Bandbreite an Schwerpunkten zuzulassen und dabei nicht Gefahr zu laufen, dass das zu entwickelnde Kategoriensystem (siehe Abschn. 4.2.2) zur Beschreibung von Erklärungen themenabhängig ist. Zudem sollten für die Studie auswertbare Ergebnisse erhalten werden, auch wenn eines der Themen sich in den Analysen als schwieriger erklärbar als die anderen erweisen sollte – im Notfall könnte eines aus der Auswertung gestrichen werden. Dies überwog bei der Konzeption das Argument der besseren Vergleichbarkeit von Erklärungen durch eine geringere Anzahl an Themen.

Am Beginn wird dem Erklärenden mitgeteilt, dass es sich bei den Adressaten um einen Schüler oder eine Schülerin der zehnten Klasse handelt, der bzw. die kaum physikalisches Vorwissen zu dem Themenbereich mitbringt. Dann bekommt der Erklärende die Aufgabenstellung schriftlich sowie den Auftrag, sich auf das Erklären vorzubereiten. Zur Vorbereitung stehen dem Erklärenden verschiedene Materialien zur Verfügung. In der Aufgabenstellung steht explizit, dass er alle Materialien und Notizen mit in die Erklärung nehmen kann, dass er aber nichts davon verwenden muss.

Zu den zur Verfügung stehenden Materialien zählen Papier und Stifte, insbesondere aber vorbereitete Diagramme und Formeln. Zu jedem Thema gibt es sechs bis acht Darstellungsformen, die gezielt mit verschiedenen Abstraktionsgraden entwickelt wurden (realistische Bilder, Bildanalogien, logische Bilder (vgl. Sumfleth und Gnoyke 1995) – siehe Abb. 2) und aus denen der Erklärende auswählen kann. Diese Materialien sprechen den Kontext und physikalische Konzepte an, die zum Erklären herangezogen

werden können (z. B. eine Achterbahn und Wurfparabeln in Abb. 2). Sie liefern dem Erklärer jedoch keine fertige physikalische Erklärung des Phänomens.

Nach der Vorbereitung wird er in den Erhebungsraum geleitet. Im Erhebungsraum sitzt der Erklärende an einem Tisch neben dem Adressaten. Die Kamera befindet sich im Hintergrund. Auf dem Tisch befinden sich leere Zettel, mehrere Stifte, eine Getränkeflasche und zwei Gläser, ansonsten ist nichts außer den vom Erklärer mitgebrachten Materialien greifbar. Der Testleiter startet die Kamera, verlässt den Raum und kehrt nach zehn Minuten zurück. Dann wird die Erklärung beendet. Bei Kulgemeyer und Schecker (2013) hat sich gezeigt, dass nach zehn Minuten in der Regel noch ein Gespräch stattfindet, da die Adressaten dazu geschult sind, das Gespräch am laufen zu halten, aber bereits alle Inhaltsaspekte gestreift wurden. Ebenfalls melden die Erklärenden in einer Voruntersuchung zurück, dass sie nicht mehr Zeit zum Erklären benötigt hätten.

Training der Schülerinnen und Schüler als Adressaten der Erklärung

Dem Training der Adressaten kommt besondere Bedeutung zu. Sie müssen glaubwürdig Wissensbedarf kommunizieren, damit sich die Erklärerinnen und Erklärer ernsthaft um das Erklären bemühen. Dies funktioniert über Nachfragen, die sich im Gesprächsverlauf ergeben. Gleichzeitig müssen die Adressaten Fragen stellen, die zwischen den Erklärungssituationen eine Vergleichbarkeit schaffen und somit die Erhebungsmethode standardisieren. Diese Fragen müssen an passenden Stellen in den Gesprächsverlauf eingebunden werden. Sie sind sowohl gedacht, um die Erklärung abstrakter, als auch, um sie einfacher zu gestalten. Beispiele für solche Fragen sind:

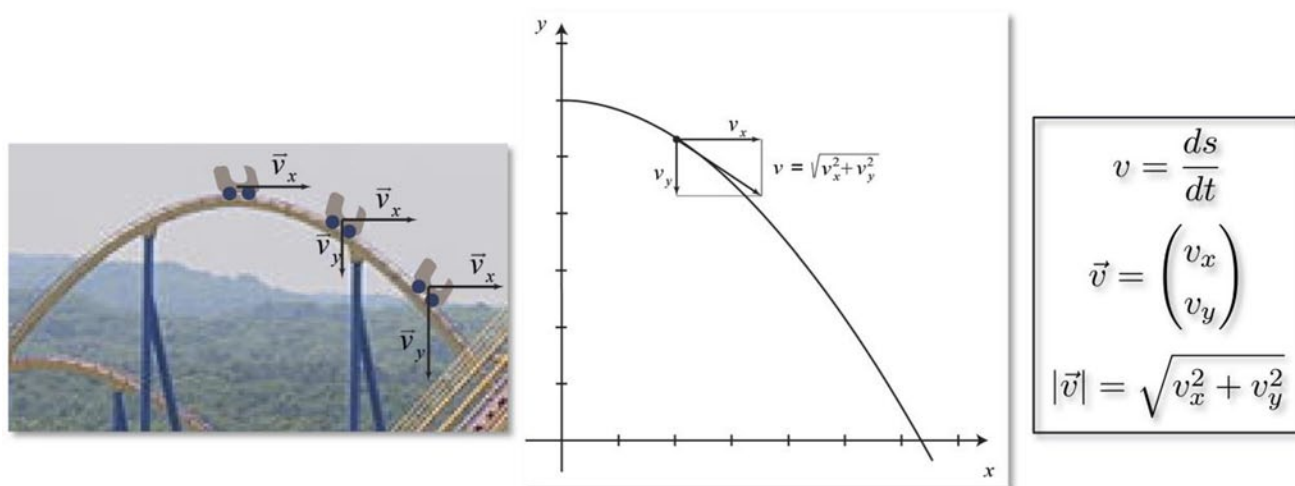


Abb. 2 Beispiele für vorgefertigte Darstellungsformen zum Thema „Schwerelosigkeit in der Achterbahn“

1. Das habe ich noch nicht verstanden. Kannst du das nochmal leichter erklären?
2. Die Grafik ist mir zu kompliziert.
3. Gibt's da noch ein anderes Beispiel für?
4. Soweit habe ich es verstanden, aber da gibt es doch sicher auch Formeln für. (Ich brauche das für eine Klausur.)

Zudem sollten die Adressaten dafür sorgen, dass die Erklärenden in den zehn Minuten alle wesentlichen physikalisch-inhaltlichen Sachverhalte des gestellten Themas (siehe 4.1.1) abdecken und sich nicht nur mit einem davon beschäftigen. Dies ist für die inhaltliche Vergleichbarkeit der Erklärungssituationen wichtig und soll es ermöglichen, die fachliche Korrektheit vergleichend bewerten zu können. Beim Thema „Kurvenfahrt“ soll so z. B. auf das Thema Kreisbewegungen gelenkt werden, wenn der Erklärende nur über Reibung spricht. Dazu könnte der Adressat in einer Gesprächspause fragen „Ist das nicht die Fliehkraft? Die drückt mich doch nach außen, wenn ich um die Kurve fahre.“ In diesem Fall wird gleichzeitig ein stark schülervorstellungshaltiger Bereich angesprochen.

Das Training der Adressaten nahm zwischen zwei und drei Stunden pro Person in Anspruch. Dabei wurden sie zunächst mit der Rollenbeschreibung, den Fragen, die sie stellen sollen und mit zwei der Erklärungsthemen vertraut gemacht. Dann haben sie die Durchführung der Erhebung mit einem Testleiter simuliert. Die Erklärungen wurden gefilmt und die Adressaten haben gemeinsam mit dem Testleiter den Film angesehen und kritisch über die Einbettung ihrer Fragen in den Verlauf der Erklärung reflektiert. Nach etwa drei solcher Probedurchläufe konnte von den Testleitern eine erfolgreiche Vorbereitung festgestellt werden. Nur in wenigen Fällen erwiesen sich einzelne Adressaten im Training als nicht geeignet die Rolle der Schülerinnen oder Schüler vergleichbar zu spielen.

Argumente für eine valide Interpretation der Ergebnisse auf Basis der Methode

Die Methode insgesamt ist sehr nah an realen Erklärungssituationen, wie sie im Unterricht von Lehrkräften beispielsweise bei Gruppenarbeiten bewältigt werden müssen. Dies legt eine valide Gestaltung des Testverfahrens nahe – auch wenn die Laborsituation nicht automatisch auf Erklärungen, die vor einer ganzen Klasse ausgeführt werden, übertragbar ist. Die ausgewählten Themen der Erklärung (siehe 4.1.1) entsprechen Themen der Vorlesungen zur Mechanik in den Grundkursen der Universitätsphysik und bilden Schwerpunkte in Lehrbüchern der Universitäts- und Schulphysik. Auch eine Analyse von Bildungsplänen bestätigt die curriculare Validität der inhaltlichen Anforderungen an die Probanden.

Die Gestaltung der Themen und Begleitmaterialien wurde in einer Pilotstudie überprüft. Dazu wurden zehn gerade examinierte Absolventen des Lehramts Physik und des Volfachs Physik getestet. Diese Stichprobe verspricht ein vergleichsweise hohes Fachwissen und bei den Lehramtsstudierenden zudem hohes fachdidaktisches Wissen. Wenn diese Studierenden die Themen nicht erklären können, wäre die Themenauswahl insgesamt für Studierende vermutlich nicht geeignet. Die zehn Erklärenden haben jeweils zwei der Themen erklärt, sodass insgesamt zwanzig Videos erstellt wurden. Wenn Probleme in den Erklärungen auftraten, wurden die Materialien kritisch gesichtet. Dies führte insbesondere zu einer Überarbeitung der zur Verfügung gestellten Diagramme. Beispielsweise waren manche Vektoren nicht eindeutig bezeichnet und ließen Verwechslungen zwischen Kraft- und Geschwindigkeitsvektoren zu.

Als erster Validierungsschritt wurde eine Expertenbefragung durchgeführt, die Unklarheiten in den gestellten Fragen der Adressaten aufdecken sollte. Liebold und Triszek (2009, S. 3) folgend wurden Experten befragt, die eine hohe praktische Erfahrung im Erklären aufweisen, aber nicht mit dem Modell vertraut sind (zwei promovierte Physikdidaktiker mit Lehrerfahrung, ein erfahrener Hochschullehrer der Experimentalphysik und ein besonders guter Erklärer aus der Pilotstudie). Den Experten wurden zunächst die Videos vorgespielt bis eine der standardisierten Fragen gestellt wurde. Dann wurden sie befragt, wie sie an Stelle des Erklärenden reagieren würden. Anschließend wurde ihnen die Reaktion der Erklärenden vorgespielt und sie sollten beurteilen, ob deren Reaktion angemessen war. So konnten Reaktionen der Erklärenden identifiziert werden, die aus Expertensicht problematisch sind. Später sollen diese Reaktionen besonders untersucht werden; Kategorien, die hier besonders häufig auftauchen, könnten negativ auf die Erklärungsqualität einwirken. Zudem wurde verglichen, ob verschiedene Reaktionen auf die Fragen der Adressaten als angemessen eingeschätzt werden; daraufhin wurden die standardisierten Adressatenfrage für Folgeerhebungen umformuliert.

In einem zweiten Schritt wurde in einer begleitenden Interviewstudie überprüft, wie gut die Adressaten ihre Rolle ausfüllen konnten (Felchow 2013). Dazu wurden sieben Erklärende im Anschluss an ihre Erklärungen interviewt. Ausgewählt wurden solche Erklärende, die 1) möglichst viele der standardisierten Fragen gestellt bekommen hatten und 2) abweichend von den intendierten Antworten reagiert haben. Diese Erklärenden wurden in ein bis zwei Wochen nach der Erhebung durchgeführten Interviews befragt. Dabei wurde ihnen ihr Videomaterial gezeigt und an kritischen Stellen – den Fragesituationen – gestoppt (Stimulated Recall). Sie wurden zunächst offen nach ihrer Wahrnehmung der Erklärungssituation allgemein befragt. Danach wurden sie zu jeder Frage des Adressaten gebeten, a) ihre Reaktion zu erläutern, b) anzugeben, ob sie wussten, was

das Problem des Adressaten war und c) einzuschätzen, ob sie den fachlichen Hintergrund zur Beantwortung der Frage kannten. Aus den Interviews ergab sich:

1. Alle Erklärenden gaben an, dass sie die Interviews als authentische Erklärungen wahrgenommen haben. Wenn Ihnen eröffnet wurde, dass die Adressaten trainiert waren, reagierten sie überrascht.
2. Zwei Erklärende gaben an, dass sie es schwierig fanden, das Vorwissen der Schülerinnen und Schüler zu erkunden.
3. Kein Erklärender gab an, aufgrund mangelnden fachphysikalischen Wissens nicht auf die Fragen reagiert haben zu können. Die Schwierigkeit lag einhellig darin, eine gute Darstellung der Sachverhalte zu finden.
4. Zwei der Erklärenden gaben an, dass sie die Erklärung nicht komplexer gestalten wollten, obwohl die Adressaten explizit z. B. nach Formeln gefragt hatten. Sie waren der Meinung, dass diese vorher nicht genügend Verständnis auf qualitativer Ebene gezeigt hätten, um dem dann höheren Abstraktionsgrad folgen zu können.

Bei Fragen der Adressaten, die zu komplexeren Ausführungen in den Erklärungen führen sollen, wird deshalb als Resultat Wert darauf gelegt, dass die Adressaten vorher Verstehen glaubhaft darstellen. Ansonsten kann die Interviewstudie von Felchow (2013) als Hinweis auf kognitive Validität gewertet werden – d. h. die Anforderungen für die Erklärenden lagen in der Art der Darstellung der zu erklärenden Sachverhalte, nicht im fachlich-physikalischen Bereich.

Stichprobe und Methodik

Um auf Basis der aus den Erklärungssituationen gewonnenen Videos ein valides, reliables und objektives Maß für Erklärens-fähigkeit zu gewinnen, gehen wir folgendermaßen vor:

1. Zunächst werden Videodaten einer geeigneten Stichprobe gesammelt (Abschn. 4.2.1)
2. Diese werden mit qualitativer Inhaltsanalyse analysiert. Ziel ist dabei, Kategorien zu finden, die Erklärungsqualität repräsentieren. Dabei dienen die Gütekriterien für Erklären (Abschn. 2.3) sowie das Modell des Erklärensprozesses (Abschn. 3) als Ausgangspunkt (Abschn. 4.2.2).
3. Das aus diesen Kategorien konstruierte quantitative Maß für Erklärens-fähigkeit benötigt Validitätsargumente. In diesem Beitrag liegt der Fokus auf Kriteriumsvalidität und Inhaltsvalidität. Inhaltsvalidität wird dadurch erreicht, dass alle Variablen des Modells im Maß repräsentiert sein sollen. Kriteriumsvalidität kann in a) konkurrenente und b) prognostische Validität unterschieden werden (Jenßen et al. 2015). Dabei wird die

konkurrenente Validität hier durch ein Expertenrating der Erklärens-fähigkeit eingeschätzt (Abschn. 4.2.3).

4. Am Ende wird das Maß test- bzw. auswertungsökonomischer gestaltet, ohne dass Reliabilität und Validität beeinträchtigt werden (Abschn. 4.2.3).
5. Anschließend wird der zweite Teil der Kriteriumsvalidität, die prognostische Validität des Maßes, überprüft, indem ein erster Einblick davon gewonnen wird, ob die Einschätzung der Erklärens-fähigkeit unabhängig vom Thema und trainierten Adressaten ist (Abschn. 4.2.4).

Stichprobe

Die Stichprobe der Erklärenden besteht aus Studierenden der Physik (Lehramt und Vollfach) im ersten und dritten Fachsemester sowie Absolventen kurz nach dem Masterexamen. Insgesamt können für die Forschungsfrage 64 Videos zur Entwicklung eines Kategoriensystems verwendet werden. Davon sind 10 von Erklärenden aus dem ersten Semester, 34 aus von solchen aus dem dritten und 20 von Absolventen. Zum Expertenrating bezüglich der Erklärens-fähigkeit wurden 16 Erklärende ausgewählt (Auswahlkriterien siehe 4.2.3). Von diesen 16 Erklärenden liegt jeweils ein weiteres Video vor, in dem sie einem anderen Adressaten einen zweiten Sachverhalt erklären. Es wurden sowohl Vollfach- als auch Lehramtsstudierende ausgewählt, um in der Stichprobe wahrscheinlicher auch ein hohes Fachwissen mit einem niedrigen fachdidaktischen Wissen gepaart zu integrieren. Die Stichprobe sollte möglichst heterogen sein, zumal nicht klar ist, welche Wirkung Fachwissen und welche fachdidaktisches Wissen auf Erklärens-fähigkeit hat.

Methoden zur Analyse der Erklärungsvideos

Die Analyse der Videos wurde mit qualitativer Inhaltsanalyse (Mayring 2003) durchgeführt. Dabei wurde als deduktiver Startpunkt von dem in Abschn. 3 vorgestellten Modell sowie den in Abschn. 2.3 angeführten Qualitätskriterien für Erklärungen ausgegangen. Das Modell wurde genutzt, um Kategorien bezüglich der Veränderungen, die während des Prozesses des Erklärens vorgenommen werden, zu bilden. Die Qualitätskriterien wurden bei Auftauchen in jeder Äußerung kodiert. Die Variablen des Modells und die Qualitätskriterien wurden dabei als Oberkategorien behandelt, die auf unterschiedliche Weise realisiert werden können. Dazu wurde in jeder Äußerung untersucht, auf welche Weise beispielsweise der Umgang mit graphischen Darstellungsformen verändert wurde (z. B. durch den gleichzeitigen Bezug auf ein Diagramm und ein Foto beim Erklären). Dies wurde als erste Kategorie behandelt. Im Verlaufe der Kategorienentwicklung wurden Kategorien, die nur einmalig aufgetaucht sind, zusammengefasst, in diesem Fall z. B.

zu „Verknüpfung mehrerer Darstellungsformen“. Zudem wurde in einem freien Kodieren nach Kategorien gesucht, die die Erklärungsqualität verschlechtern könnten.

Die erste Analyse führte so zu 45 Kategorien mit denen alle Reaktionen der Erklärenden auf die Adressatenfragen beschrieben werden konnten. Die Kategorien ließen sich allen vier Variablen aus dem Modell (Abschn. 3) – Graphische Darstellungsform, Mathematisierung, Beispiele/Analogien sowie sprachlicher Code (s. auch Tab. 1) – zuordnen. Zudem konnten für die Gütekriterien (Abschn. 2.3, z. B. Evaluation des Verständnisses nach einer Erklärens-passage) Kategorien gebildet werden. Die hohe Anzahl von 45 Kategorien ergab sich, weil die vier Modellvariablen auf unterschiedliche Weisen in den Erklärungssituationen konkretisiert werden konnten (z. B. für graphische Darstellungsform: Verknüpfung mehrerer Darstellungsformen, Verwendung unterschiedlicher Abstraktionsgrade von Darstellungsformen,...). Ebenso konnten die Gütekriterien auf unterschiedliche Arten in den Videos erscheinen (z. B. Evaluation des Verständnisses nach einer Erklärens-passage: verständnistestende Aufgaben, direkte Fragen nach Verstehen,...). Es wurden auch vier Kategorien gefunden, die der Erklärungsqualität vermutlich nicht zuträglich sind (z. B. herablassende Äußerungen, Ignorieren von Fragen). Über die Kodierung dieser 45 Kategorien für jede Erkläreräußerung konnte bei allen 16 Fällen, die für das Expertenrating herangezogen wurden (siehe 4.2.1 und 4.2.3), zwischen zwei Ratern ein Konsens gefunden werden. Bei Katego-

rien unter Cohens Kappa $\kappa < 0,7$ haben sich die Rater in der Kommunikation geeinigt und das Kodiermanual verbessert.

Expertenrating zur Validierung

Das übergeordnete Ziel des Vorhabens bestand darin, ein Maß für Erklärens-fähigkeit zu formulieren. Man benötigt dafür einen Maßstab, an dem die „Kalibrierung“ solch einer Auswertung der Videos valide vorgenommen werden kann. Dazu wurde ein Expertenrating durchgeführt. Wiederum wurden Liebold und Trinszek (2009, S. 3) folgend Expertinnen und Experten ausgewählt, die eine hohe praktische Erfahrung im Erklären aufweisen. Selbstverständlich kann selbst von Expertinnen und Experten nicht erwartet werden, dass sie so ein komplexes Merkmal wie Erklärens-fähigkeit direkt einstufen oder gar Ränge von Erklärern bilden können. Deshalb wurden Kontraste gebildet. Die Expertinnen und Experten sollten jeweils für ein Paar von zwei ausgewählten Videos entscheiden, welcher der beiden Erklärer besser erklärt hat (Auftrag: „Wer war in diesen Videos der bessere Erklärer bzw. die bessere Erklärerin?“). Ein analoges Verfahren des Paarvergleichs bei komplexen Merkmalen wurde bereits von Kulgemeyer (2010, S. 184) für Kommunikationskompetenz und von Schreiber (2012, S. 103) für Experimentierkompetenz angewendet. Hochinferente Expertenratings werden für komplexe Merkmale in der Literatur als angemessen charakterisiert (Langer und Schulz von Thun 2007). Man kann ihnen jedoch nur Aus-

Tab. 1 Liste der Kategorien, die in das Maß eingehen sowie deren Zugehörigkeit zu den sich aus dem Modell ergebenden Oberkategorien

| | Oberkategorie | Kategorie | Erläuterung |
|-----------------------------|--|---|---|
| Variablen einer Erklärung | Sprache Beispiele/Analogien Mathematisierung | 1. Umschreibung von Fachbegriffen 2. Unangemessenes Beispiel 3. Zahlenbeispiel | 1. Meint die direkte Erläuterung eines Fachbegriffs in Alltagssprache 2. Meint Beispiele, die ein falsches Bild der Realität vermitteln 3. Meint z. B. physikalische Größen, die durch einen Wert veranschaulicht werden |
| | Graphische Darstellungsform | 4. Verknüpfen von nonverbalen Elementen 5. Gegenständliche Hilfsmittel verwenden 6. Gegenstände mit Erklärung verknüpfen 7. Ad-hoc Experimente | 4. Meint z. B. das Aufeinanderbeziehen verschiedener Graphiken („Hier sieht man... und da sieht man...“) 5. Meint das Einbeziehen von Gegenständen (z. B. Wasserflasche) in die Erklärung zur Nachstellung von Situationen oder Bewegungen 6. Meint das Einbeziehen von Gegenständen durch Benennung, was diese repräsentieren („Das bist du...“) 7. Meint das spontane Durchführen von „Experimenten“, z. B. Werfen eines Papierknäuels zur Veranschaulichung der Wurfparabel |
| Perspektive | Adressatengemäßheit diagnostizieren | 8. Diagnostizierende Verständnisversicherung 9. Aufforderung des Adressaten zur Handlung | 8. Meint das Diagnostizieren des Erklärungserfolgs durch Wissensfragen oder Aufgaben – NICHT: „Hast du das verstanden?“ 9. Meint die Aufforderung, eine Handlung wie z. B. zeichnen zu vollführen |
| Verknüpfung von Erklärungen | Strukturelle Elemente | 10. Strukturierender Rückblick 11. Zusammenfassung/ Wiederholung 12. Ermunterung und Lob | 10. Meint, dass der Erklärer explizit darauf verweist, dass etwas bereits erklärt wurde und jetzt eine Rolle spielt („Du hast ja gerade von der Zentrifugalkraft gesprochen...“) 11. Meint das Zusammenfassen eines bereits erklärten Sachverhalts 12. Meint Lob bei erfolgreichem Handeln des Schülers oder Verständnis für etwaige Schwierigkeiten |

sagekraft zumessen, wenn die Experten mit hoher Interraterreliabilität entscheiden.

Um zu belastbaren Ergebnissen zu gelangen, wurden deutlich mehr als die oft üblichen zwei Rater pro Paarvergleich angesetzt. Für die Expertenbefragung wurden 16 der Videos ausgewählt und je zwei Videos zu einem Thema gepaart. Die Paare wurden so konfiguriert, dass die Erklärungen in ähnlichem Maße physikalisch korrekt (oder falsch) waren und sowohl korrekte als auch falsche Paarungen vorkamen. Zudem wurden die Experten explizit darauf hingewiesen, nicht auf Basis fachlicher Korrektheit zu entscheiden, sondern über die Adressatengemäßheit der Erklärung zu urteilen. Wenn mehrere Videos für die Paarbildung in Frage kamen, wurden Paare von Videos gebildet, die sich nach der ersten Analyse (Abschn. 4.2.2) in ihrer Qualität möglichst unterscheiden. Dies wurde vorgenommen, um den Kontrast zu verstärken und die hochinferenten Urteile zu erleichtern. Da in solche Urteile vermutlich vielerlei Abwägungen eingehen, soll so die Entscheidung auf Basis der Erklärungsqualität stärker zutage treten. Über jedes der acht Paare wurden fünf Expertenurteile eingeholt. Für diese acht Paare wurde die Einigkeit mit Fleiss' Kappa (Wirtz und Caspar 2002) berechnet. Dazu wurde das Expertenurteil dichotom kodiert (1: Video A besser, 0: Video B besser). Dabei ergab sich eine sehr gute Übereinstimmung der Expertinnen und Experten (90%, $\kappa=0,798$).

Methoden zur Gewinnung eines Maßes für Erklärens-fähigkeit und zur Analyse der Validität

Mit den 45 Kategorien konnten die Erklärungen qualitativ gut beschrieben werden. Nun wurden sieben Schritte unternommen, um daraus ein quantitatives Maß zu entwickeln bzw. Argumente für die Validität der Interpretation dieses Maßes als Maß für Erklärens-fähigkeit zu gewinnen:

1. Der erste Schritt ging von der Annahme aus, dass eine höhere Erklärens-fähigkeit damit einhergeht, dass der Erklärende mehr Kategorien guten Erklärens verwendet und weniger Kategorien, die der Erklärungsqualität vermutlich nicht zuträglich sind. Wie bei Kulgemeyer und Schecker (2013) wurden für jede Erklärungssituation das Auftreten der vermutlich für die Erklärungsqualität positiven Kategorien summiert (hier: X). Ebenso wurde die Summe des Auftretens der negativ gepolten Kategorien gebildet (hier: Y). Die Berechnungsvorschrift für den „Erklärens-fähigkeitsindex“ $EFI=X-Y+4$ (es gibt vier negative Kategorien) führt stets zu einem Wert größer null. Durch die Verschiebung des Maßes in den positiven Bereich soll die Skala anschaulicher werden. Alle Kategorien gehen gleichgewichtet in das Maß ein, da für Gewichtungsfaktoren keine Grundlage vorliegt.

2. Im nächsten Schritt wurden die Bedingungen festgelegt, die zu einer validen Interpretation des Maßes erfüllt sein sollen. Als Validitätskriterium wurde gewählt, dass das Maß die Expertenentscheidung für den besseren Erklärer in den in Abschn. 4.2.3 beschriebenen Paaren vorhersagen kann (kriteriale Validität).
3. Als statistisches Maß für die Übereinstimmung zwischen der Expertenentscheidung und dem höheren Score im Maß wurde Cohens Kappa gewählt. Dabei soll $\kappa>0,7$ (eine sehr gute Übereinstimmung nach Wirtz und Caspar 2002) angestrebt werden.
4. Zudem sollte das Maß reliabel sein, in diesem Fall wurde ein Maß angestrebt, das die übliche Bedingung Cronbachs $\alpha>0,7$ erfüllt.
5. Alle Modellvariablen (s. Abschn. 3) und möglichst viele Gütekriterien (s. Abschn. 2.3) sollen durch mindestens eine Kategorie repräsentiert werden (Inhaltsvalidität).
6. Bei der Konstruktion des Maßes wird nun ab Punkt 3 schrittweise vorgegangen. Es wird ständig die Übereinstimmung des EFI mit der Expertenentscheidung berechnet. Gleichzeitig wird Cronbachs Alpha sowie die Item-Skalen-Korrelation zwischen dem Auftreten der Kategorien und der EFI-Skala berechnet. Schrittweise werden die Kategorien aus dem Maß entfernt, die eine negative Item-Skalen-Korrelation aufweisen. Wenn dies zu einer Verschlechterung der Übereinstimmung mit dem Expertenentscheid führt, werden sie wieder hinzugefügt. Es werden also nur solche Kategorien aus dem Maß entfernt, die sowohl die Validität als auch die Reliabilität des Maßes negativ beeinflussen.
7. Im letzten Schritt wird die Stabilität dieser Fähigkeits-einschätzung überprüft. Alle 16 Erklärenden, deren Videos von den Experten eingeschätzt wurden, haben noch ein zweites Mal direkt im Anschluss an ihre erste Erklärung einen anderen Sachverhalt einem anderen Adressaten erklärt. Für diesen zweiten Testzeitpunkt wurde ebenfalls der EFI berechnet. Wenn die Fähigkeitseinschätzung stabil ist, sollten sich beim zweiten Testzeitpunkt wieder dieselben Urteile über die bessere Erklärung wie beim ersten Testzeitpunkt ergeben. Dies ist eine Frage prognostischer Validität: kann auf Basis des höheren Maßes die Performanz in einem neuen Thema bei einem neuen Adressaten der Erklärung vorhergesagt werden?

Entwicklung eines kategorienbasierten Maßes zur Quantifizierung von Erklärens-fähigkeit anhand eines Videos

Nach dem in 4.2.3 beschriebenen sukzessiven Verfahren konnte die Anzahl der notwendigen Kategorien zur Berechnung des Erklärens-fähigkeitsindex (EFI) sukzessive von

45 auf 12 reduziert werden. Diese 12 Kategorien bilden offenbar den Kern des ursprünglichen Kategorien-Satzes. Tabelle 1 stellt diese Kategorien dar. Es zeigt sich, dass alle Variablen des Modells durch Kategorien vertreten sind. Die Kategorie „Unangemessenes Beispiel“ gilt dabei als negativ gepolte Kategorie. Gemeint ist damit, dass das gewählte Beispiel des Erklärers nicht zu der zu erklärenden Situation passt. Das resultierende Maß für Erklärens-fähigkeit spannt demzufolge nach der Berechnungsvorschrift $EFI = X - Y + 1$ von 0 bis 12 und ist wiederum aus Gründen der Anschaulichkeit in den positiven Bereich verschoben.

In Tab. 2 findet sich eine Gegenüberstellung des EFI für die Erklärenden mit den Expertenurteilen.

Für Videopaar 1 heißt das, dass die fünf Expertenentscheidungen einhellig für Erklärer A als besseren Erklärende ausgefallen sind. Nach EFI erreicht Erklärer A 9 Punkte und Erklärer B 2. Die Übereinstimmung zwischen dem Expertenentscheid und dem EFI liegt insgesamt bei

Tab. 2 Vergleich der Ergebnisse des Expertenentscheids über den besseren Erklärer in einem Paarvergleich (A oder B) mit den Ergebnissen der Einstufung der Erklärens-fähigkeit durch das konstruierte Maß. Zweite Zeile: Fleiss' Kappa für die Übereinstimmung der Experten, Cronbachs Alpha für das Maß. Letzte Zeile: Cohens Kappa für die Übereinstimmung des Expertenentscheids und des Urteils auf Basis des Maßes

| Videopaar | Expertenentscheid | | EFI | |
|-----------|-------------------|---|----------------|---|
| | $\kappa=0,798$ | | $\alpha=0,772$ | |
| | A | B | A | B |
| 1 | 5 | 0 | 9 | 2 |
| 2 | 0 | 5 | 0 | 3 |
| 3 | 5 | 0 | 7 | 3 |
| 4 | 4 | 1 | 9 | 6 |
| 5 | 4 | 1 | 2 | 3 |
| 6 | 0 | 5 | 6 | 8 |
| 7 | 0 | 5 | 4 | 7 |
| 8 | 0 | 5 | 2 | 7 |
| | $\kappa=0,750$ | | | |

Tab. 3 Vergleich der Ergebnisse von Messzeitpunkt 1 und Messzeitpunkt 2 (die Messzeitpunkte unterschieden sich darin, dass dieselbe Person jeweils ein anderes Thema einem anderen Adressaten erklärt). Angegeben ist jeweils der Erklärens-fähigkeitsindex (EFI) für die Videopaare sowie Cohens Kappa für die Übereinstimmung des Urteils über die bessere Erklärung (grau hervorgehoben: Abweichungen)

| Videopaar | EFI (Z1) | | EFI (Z2) | |
|-----------|----------------|---|----------|---|
| | A | B | A | B |
| 1 | 10 | 3 | 6 | 4 |
| 2 | 1 | 4 | 4 | 4 |
| 3 | 8 | 4 | 7 | 3 |
| 4 | 10 | 7 | 6 | 4 |
| 5 | 3 | 4 | 3 | 6 |
| 6 | 7 | 9 | 2 | 4 |
| 7 | 5 | 8 | 6 | 9 |
| 8 | 3 | 8 | 4 | 4 |
| | $\kappa=0,600$ | | | |

$\kappa=0,750$ (87,5% Übereinstimmung). Der EFI erreicht zudem $\alpha=0,772$.

Hervorzuheben ist das Videopaar 5. Hier votieren die Experten mit 4 zu 1 für den Erklärenden A als den besseren Erklärer. Der EFI ergibt mit 3 zu 2 Punkten, dass der Erklärende B die bessere Leistung gezeigt hat. Hier ist das Maß also nicht in der Lage, den Expertenentscheid zu prognostizieren. Es ist allerdings festzustellen, dass die Experten bei gerade diesem Video uneiniger waren als bei den anderen (bis auf Paar 4).

In Tab. 3 finden sich die Ergebnisse für den Vergleich des EFIs von Messzeitpunkt 1 und Messzeitpunkt 2.

Es zeigt sich, dass trotz eines neuen Themas und eines anderen Adressaten die Urteile über die bessere Erklärung nicht wesentlich anders ausfallen. Die Übereinstimmung des Urteils über den besseren Erklärenden liegt bei $\kappa=0,600$ (75% Übereinstimmung).

Diskussion und Ausblick

In der hier beschriebenen Studie wurde mit einem Modell für Unterrichtserklärungen ein Maß (der EFI) entwickelt, mit dem die Fähigkeit angehender Physiklehrkräfte, adressatengemäß zu erklären, quantifiziert werden kann. Das Maß bezieht sich dabei auf die kategorienbasierte Auswertung des Handelns in unterrichtsnahen Erklärsituationen.

Diskussion der Reliabilität und Validität des EFI

Der EFI ist reliabel in dem Sinne, dass er ein Cronbachs Alpha von $\alpha=0,772$ erreicht. Es zeigt sich zudem, dass die in Abschn. 4.2.4 gestellten Anforderungen an die Validität des EFI erfüllt werden konnten. Er erlaubt eine valide Interpretation seiner Ausprägung hinsichtlich der Erklärens-fähigkeit in dem Sinne, dass er die Entscheidung von Experten über die bessere von zwei Erklärungen mit einer Übereinstimmung von 90% ($\kappa=0,750$) vorhersagt. Dies kann nach Wirtz und Caspar (2002, S. 59) als eine sehr gute Übereinstimmung interpretiert werden. Zu diskutieren ist dabei die Abweichung zwischen dem Expertenurteil und dem EFI in Videopaar 5 (Tab. 2). Hier entscheidet der EFI mit 2 zu 3 Punkten für Video B während die Experten Video A mit 4 zu 1 Stimmen als besseres beurteilen. Der sehr geringe Abstand zwischen den Videos im EFI könnte dazu führen, dass das Expertenurteil schwieriger ist als bei den anderen Paaren, bei denen die Experten (Ausnahme: Paar 4) immer einhellig entscheiden. Auf Basis der Kategorien, aus denen sich die EFIs bei diesem Paar zusammensetzt, lässt sich keine besondere Kombination ausmachen, die sich von den anderen Paaren unterscheidet. Insgesamt sind beide Erklärende jedoch als eher schlecht beurteilt worden. Möglicher-

weise ist ein Expertenurteil gerade bei der Unterscheidung zwischen zwei schlechten Erklärenden schwieriger.

Für eine valide Interpretation des EFI spricht neben dem Expertenrating, dass alle Variablen des Modells durch Kategorien repräsentiert werden und die aus der Literatur bekannten Gütekriterien für Erklärungen zudem in der Breite dargestellt werden (Inhaltsvalidität). Ein Vergleich der Tab. 1 mit den Gütekriterien aus Abschn. 2.3 zeigt, dass die Punkte „Gebrauch von Darstellungsformen“, „Einbezug des Adressaten“, „Evaluation des Verständnisses“, „Beispiele und Analogien“ direkt im EFI vertreten sind und die „Betonung besonders relevanter Punkte“ sowie der „logische Bezug“ durch die strukturellen Elemente repräsentiert werden.

Auch prognostisch valide Einschätzungen scheinen möglich. Das Urteil über den besseren Erklärenden mithilfe des EFIs ist weitgehend themenunabhängig sowie unabhängig vom Adressaten der Erklärung. Dazu wurden die 16 Erklärenden, deren Videos im Expertenrating untersucht wurden (Messzeitpunkt 1), noch ein weiteres Mal beim Erklären eines anderen Themas mit einem anderen Adressaten gefilmt (Messzeitpunkt 2). Die Übereinstimmung des Urteils über den besseren Erklärer eines Paares auf Basis des EFI liegt zwischen Messzeitpunkt 1 und Messzeitpunkt 2 bei $\kappa=0,600$ (75%). Das wird hier als gute Übereinstimmung gewertet (vgl. Wirtz und Caspar 2002, S. 59), da das zu erfassende Merkmal komplex und schwierig zu erfassen ist. Zudem ergab sich in Messzeitpunkt 2 keine gegenteilige Einschätzung zu Messzeitpunkt 1. Wenn man davon ausgeht, dass sich Erklärens-fähigkeit themenunabhängig sowie unabhängig vom Adressaten in der Performanz manifestiert, ist dies das zu erwartende Ergebnis – aus der Performanz in Messzeitpunkt 1 lässt sich prognostisch valide die Performanz in Messzeitpunkt 2 vorhersagen.

Beim Verfahren der Skalierung wurden Reliabilität und Validität in einem sukzessiven Prozess optimiert, um von 45 Kategorien zu 12 notwendigen zu gelangen. Alternativ hätten Faktorenanalysen zur Dimensionsfindung auf alle 45 Kategorien angewendet werden können, ebenso wären mehrdimensionale Rasch-Analysen eine Alternative gewesen. Bei beiden Verfahren wäre allerdings der Vergleich mit der Experteneinschätzung nicht direkt eingegangen – zudem wären deutlich mehr Erklärende notwendig gewesen.

Dem Maß kann also eine gute theoretische Einbettung sowie Validität und Reliabilität in den hier betrachteten Facetten unterstellt werden. Es geht durch seine eher niedrigeren Einstufungen deutlich über die bei Kulgemeyer und Schecker (2013) oder Sevian und Gonsalves (2008) beschriebenen Punkte hinaus. Dennoch ist die vorliegende Evidenz nur als vorläufiges Ergebnis zu werten. Die letztlich im Hinblick auf Erklärungen im Unterricht entscheidende prognostische Validierung – der Vergleich des Erklärens-fähigkeitsmaßes mit einem durch Pre- und Post-Test

diagnostizierten Fachwissenszuwachses der Adressaten – konnte hier nicht durchgeführt werden, da die Adressaten wegen der Notwendigkeit einer Standardisierung der Erhebungsumgebung zum Zwecke der Testentwicklung vorher geschult werden mussten. Das Maß beschreibt zudem nur adressatengemäßes Erklären. Die sachgerechte Komponenten ist bislang nicht berücksichtigt: die fachliche Korrektheit der Erklärung oder die Struktur der Erklärung (z. B. unter deduktiv-nomologischen Gesichtspunkten) benötigen ein getrenntes Maß, das Gegenstand laufender Arbeiten ist.

Ausblick

Zusätzlich zu den in diesem Beitrag beschriebenen 64 Videos sind weitere 134 Erklärungsvideos von Studierenden aller Semester aufgenommen worden. Die Studierenden sind außerdem zum fachdidaktischen Wissen und zum Fachwissen getestet worden (Projekt ProfiLe-P s. Riese et al. 2015). Über diese Zusammenhänge sind weitere Erkenntnisse über die Konstruktvalidität zu erwarten. Außerdem soll der Einfluss professionellen Wissens auf die Performanz beim Erklären analysiert werden.

Mit dem formulierten EFI ist die Grundlage für weitere Analysen geschaffen. Die bei Geelan (2012) geforderten Studien über Lehrerklärungen könnten auf Basis dieses Maßes oder auf Basis der Kategorien durchgeführt werden. Beide müssen jedoch für Erklärungen vor Gruppen möglicherweise noch adaptiert werden. So könnte beispielsweise die Frage, wie eine Lehrkraft allen Schülerinnen und Schülern gleichermaßen die Aufmerksamkeit widmet oder welche Verständnisprobleme zum Anlass der Erklärungsvariation genutzt werden, Berücksichtigung finden. Diagnostik von Verständnisschwierigkeiten in einer Gruppe verläuft sicherlich anders. Möglicherweise könnte die Analyse der Qualität von Lehrerklärungen nach solchen Kriterien auch weitere Anhaltspunkte geben, um die Schülerperformance in Leistungstests zu erklären.

Die vorliegenden Kategorien zur Einstufung der Erklärens-fähigkeit können zudem als relevantes Ergebnis für die fachdidaktische Ausbildung von Physiklehrkräften gewertet werden: von der Deutschen Physikalischen Gesellschaft wird beispielsweise in den Empfehlungen für das Lehramtsstudium ein Modul „Physik erklären“ beschrieben (Großmann und Hertel 2014), in dem diese Kategorien thematisiert werden könnten.

Generell sollte die Frage nach der Rolle von Lehrerklärungen in einem modernen Naturwissenschaftsunterricht diskutiert werden. Wir sind der Auffassung, dass Erklären mit konstruktivistisch ausgerichtetem Unterricht vereinbar ist, solange darunter eine kommunikative Interaktion von Lehrenden und Lernenden verstanden wird – aber Geelans (2012) Auffassung, dass ein großer Mangel an Studien in diesem Bereich zu konstatieren ist, kann nur unterstrichen werden.

Literatur

- Aufschnaiter, C. v., & Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften*, 16, 361–367.
- Bartelborth, T. (2007). *Erklären*. Berlin: De Gruyter.
- Berland, L. K., & McNeill, K. L. (2012). For whom is argument and explanation a necessary distinction? A response to Osborne and Patterson. *Science Education*, 96(5), 808–813.
- Brewer, W., Chinn, C., & Samarapungavan, A. (2000). Explanation in scientists and children. In F. Keil & R. Wilson (Hrsg.), *Explanation and cognition* (S. 279–298). Cambridge: The MIT Press.
- Brown, G. (2006). Explaining. In O. Hargie (Hrsg.), *The handbook of communication skills* (S. 195–228). East Sussex: Taylor & Francis.
- Clark, H., & Brennan, S. (1991). Grounding in communication. In L. Resnick, J. Levine, & S. Teasley (Hrsg.), *Perspectives on socially shared cognition* (S. 127–149). Washington: American Psychological Association.
- Felchow, S. (2013). *Validierung einer qualitativen Methode zur Diagnostik von Erklärungswissen bei Physikstudierenden*. Unveröffentlichte Masterarbeit Universität Bremen. (Auf Anfrage erhältlich).
- Fleischhauer, J. (2013). *Wissenschaftliches Argumentieren und Entwicklung von Konzepten beim Lernen von Physik*. Berlin: Logos.
- Gage, N. (1968). The microcriterion of effectiveness in explaining. In N. Gage (Hrsg.), *Explorations of the teacher's effectiveness in explaining*, Technical Report No. 4 (S. 1–8). Stanford Center for Research and Development in Teaching.
- Geelan, D. (2012). Teacher explanations. In B. Fraser, K. Tobin, & C. McRobbie (Hrsg.), *Second international handbook of science education* (S. 987–999). Dordrecht: Springer.
- Großmann, S., & Hertel, I. (Hrsg.). (2014). *Zur fachlichen und fachdidaktischen Ausbildung für das Lehramt Physik*. Bad Honnef: DPG.
- Hempel, C., & Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science*, 15(2), 135–175.
- Jenßen, L., Dunekacke, S., & Blömeke, S. (2015). Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis. In Blömeke, S. & Zlatkin-Troitschanskaia, O. (Hrsg.), *Kompetenzen von Studierenden: 61. Beiheft der Zeitschrift für Pädagogik* (S. 11–31). Weinheim: Beltz.
- Kiel, E. (1999). *Erklären als didaktisches Handeln*. Würzburg: Ergon.
- Kitcher, P. (1981). Explanatory unification. *Philosophy of Science*, 48(4), 507–531.
- Kulgemeyer, C. (2010). *Physikalische Kommunikationskompetenz. Modellierung und Diagnostik*. Berlin: Logos.
- Kulgemeyer, C. (2013). Gelingensbedingungen physikalischer Erklärungen. Zu einer konstruktivistischen Auffassung des Erklärens. *PhyDid B – Didaktik der Physik*.
- Kulgemeyer, C., & Schecker, H. (2009a). Kommunikationskompetenz in der Physik: Zur Entwicklung eines domänenspezifischen Kompetenzbegriffs. *Zeitschrift für Didaktik der Naturwissenschaften*, 15, 131–153.
- Kulgemeyer, C., & Schecker, H. (2009b). Physikalische Darstellungsformen. Ein Beitrag zur Klärung von Kommunikationskompetenz. *Der mathematische und naturwissenschaftliche Unterricht* 62(6), 331–336.
- Kulgemeyer, C., & Schecker, H. (2012). Physikalische Kommunikationskompetenz – Empirische Validierung eines normativen Modells. *Zeitschrift für Didaktik der Naturwissenschaften*, 18, 29–54.
- Kulgemeyer, C., & Schecker, H. (2013). Students explaining science – assessment of science communication competence. *Research in Science Education*, 43, 2235–2256.
- Kulgemeyer, C., & Starauschek, E. (2014). Analyse der Verständlichkeit naturwissenschaftlicher Fachtexte. In D. Krüger, I. Parchmann, & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 241–253). Heidelberg: Springer.
- Kulgemeyer, C., Borowski, A., Fischer, H., Gramzow, Y., Reinhold, P., Riese, J., Schecker, H., Tomczaszyn, E., & Walzer, M. (2012). ProfiLe-P – Professionswissen in der Lehramtsausbildung Physik. *PhyDid B – Didaktik der Physik – Beiträge zur DPG-Frühjahrstagung*.
- Krey, O. (2012). *Zur Rolle der Mathematik in der Physik. Wissenschaftstheoretische Aspekte und Vorstellungen Physiklernender*. Berlin: Logos.
- Langer, I., & Schulz von Thun, F. (2007). *Messung komplexer Merkmale in Psychologie und Pädagogik*. Münster: Waxmann.
- Liebold, R., & Trinczek, R. (2009). Experteninterviews. In: S. Kühl, P. Strotholz, & A. Taffertshofer (Hrsg.), *Handbuch Methoden der Organisationsforschung: quantitative und qualitative Methoden* (S. 32–56). Wiesbaden: VS Verlag für Sozialwissenschaften.
- Mayring, P. (2003). *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.
- Merzyn, G. (2005). Junge Lehrer im Referendariat. *Der mathematische und naturwissenschaftliche Unterricht*, 58(1), 4–7.
- Miller, G. E. (1990). The assessment of clinical skills/competence/performance. *Academic medicine. Journal of the Association of American Medical Colleges*, 65, 63–67.
- Nagel, E. (1961). *The structure of science: Problems in the logic of scientific explanation*. New York: Harcourt, Brace & World.
- Norris, S. P., Guilbert, S. M., Smith, M. L., Hakimelahi, S., & Phillips, L. M. (2005). A theoretical framework for narrative explanation in science. *Science Education*, 89(4), 535–563.
- Ogborn, J., Kress, G., Martins, I., & McGillicuddy, K. (1996). *Explaining science in the classroom*. Buckingham: Open University Press.
- Ohle, A., Fischer, H. E., & Kauertz, A. (2011). Der Einfluss des physikalischen Fachwissens von Primarstufenlehrkräften auf Unterrichtsgestaltung und Schülerleistung. *Zeitschrift für Didaktik der Naturwissenschaften*, 17, 357–390.
- Olszewski, J. (2010). *The impact of physics teachers' pedagogical content knowledge on teacher action and student outcomes*. Berlin: Logos.
- Osborne, J. F., & Patterson, A. (2011). Scientific argument and explanation: A necessary distinction? *Science Education*, 95(4), 627–638.
- Rescher, N. (1962). The stochastic revolution and the nature of scientific explanation. *Synthese*, 14, 200–215.
- Riese, J., Kulgemeyer, C., Borowski, A., Fischer, H., Gigl, F., Gramzow, Y., Schecker, H., Tomczaszyn, E., & Zander, S. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. In Blömeke, S. & Zlatkin-Troitschanskaia, O. (Hrsg.), *Kompetenzen von Studierenden: 61. Beiheft der Zeitschrift für Pädagogik* (S. 55–79). Weinheim: Beltz.
- Rusch, G. (1999). Eine Kommunikationstheorie für kognitive Systeme. In G. Rusch & S. Schmidt (Hrsg.), *Konstruktivismus in der Medien- und Kommunikationswissenschaft* (S. 150–184). Frankfurt a. M.: Suhrkamp.
- Schreiber, N. (2012). *Diagnostik experimenteller Kompetenz*. Berlin: Logos.
- Sevian, H., & Gonsalves, L. (2008). Analysing how scientists explain their research: A rubric for measuring the effectiveness of scientific explanations. *International Journal of Science Education*, 30(11), 1441–1467.
- Shulman, L. (1987). Knowledge and teaching: Foundations of the new reform. *Harvard Education Review*, 57(1), 1–22.
- Sumfleth, E., & Gnoyke, A. (1995). Die Bedeutung bildlicher Symbolsysteme für Theoriebildungen in der Chemie. *Der mathematische und naturwissenschaftliche Unterricht*, 48(1), 14–21.
- Tomas, L., & Ritchie, S. (2014). The challenge of evaluating students' scientific literacy in a writing-to-learn construct. *Research in Science Education*. doi:10.1007/s11165-014-9412-3.

- Toulmin, S. (1958). *The uses of argument*. Cambridge: Cambridge University Press.
- Treagust, D., & Harrison, A. (1999). The genesis of effective science explanations for the classroom. In J. Loughran (Hrsg.), *Researching teaching: Methodologies and practices for understanding pedagogy* (S. 28–43). Abingdon: Routledge.
- Vogelsang, C. (2014). *Validierung eines Instruments zur Erfassung der professionellen Handlungskompetenz von (angehenden) Physiklehrkräften – Zusammenhangsanalysen zwischen Lehrerkompetenz und Lehrerperformanz*. Berlin: Logos.
- Vogelsang, C., & Reinhold, P. (2013). Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften*, 19, 129–157.
- Wellenreuther, M. (2005). *Lehren und Lernen – aber wie? Empirisch-experimentelle Forschung zum Lehren und Lernen im Unterricht*. Hohengehren: Schneider.
- Wilson, H., & Mant, J. (2011a). What makes an exemplary teacher of science? The pupils' perspective. *School Science Review*, 93(342), 121–125.
- Wilson, H., & Mant, J. (2011b). What makes an exemplary teacher of science? The teachers' perspective. *School Science Review*, 93(343), 115–119.
- Wirtz, M., & Caspar, F. (2002). *Beurteilübereinstimmung und Beurteilerreliabilität*. Göttingen: Hogrefe.