

## Entwicklung eines interaktiven computerbasierten Tests für Erklärfähigkeit

Hauke Bartels, Christoph Kulgemeyer

Institut für Didaktik der Naturwissenschaften, Abt. Physikdidaktik,  
Universität Bremen FB 1, Postfach 330440, 28359 Bremen  
hauke.bartels@uni-bremen.de, kulgemeyer@physik.uni-bremen.de

### Kurzfassung

Das Erklären physikalischer Sachverhalte wird als eine wichtige Fähigkeit von Physiklehrkräften betrachtet. Gleichzeitig zeigen Studien, dass angehende Lehrkräfte das Erklären solcher Sachverhalte als anspruchsvoll wahrnehmen. Vor diesem Hintergrund wurde von [1] bereits ein Modell entwickelt, das den Prozess des Erklärens in einem Dialog zwischen einem Experten (Erklärer / Lehrkraft) und einem Novizen (Adressat / Schüler) beschreibt. Dazu wurde zudem ein Test entwickelt, der die Erklärfähigkeit von Probanden misst [2]. Der Test beruht auf Videoaufzeichnungen des Erklärens von Probandinnen und Probanden. Die aus diesem Test gewonnenen Erkenntnisse können als besonders valide angesehen werden; der Test ist jedoch in Durchführung und Auswertung aufwändig. Im hier vorgestellten Projekt soll ein computerbasierter Kompetenztest entwickelt werden, bei dem die Erklärfähigkeit von Physiklehrkräften in großen Stichproben erhoben werden kann. Der Test basiert auf Videovignetten von Erklärersituationen, nach deren Ende die Probanden im *two tier multiple choice*-Design aufgefordert werden, das weitere, optimale Vorgehen bei der Erklärung auszuwählen und anschließend zu begründen. Mit diesem Test soll der Zusammenhang der Erklärfähigkeit von Lehrkräften und der Lernwirkung bei Schülerinnen und Schülern exploriert werden. Wir erläutern grundlegende Überlegungen bei der Testkonstruktion und stellen eine erste Version des Instruments vor.

### 1. Einleitung

Werden Studienreferendare des Lehramts Physik zu ihren größten Herausforderungen in der Praxis befragt, nennen sie häufig den als schwierig empfundenen Prozess des Erklärens physikalischer Inhalte [3]. Gleichzeitig spricht die Studienlage dafür, dass Lehrkräfte, die gut erklären können, von Schülerinnen und Schülern als besonders gut beurteilt werden [4].

Nach [5] ist der Wissenszuwachs, den Schülerinnen und Schüler im naturwissenschaftlichen Unterricht durch die ihnen angebotenen Erklärungen erhalten, ein wichtiges Maß für Unterrichtsqualität. Das Anbieten guter Erklärungen im Unterricht ist relevant für den Lernerfolg der Schülerinnen und Schüler.

Zum Erklären von Lehrkräften liegen nur wenige naturwissenschaftsdidaktische Studien vor. Dies könnte daran liegen, dass Erklärungen durch Lehrkräfte häufig mit frontalem, lehrerzentriertem oder sogar dozierendem Unterricht gleichgesetzt werden [6].

Davon abgrenzend vertreten wir ein mit konstruktivistischen Grundannahmen vereinbares Verständnis von Lehrererklärung im Unterricht, wie es in [7] und [2] beschrieben wird. Der Kern liegt dabei in der Interaktion zwischen Erklärern und Adressaten, im Zuge des Erklärens muss ständig das Verstehen evaluiert werden um daraus Schlüsse für das weitere Vorgehen zu ziehen. Im vorliegenden Beitrag werden

zunächst theoriebasierte Abgrenzungen unterschiedlicher Auffassungen des Erklärbegriffs getroffen, sowie ein empirisch abgesichertes Modell zum Erklären aufgegriffen und bisherige Ansätze zum Messen von Erklärfähigkeit vorgestellt. Anschließend wird auf die Möglichkeiten computerbasierter Tests zum Erklären eingegangen und der aktuell verfolgte Ansatz erläutert.

### 2. Erklären im Kontext Physikunterricht

#### 2.1. Abgrenzungen

Im Folgenden sollen zunächst zwei Abgrenzungen getroffen werden, die eine Einengung des Erklärbegriffs für den Kontext des Physikunterrichts ermöglichen.

#### *Erklären vs. Argumentieren*

In ihrer Bedeutung verwandt und mancherorts überlappend sind die Konzepte *Argument* und *Erklärung* [5]. [8] nennen als wesentliche Bestandteile einer Erklärung das *Explanandum*, (das Phänomen, das es zu erklären gilt) und das *Explanans* (das allgemeingültige Gesetz, welches dem Phänomen zugrunde liegt). Die Erklärung besteht nun darin, unter Nennung der nötigen Randbedingungen eine logische Verknüpfung zwischen *Explanandum* und *Explanans* herzu-

stellen. [5] folgen diesem Verständnis und grenzen das Konzept des Arguments davon ab. Statt einen kausalen Zusammenhang zwischen *Explanandum* und *Explanans* darzulegen, habe das Argument die Intention, eine Behauptung durch Präsentation geeigneter Daten zu stützen.

### Wissenschaftliche Erklärungen vs. Unterrichtserklärungen

Auf [9] geht die Unterscheidung von *science explanation* und *science teaching explanation* zurück. Sie wurde von [10] aufgegriffen und bildet die Basis für die ins Deutsche übertragene Unterscheidung zwischen *wissenschaftlicher Erklärung* und *Unterrichtserklärung* nach [2].

Die *Wissenschaftliche Erklärung* entspricht im Wesentlichen den obigen Ausführungen, das Modell von [8] ist das für die Physik wohl bedeutendste Modell wissenschaftlicher Erklärungen. Die für diesen Beitrag relevante *Unterrichtserklärung* hat gegenüber dem Argumentieren und der wissenschaftlichen Erklärung das Alleinstellungsmerkmal einer Vermittlungsabsicht. Sie beruht auf einem Wissensgefälle zwischen einem Erklärer (z. B. einer Lehrkraft) und einem Novizen (z. B. einem Schüler), der Adressat der Erklärung ist. Um das Wissensgefälle zu überbrücken, muss der Experte Überlegungen hinsichtlich geeigneter Kommunikation anstellen.

### 2.2. Ein Modell des Erklärens im naturwissenschaftlichen Unterricht

Wie eingangs erwähnt, wird die „Lehrererklärung“ häufig mit dozierendem Unterricht gleichgesetzt. Tatsächlich gibt es aber eine Vielzahl von Situationen, in denen eine Lehrkraft auch in stark schülerzentrierten Lernumgebungen Erklärungen anbieten muss. Das können z.B. sein: das „Besprechen von Lösungsstrategien“, die „Anleitung zu Schülerexperimenten“ oder das „Strukturieren und Zusammenfassen von Wissen“ [7].

In diesem Beitrag wird Erklären als ein konstruktivistischer Prozess im Rahmen des Kommunikationsmodells von [1,11] verstanden, das empirisch gut abgesichert wurde und davon ausgeht, dass Erklären ein wesentlicher Bestandteil von Kommunikationskompetenz ist. Nach diesem Verständnis unterbreitet der Erklärer dem Adressaten ein Informationsangebot. Der Adressat kann dieses Angebot nun annehmen oder ablehnen. Ein direkter Wissenstransfer ist nicht möglich. In jeden Fall erhält der Erklärer eine Rückmeldung, die er dazu nutzen kann, seine Erklärung an die Bedürfnisse des Adressaten anzupassen, um so die Wahrscheinlichkeit zu erhöhen, dass der Adressat aus der Erklärung eine Bedeutung konstruieren kann. Dies deckt sich auch mit dem Konzept des *common ground* nach [12], welches das Herstellen eines gemeinsamen Referenzrahmens für die Teilnehmer einer Kommunikationssituation fordert: Nur wenn Erklärer und Adressat eine gemeinsame Basis für ihre

Kommunikation finden können, kann eine Erklärung erfolgreich verlaufen.

### Sachgerechtheit und Adressatengemäßheit

Dieser Ansatz wurde nach empirischen Absicherungen von [1] durch [2] weiterentwickelt und führte zum Modell des dialogischen Erklärens, das in Abbildung 1 dargestellt ist. In diesem Modell muss der Erklärer beim Erstellen einer Erklärung zwei Perspektiven berücksichtigen.

Die erste Perspektive ist die *Sachgerechtheit* und bezieht sich auf die fachliche Korrektheit und Vollständigkeit der Erklärung. Die zweite Ebene ist die *Adressatengemäßheit*. Sie beinhaltet die Berücksichtigung von Vorwissen und Interesse des Adressaten. Sachgerechtheit und Adressatengemäßheit können durch vier „Variablen“ beeinflusst werden. Auf Basis der Rückmeldungen des Adressaten müssen diese Variablen im Zuge eines Erklärprozesses angepasst und verändert werden. Diese sind:

- **Sprachlicher Code** (z. B. Fachsprache vs. Alltagssprache)
- **Mathematisierungsgrad** (z. B. Formel vs. Zahlenbeispiel vs. qualitativer Zusammenhang)
- **Beispiele und Analogien** (z.B. fachnahe Beispiele wie „schiefe Ebene“ vs. alltagsnahe Beispiele wie „Rampe“)
- **Graphische Darstellungsformen** (z. B. realitätsnahes Bild vs. Diagramm)

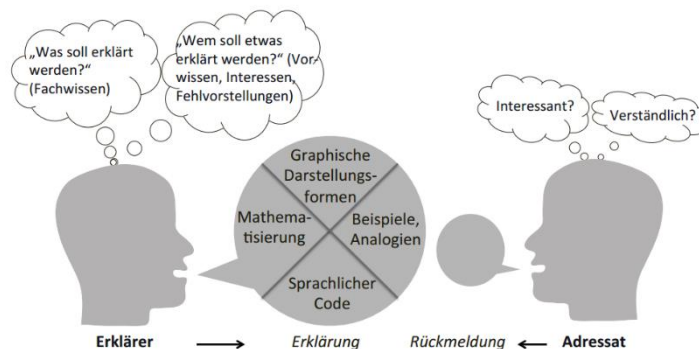


Abb. 1: Dialogisches Modell des Erklärens nach [2]

### 2.3. Einordnung und Definition von „Erklärbarkeit“

Erklären ist eine Standardsituation des Physikunterrichts, zu deren Bewältigung die Lehrkraft sowohl auf Fachwissen, als auch auf fachdidaktisches und pädagogisches Wissen zurückgreifen muss [13]. Ein unmittelbar positiver Effekt von fachdidaktischem Wissen auf Erklärbarkeit konnte inzwischen gezeigt werden. Fachwissen ist ebenfalls förderlich für Erklärbarkeit, hat aber lediglich einen indirekten Einfluss, der durch fachdidaktisches Wissen vermittelt wird (z. B. [14]).

Dies führt uns vor dem Hintergrund des in 2.2 vorgestellten Modells zu folgender Definition (vgl. hierzu auch [7]):

Erklärfähigkeit ist die durch professionelles Wissen beeinflusste Fähigkeit von Lehrkräften zur Bewältigung von Erklärsituationen. Personen mit hoher Erklärfähigkeit können physikalische Sachverhalte unter Berücksichtigung von Aspekten der Sachgerechtigkeit und Adressatengemäßheit so aufbereiten, dass das Kommunikationsangebot bei einem Adressaten mit hoher Wahrscheinlichkeit zur Konstruktion von Bedeutung führt.

### 3. Messen von Erklärfähigkeit

#### 3.1. Zum Problem der Handlungsvalidität beim Testen professionellen Wissens

[15] unterteilt die Messung professioneller, d.h. berufsbezogener Fähigkeiten in vier Kategorien. Diese lauten mit ansteigendem Grad realer Handlung und typischerweise absteigender Standardisierung: Wissenstests (Test von deklarativem Wissen), Kompetenztests (i.d.R. schriftliche Tests zu handlungsbezogenem Wissen), Performanztests (Handlungsbezogene Tests in realen, aber streng standardisierten Testumgebungen) und Handlungstests (Tests in realen Situationen wie Videografieren im Unterricht).

Obwohl schriftliche Tests den niedrigsten Grad an Authentizität bieten, werden sie regelmäßig eingesetzt, um Professionswissen zu erheben, denn sie sind aufgrund des hohen Grads an Standardisierung und der möglichen Umsetzung in komplett geschlossenem Antwortformat für Large-Scale Assessments besonders gut geeignet (z. B. [16]).

Es ist jedoch zu bedenken, dass der Mangel an Authentizität auch einen Mangel an (Handlungs-) Validität nach sich ziehen kann, sodass schriftliche Tests eben gerade keine gute Vorhersage über tatsächliches professionelles Handeln treffen können [17]. Auch [15] weist darauf hin, dass der Erfolg in schriftlichen Tests kein hinreichender Prädiktor für die Handlungsqualität in realen Situationen sei.

#### 3.2. Erklären im Performanztest

Vor diesem Hintergrund haben [2] einen Performanztest für Erklärfähigkeit entwickelt. Sie haben in einer Videostudie etwa 100 Lehramtsstudierende der Physik in standardisierten, dialogischen Erklärsituationen untersucht. Die Probanden wurden dabei einzeln getestet und hatten standardisiert 10 Minuten zur Verfügung, um sich auf die Erklärung eines gegebenen Themas mit curricularem Anschluss zum Physikunterricht der Mechanik (z. B. „Warum rutscht ein Motorradfahrer bei nasser Straße häufiger aus der Kurve?“) vorzubereiten. In weiteren 10 Minuten erklärten sie dann einem Schüler oder einer Schülerin dieses Szenario. Die Erklärung wurde videografiert.

Als geschulte Adressaten waren die Schülerinnen und Schüler dazu angeleitet, standardisierte Impulse zur Variation der vier oben dargestellten Variablen in die Erklärung einzustreuen, etwa „das war mir zu kompliziert, können Sie das nochmal einfacher erklären?“ oder „können wir uns das nochmal an einem Bild anschauen?“. Die Videos der Erklärungen wurden kategorienbasiert kodiert. Daraus wurde die Erklärfähigkeit abgeleitet und quantifiziert.

Vorteile dieses Testformats liegen im starken Bezug zu realen Unterrichtssituationen, was die Handlungsvalidität im Sinne von [17] erhöht. Gleichzeitig leidet das Instrument jedoch unter den logistischen Anforderungen an die Durchführung und dem enormen Zeitaufwand der Datenerhebung und -auswertung [2,14].

#### 3.3. Videovignetten als testökonomische Alternative

Vignetten sind kurze Sequenzen, die einem Probanden z. B. als Text oder Video zur Verfügung gestellt werden, um ihn in eine bestimmte Situation hineinzusetzen. Dabei handelt es sich wahlweise um reale oder fiktional konstruierte, in sich geschlossene Szenen. Typischerweise endet die Sequenz an einer definierten, häufig auch problematischen Stelle und der Proband wird aufgefordert, sie zu bewerten oder ein weiteres, aus seiner Sicht sinnvolles Handeln zu beschreiben [z. B. 18]. Nach [19] haben sich Vignetten-tests als guter Mittelweg zwischen Authentizität und Testaufwand beim Vorhersagen der Qualität professionellen Handelns erwiesen (vgl. auch [20]).

[21] beschreibt insbesondere Formate mit Videovignetten als interessant, da sie professionelles Wissen handlungsnah testeten und so über die Aktivierung von deklarativem Wissen hinausgingen. Dadurch ergäbe sich ein großes Potential für die empirische Beforschung der Lehrerbildung. Gleichzeitig müssten gegenüber Performanztests in Realsituationen jedoch einige Abstriche gemacht werden. Die Handlungsfreiheit sei deutlich eingeschränkt, nur vorhergesehene Handlungswege seien möglich. Ferner könne auch bei umfangreicher Angabe von Informationen über die Vignette das Wissen, das eine Lehrkraft in der Realität über ihre Lernenden hat, nicht vollständig ausgeglichen werden.

Eine weitere Hürde ist es häufig, geschlossene Antwortformate zu integrieren. Welche Lösungen in einer Vignette (z. B. Auswahl des weiteren Verhaltens einer Lehrkraft) sind richtig? Typischerweise werden richtige Lösungen zunächst theoriegeleitet definiert und dann durch normative Expertenurteile bestätigt [19,22].

Diesen Schwierigkeiten zum Trotz gibt es eine große Zahl von Studien, die in vielfältiger Weise empirische Daten zur Lehrerbildung unter Einsatz von Videovignetten liefern (vgl. z. B. das Review von [23]). Dabei zeigt sich, dass unter anderem im Grad der Standardisierung erhebliche Unterschiede bestehen. Um die Spannweite aufzuzeigen sei hier auf zwei Studien

verwiesen: [24] untersuchen das spontane Handlungsvermögen von Mathematiklehrkräften. Dazu zeigen sie den Probanden Vignetten mit Schülergesprächen aus dem Mathematikunterricht. Am Ende der Vignetten werden die Probanden aufgefordert z. B. die Lernenden von ihrer Fehlvorstellung abzubringen oder sie beim Lösen einer Aufgabe zu unterstützen. Die Reaktionen der Probanden werden videografiert, und qualitativ ausgewertet. [25] zeigen im Sinne ihres sogenannten *Advokatorischen Ansatzes* den Probanden Vignetten mit Unterrichtssituationen aus der Berufsschule und lassen sie das Verhalten der Lehrkraft auf Likertskalen bewerten.

Hier wird erkennbar, dass auch innerhalb des Einsatzes von Videovignetten große Unterschiede hinsichtlich der Nähe zu realen Handlungssituationen denkbar sind. Ferner ist naheliegend, dass nur Designs mit geschlossenen Antwortformaten einen testökonomischen Vorteil gegenüber Videostudien wie zum Beispiel dem Performanztest von [2] bieten.

#### 4. Computerbasiertes Messen von Erklärfähigkeit mit Videovignetten

##### 4.1. Forschungsziele

Im hier vorgestellten Projekt sollen folgende Forschungsfragen beantwortet werden:

FF 1: (Wie) kann Erklärfähigkeit mit einem standardisierten, computerbasierten Test gemessen werden?

FF2: Gibt es Hinweise auf einen positiven Einfluss der Erklärfähigkeit von Lehrkräften auf den Lernerfolg, den Sie in kontrollierten Erklärsituationen bei ihren Schülerinnen und Schülern erreichen?

Dazu soll ein interaktiver, computerbasierter Test für Erklärfähigkeit erstellt werden, der auf Videovignetten basiert und dazu geeignet ist, Erklärfähigkeit als handlungsnahe Ressource valide zu messen (FF 1). Gleichzeitig soll das Instrument hinsichtlich Testdurchführung und –Auswertung so einfach durchführbar sein, dass auch ohne hohen personellen Aufwand Large-Scale Assessments möglich sind.

So sollen weitere Einblicke die Wirkung des Erklärens im Physikunterricht möglich werden. Dabei gilt ein besonderes Augenmerk der Untersuchung des Einflusses der Erklärfähigkeit von Lehrkräften auf den Lernerfolg der von ihnen unterrichteten Schülerinnen und Schülern (FF 2).

##### 4.2. Design des Testinstruments

Unser Test basiert auf Videovignetten, wie sie von [19] beschrieben werden. Bei den Vignetten handelt es sich um Sequenzen einer Erklärsituation, die an das Design des Performanztests von [2] angelehnt sind. Sie stellen also den Dialog zwischen einer Lehrkraft („Herr Lehmann“) und einer Schülerin („Pauline“) dar, in dem die Lehrkraft versucht, der Schülerin einen bestimmten Sachverhalt zu erklären und dabei

auf die Äußerungen der Schülerin reagieren muss. Auch hier handelt es sich bei diesen Äußerungen um Impulse, die eine Variation der Erklärung hinsichtlich der vier Variablen des Erklärens motivieren soll. Z. B. „Das klingt ganz schön kompliziert, können Sie das nicht ‘mal an einem Beispiel erklären?’“ Der hier vorgestellte Videovignettentest testet daher insbesondere die Komponente „Adressatengemäßheit“ der Erklärfähigkeit. Die Fähigkeit zu sachgerechter Erklärung wird vor Beginn des Vignettentests in einem rein schriftlichen Test abgeprüft, der sowohl Aspekte der fachlichen Korrektheit (Fachwissenstest), als auch der Vollständigkeit (Welche Themenbereiche müssen der Schülerin zum Verständnis des Phänomens erklärt werden?) beinhaltet. Abbildung 2 zeigt eine typische Darstellung innerhalb einer Vignette.



Abb. 2: Erklärsituation in einer Videovignette.

Der Vignettentest basiert auf einem two tier multiple choice-Design (vgl. z. B. [26,27]). Er besteht aus zweistufigen Fragen. In der ersten Stufe wird der Proband gebeten, aus gegebenen Möglichkeiten, die Erklärung fortzusetzen, jene zu wählen, die seiner Meinung nach am besten geeignet ist, um das Verständnis des Adressaten zu erhöhen. Die erste Stufe wird mit geschlossenen Aufgaben realisiert. In der zweiten Stufe wird er gebeten, das ausgewählte Vorgehen zu begründen. Die zweite Stufe ist in der ersten Version des Tests mit einem offenen Aufgabenformat realisiert.

Nach dieser Auswahl wird der tatsächliche Fortgang der Erklärsituation dargestellt, wobei das gezeigte Verhalten einer der Antwortmöglichkeiten entspricht, aber nicht immer die beste Lösung ist. Diese Sequenz stellt bereits die nächste Videovignette dar, denn am Ende folgt wieder eine Reaktion des Adressaten, die zur Modifikation der Erklärung auffordert und somit zum nächsten Item überleitet. Insgesamt wird so mit etwa 20 Videovignetten in einer Testzeit von ca. 45 Minuten eine zusammenhängende Erklärsituation dargestellt, in deren Verlauf die vier Variablen des Modells zum dialogischen Erklären aufgegriffen werden (siehe 4.3 und 4.4).

Eine Herausforderung besteht in der Wahl geeigneter Antwortmöglichkeiten, die sowohl realistisch als auch theoretisch fundiert richtig oder falsch sind und die Items gleichzeitig nicht zu einfach machen. Neben theoriegeleiteten Überlegungen wurden Experten mit Lehr- und (physikdidaktischer) Forschungserfahrung zurate gezogen, die die Korrektheit der Antworten verifizierten. Ein umfangreicheres Expertenpanel

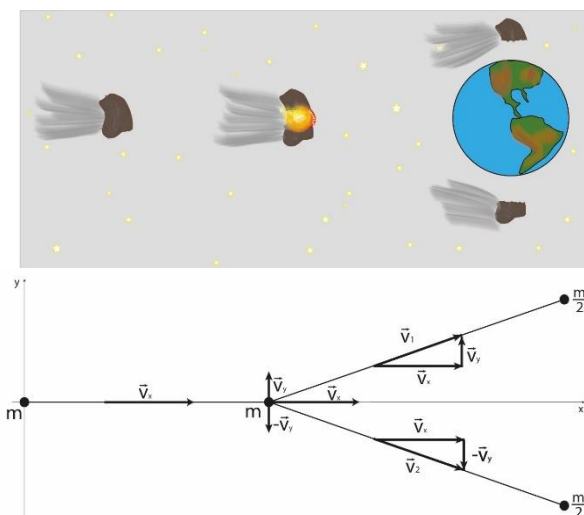
wird zu einem späteren Zeitpunkt durchgeführt.

Die Bewertung der Erklärfähigkeit ergibt sich aus der Summe der richtig getroffenen und richtig begründeten Entscheidungen. In jedem Item erfolgt eine getrennte Bewertung für die Auswahl der richtigen Antwortoption und für die dazu gewählte Begründung. Ein Proband, der die richtige Auswahl für den Fortgang der Erklärung trifft und diese Auswahl auch noch didaktisch fundiert begründet, wird besser bewertet, als ein Proband, der die richtige Auswahl trifft, diese aber nicht sinnvoll begründen kann.

Die Vignetten sind dabei in einer online verfügbaren Testumgebung mit der quelloffenen Software *Li-meSurvey* eingebettet (vgl. [28]). Der Test kann betriebssystemunabhängig von jedem internetfähigen Computer aus durchgeführt werden, wobei die Testumgebung auf unseren Servern liegt, was eine hohe Datensicherheit ermöglicht.

### 4.3. Beispielszenario

Ein in unserem Test eingesetztes Szenario ist die Rettung der Erde vor einem herannahenden Asteroiden. Wie in einem bekannten Kinofilm soll der Asteroid in zwei Hälften gesprengt werden, die durch den so erhaltenen Zusatzimpuls schräg an der Erde vorbeifliegen. Das Szenario hat einen engen inhaltlichen Bezug zur Mechanik im Physikunterricht der 10. Klasse (z. B. Impulserhaltung, Superposition). Die Probanden werden durch eine kurze Präsentation des Szenarios auf den Test vorbereitet. In einigen der Items steht ihnen eine Auswahl unterschiedlich komplexer Darstellungen zur Verfügung. Abbildung 3 zeigt Beispiele für zwei Darstellungen von unterschiedlichem Abstraktionsgrad.



**Abb. 3:** Typische Darstellungsformen des Beispielszenarios.

### 4.4 Beispielitem

Das folgende (gekürzte) Item ist in das in 4.3 dargestellte Szenario der Rettung der Erde vor einem Asteroideneinschlag eingebettet und bezieht sich auf den Einsatz von Beispielen. Der Dialog der Videovignette

ist transkribiert. Danach folgt das zweistufige Item. In der ersten Stufe sind drei der vier Auswahlmöglichkeiten dargestellt. Für die zweite Stufe wurden noch keine geschlossenen Antworten entwickelt. In der Erprobungsphase wird hier in Freitext geantwortet. Ein Screenshot der Videovignette wurde bereits in Abbildung 2 gezeigt.

L:	<i>[...] Die Bombe wird so platziert, dass der Asteroid in zwei Hälften bricht. Durch die Sprengung bekommen die beiden Hälften einen Kraftstoß und bewegen sich auseinander. Das kannst Du dir vorstellen wie einen Stoß. Der Asteroid bricht in zwei gleichgroße Hälften und die werden voneinander weggestoßen.</i>
S:	<i>Okay... Das verstehe ich irgendwie noch nicht, was da mit den Asteroidenhälften passiert.</i>
L:	<i>Ich versuche es mal an einem Beispiel!</i>

#### Stufe 1 (weiteres Vorgehen auswählen)

*Wie sollte der Lehrer vorgehen, damit die Schülerin ihn besser versteht?*

- Stell dir vor, wir rollen nebeneinander auf Bürostühlen auf einen Tisch zu. Kurz bevor wir den Tisch erreichen, stoßen wir uns voneinander ab und rollen dadurch schräg rechts und links am Tisch vorbei. Wir sind dann die Hälften des Asteroids und der Tisch ist die Erde, an der die Asteroidenhälften vorbeifliegen.
- Nehmen wir an du stehst am Billardtisch und eine Kugel rollt an dir vorbei. Stell dir vor, du würdest sie kurz mit dem Finger antippen, während sie an dir vorbeirollt, und zwar senkrecht zur Bewegungsrichtung. Dann würde sie sich ja auch zur Seite bewegen, wie die Asteroidenhälfte nach der Sprengung.
- Stell dir vor, du befindest dich auf einem Ruderboot auf dem See. Du hast deine Paddel verloren, aber ein paar Steine im Boot. Wenn Du die Steine jetzt hinten herauswirfst, kannst Du dich auf das Ufer zubewegen, weil Du die Steine zwar beim Werfen wegdrückst, aber er gleichzeitig auch dich.

#### Stufe 2 (Begründung geben)

*Wieso haben Sie sich für dieses Vorgehen entschieden?*

In der ersten Stufe ist Antwort a die richtige Lösung. Sie ist das einzige Beispiel, das die Situation der Sprengung in zwei Teile beschreibt und gleichzeitig einen konkreten sprachlichen Bezug zwischen der realen Situation der Sprengung und dem Modell (der rollenden Stühle) herstellt.

## 5. Projektverlauf

### 5.1. Aktueller Stand

Unser Projekt ist in drei Phasen angelegt (siehe Abbildung 4). Die erste Phase mit technischen Vorarbeiten und der Entwicklung eines Pilottests ist abgeschlossen.

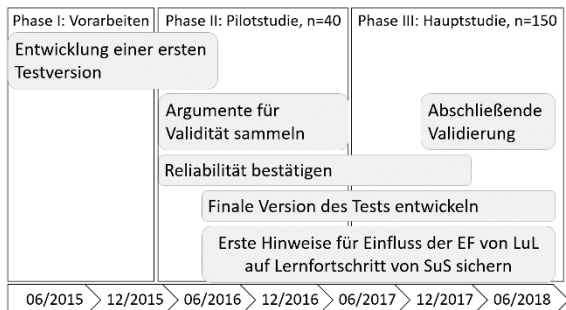


Abb. 4: Verlaufsplan des Projekts.

Zurzeit befinden wir uns am Beginn der zweiten Phase und pilotieren mit  $N=40$  Probanden das vorläufige Instrument mit geschlossenen Items in der ersten Stufe der Fragen (*multiple choice*) und offenen Fragen in der zweiten Stufe (schriftliche Begründung). Bei den Probanden handelt es sich um Testpersonen aus der ersten und zweiten Phase der Lehramtsausbildung im Fach Physik. Darüber hinaus werden in einem *stimulated recall* Verfahren Interviews mit einzelnen Probanden geführt, um weitere Hinweise z. B. über die Durchführbarkeit des Tests zu erhalten.

Die so erhaltenen Informationen fließen in die weitere Testentwicklung ein und bilden die Grundlage zur Überarbeitung des Pilottests und zur Entwicklung weiterer Items. Der Pilottest beinhaltet nur das fachliche Szenario, das in 4.3 dargestellt ist. Weitere Szenarien werden zeitnah entwickelt. Wesentliche Meilensteine der aktuellen Projektphase sind eine anfängliche Untersuchung zur Validität und erste Schritte in einer Bestätigung der Reliabilität.

### 5.2. Studien zur Validität

Wir ergreifen zwei von [19] vorgeschlagenen Maßnahmen zur Erhöhung der Validität:

Die erste Maßnahme bezieht sich auf die Erstellung des Videomaterials. Alle Antwortmöglichkeiten, die in Fragen des Instruments auftauchen, sind aus real beobachteten Handlungen entnommen. Die im Test dargestellten Dialoge sowie die Auswahlmöglichkeiten zur Wahl des weiteren Vorgehens in der Erklärung entsprechen den Verhaltensweisen verschiedener Probanden der Videostudie von [2]. Für unseren Test wurden die Dialoge lediglich mit Schauspielern neu verfilmt, um bessere Bild- und Tonqualität zu gewährleisten sowie die Anonymität der Probandinnen und Probanden zu wahren. Die Auswahlmöglichkeiten zur Begründung des gewählten Vorgehens werden zurzeit aus qualitativ ausgewerteten Freitextantworten der Pilotstudie erzeugt, wobei theoretische Überlegungen mit einbezogen werden. In diesem Test tauchen also zum überwiegenden Teil nur Handlungsoptionen als Distraktoren bzw. Attraktoren auf,

die unter realen Bedingungen auch tatsächlich auftreten.

Die zweite Maßnahme bezieht sich auf die Erzeugung von Handlungsdruck, um einer realen Erklärsituation gerecht zu werden. Die Bearbeitung für die einzelnen Items ist jeweils begrenzt. Die zur Verfügung stehende Zeit ist dabei an durchschnittlichen Bearbeitungszeiten von Probanden aus einer Pilotstudie orientiert und für jedes Item individuell. Sie lässt genug Zeit, um alle Antwortmöglichkeiten zu lesen, stellt aber dennoch sicher, dass der Proband seine Entscheidung zum weiteren Vorgehen in einem Zeitraum treffen muss, der ihm prinzipiell auch in realen Unterrichtssituationen noch zur Verfügung stünde. Die Videovignetten können ferner nur jeweils einmal betrachtet werden.

Ferner nutzen wir den Performanztest von [2] zur konkurrenten Validierung, in dessen Zusammenhang bereits umfangreiche Studien zur Konstruktvalidität vorgenommen wurden. Die mit diesem Performanztest gewonnenen Daten lassen Rückschlüsse auf die Erklärbarkeit zu. Daher lassen wir die Probanden der Pilotstudie sowohl den Performanztest als auch unseren computerbasierten Test durchführen und untersuchen den Zusammenhang zwischen den Ergebnissen. Diese Untersuchung soll bis zum Ende der aktuellen Projektphase abgeschlossen sein, um das weitere Vorgehen auf starke Argumente der Validität stützen zu können. Ferner sollte Erklärbarkeit bis zu einem gewissen Grad unabhängig vom erklärten fachlichen Szenario sein. Daher werden zu einem späteren Zeitpunkt die Ergebnisse der Probanden in unterschiedlichen fachlichen Szenarien untersucht. Es sollten dieselben Fähigkeitsausprägungen diagnostiziert werden.

### 5.3. Ausblick

In der dritten Phase des Projekts soll für die Hauptstudie ein finaler Test entwickelt und erprobt werden, der unterschiedliche fachliche Szenarien enthält und komplett auf ein geschlossenes *two tier multiple choice*-Design zurückgreift. Dieses Instrument soll mit  $N=150$  Probanden durchgeführt werden, wobei hier fertig ausgebildete Lehrkräfte der Physik eingesetzt werden sollen. Die Auswertung der Testergebnisse soll dabei vollautomatisch erfolgen und sowohl die Korrektheit der getroffenen Entscheidungen zum weiteren Verlauf der Erklärung, als auch die dazu abgegebenen Begründungen berücksichtigen.

In dieser Phase werden auch Untersuchungen zum Zusammenhang der gemessenen Erklärbarkeit von Lehrkräften und dem Lernzuwachs der von ihnen unterrichteten Schülerinnen und Schülern in Erklärsituationen durchgeführt. Auch wenn hier keine vollständige Analyse des Kausalzusammenhangs möglich sein wird, streben wir doch zumindest eine Exploration mit ersten Hinweisen auf einen positiven Effekt der Erklärbarkeit an. Dazu werden Probanden gebeten in kontrollierten Erklärsituationen vor ihrer Klasse ein physikalisches Phänomen zu erklären. Der

Lernzuwachs der Lernenden wird in einem Pre-Post Design erhoben. Anschließend führen die Lehrkräfte den computerbasierten Test zur Erklärfähigkeit aus und Zusammenhänge zwischen dem Lernzuwachs der Schüler und der Erklärfähigkeit der Lehrkräfte werden untersucht.

## 6. Literatur

- [1] Kulgemeyer, C., Schecker, H. (2013). Students Explaining Science - Assessment of Science Communication Competence. *Res Sci Educ* 43/6, 2235–2256.
- [2] Kulgemeyer, C., Tomczyszyn, E. (2015). Physik erklären. Messung der Erklärens-fähigkeit angehender Physiklehrkräfte in einer simulierten Unterrichtssituation. *Zeitschrift für Didaktik der Naturwissenschaften* 21/1, 111–126.
- [3] Merzyn, G. (2005). Junge Lehrer im Referendariat. Der mathematische und naturwissenschaftliche Unterricht 1, 4–7.
- [4] Wilson, H., Mant, J. (2011). What makes an exemplary teacher of science? The pupil's perspective. *School Science Review* 93/343, 121–125.
- [5] Osborne, J. F., Patterson, A. (2011). Scientific argument an explanation: A necessary distinction? *Science Education* 4, 627–638.
- [6] Geelan, D. (2012). Teacher Explanations. In: *Second International Handbook of Science Education*. Fraser, B., Tobin, K., McRobbie, C. (Hrsg.). Springer, Dordrecht, 987–999.
- [7] Kulgemeyer, C. (2013). Gelingensbedingungen physikalischer Erklärungen. Zu einer konstruktivistischen Auffassung des Erklärens. *PhyDid B - Didaktik der Physik - Beiträge zur DPG Frühjahrstagung*.
- [8] Hempel, C., Oppenheim, P. (1948). Studies in the logic of explanation. *Philosophy of Science* 15/2, 135–175.
- [9] Martin, J. (1970). *Explaining, Understanding and Teaching*. McGraw-Hill, New York.
- [10] Treagust, D., Harrison, A. (1999). The Genesis of Effective Science Explanations for the Classroom. In: *Researching Teaching: methodologies and Practices for Understanding Pedagogy*. Loughran, J. (Hrsg.). Routledge, Abingdon, 28–43.
- [11] Kulgemeyer, C., Schecker, H. (2009). Kommunikationskompetenz in der Physik. Zur Entwicklung eines domänenspezifischen Kommunikationsbegriffs. *Zeitschrift für Didaktik der Naturwissenschaften*; 15, 131–153.
- [12] Clark, H., Brennan, S. (1991). Grounding in Communication. In: *Perspectives on socially shared cognition*. Resnick, J., Levine, J., Teasley, S. (Hrsg.). American Psychological Association, Washington, 127–149.
- [13] Riese, J., Kulgemeyer, C., Zander, S., Borowski, A., Fischer, H. E., Gramzow, Y., Reinhold, P., Schecker, H., Tomczyszyn, E. (2015). Modellierung und Messung des Professionswissens in der Lehramtsausbildung Physik. *Zeitschrift für Pädagogik* 61, 55–79.
- [14] Kulgemeyer, C., Tomczyszyn, E., Schecker, H. (2016). Was beeinflusst die Performanz beim Erklären von Physik? Fachwissen und fachdidaktisches Wissen im unterrichtlichen Handeln. In: *Authentizität und Lernen - das Fach in der Fachdidaktik*. Bernholt, S. (Hrsg.).
- [15] Miller, G. (1990). The Assessment of Clinical Skills / Competence / Performance. *Journal of the Association of American Medical Colleges* 65/9, 563–567.
- [16] Tepner, O., Borowski, A., Dollny, S., Fischer, H. E., et al. (2012). Modell zur Entwicklung von Testitems zur Erfassung des Professionswissens von Lehrkräften in den Naturwissenschaften. *Zeitschrift für Didaktik der Naturwissenschaften* 18, 7–28.
- [17] Vogelsang, C., Reinhold, P. (2013). Zur Handlungsvalidität von Tests zum professionellen Wissen von Lehrkräften. *Zeitschrift für Didaktik der Naturwissenschaften* 19, 103–128.
- [18] Schratz, M., Schwarz, J. F., Westfall-Greiter, T., Rumpf, H. (2012). Lernen als bildende Erfahrung. Vignetten in der Praxisforschung. StudienVerlag, Innsbruck.
- [19] Rehm, M., Bölsterli, K. (2014). Entwicklung von Unterrichtsvignetten. In: *Methoden in der naturwissenschaftsdidaktischen Forschung*. Krüger, D., Parchmann, I., Schecker, H. (Hrsg.). Springer Spektrum, Berlin, Heidelberg, 213–225.
- [20] Aufschnaiter, C. von, Blömeke, S. (2010). Professionelle Kompetenz von (angehenden) Lehrkräften erfassen – Desiderata. *Zeitschrift für Didaktik der Naturwissenschaften* 16, 361–367.
- [21] Neuweg, G. H. (2015). Kontextualisierte Kompetenzmessung. *Zeitschrift für Pädagogik* 61/3, 377–383.
- [22] König, J., Blömeke, S., Klein, P., Suhl, U., Busse, A., Kaiser, G. (2014). Is teachers' general pedagogical knowledge a premise for noticing and interpreting classroom situations? A video-based assessment approach. *Teaching and Teacher Education* 38, 76–88.
- [23] Lindmeier, A. (2013). Video-vignettenbasierte standardisierte Erhebung von Lehrerkognitionen. In: *Videobasierte Kompetenzforschung in den Fachdidaktiken*. Macha, K., Riegel, U. (Hrsg.). Waxmann, Münster, 45–61.
- [24] Lindmeier, A. M., Heinze, A., Reiss, K. (2013). Eine Machbarkeitsstudie zur Operationalisierung aktionsbezogener Kompetenz von Mathematiklehrkräften mit videobasierten Maßen. *J Math Didakt* 34/1, 99–119.
- [25] Oser, F., Forster-Heinzer, S. (2015). Wer setzt das Maß? Eine kritische Auseinandersetzung mit dem Advokatorischen Ansatz. *Zeitschrift für Pädagogik* 61/3.
- [26] Tan, K. C. D., Goh, N. K., Chia, L. S., Treagust, D. F. (2002). Development and Application of a Two-Tier Multiple Choice Diagnostic Instrument To Assess High School Students'

- Understanding of Inorganic Chemistry Qualitative Analysis. *Journal of Research in Science Teaching* 39/4, 283–301.
- [27] Urban-Woldron, H., Hopf, M. (2012). Entwicklung eines Testinstruments zum Verständnis in der Elektrizitätslehre. *Zeitschrift für Didaktik der Naturwissenschaften* 18, 201–227.
- [28] LimeSurvey Project Team (2015). *Lime Survey: An Open Source Survey Tool*. Hamburg.