

Entwicklung und Evaluation eines zweistufigen Testinstruments für Schülervorstellungen zur Anfangsoptik

Albert Teichrew*, Roger Erb*

*Institut für Didaktik der Physik, Goethe-Universität Frankfurt am Main
teichrew@physik.uni-frankfurt.de

Kurzfassung

Die Kenntnis von alternativen Vorstellungen bei Schülerinnen und Schülern ist eine Voraussetzung für die Entwicklung von Lernumgebungen zur Förderung physikalisch gültiger Konzepte. Die erste Begegnung mit der Physik stellt häufig die Optik dar, weshalb hier eine frühzeitige Etablierung von anschlussfähigen Denkmodellen für die weitere Beschäftigung mit der Physik ausschlaggebend sein kann. Um gedankliche Strukturen aufzudecken, sind offene Erhebungsmethoden wie Interviews oder Concept Maps zielführend. Um die Wirksamkeit von Lernumgebungen zu evaluieren, werden jedoch verfahrensökonomischere Methoden benötigt. Allerdings ist die Aussagekraft von Multiple-Choice Tests bezüglich der Vorstellungen, auf denen eine Auswahl basiert, oft eingeschränkt. Einen Lösungsansatz stellen zweistufige Testverfahren dar, die mithilfe bereits bekannter Schülervorstellungen entwickelt werden. Als Indikator einer bestimmten Vorstellung wird dabei die übereinstimmende Kombination aus Vorhersage und Begründung einer dargestellten Ausgangssituation gesehen. Der Beitrag behandelt das Testverfahren sowie die Testentwicklung und präsentiert Ergebnisse aus Erhebungen und Itemanalysen.

1. Hintergrund

Schülerinnen und Schüler greifen beim Erlernen neuer Inhalte oder beim Lösen von Aufgaben häufig auf Vorstellungen zurück, die sie vor dem Unterricht zu einem bestimmten Sachverhalt erworben haben. Diese sich im Alltag als fruchtbar erwiesenen Vorerfahrungen oder auch unterrichtsinduzierte Ansichten zu naturwissenschaftlichen Phänomenen, Begriffen und Prinzipien werden unter dem Begriff „Schülervorstellungen“ seit der Mitte der 70er Jahre intensiv untersucht [1]. Obwohl sie sich selten mit den geförderten Konzepten überschneiden und zu Lernschwierigkeiten führen können, birgt ihre Kenntnis Verbesserungspotential für den Unterricht.

Im Rahmen eines Promotionsvorhabens zu physikalischer Modellbildung mithilfe Dynamischer Geometrie-Software wird ein Schülerlabor etabliert, das mit dynamischen Modellen [2] und Experimenten physikalisch gültige Konzepte der Anfangsoptik bei Schülerinnen und Schülern der Klassenstufe 7 fördert. Der im Folgenden vorgestellte Test wurde mit dem Ziel entwickelt, die in der Zielgruppe am weitesten verbreiteten Schülervorstellungen als Lernvoraussetzungen zu erheben, die Ergebnisse bei der Entwicklung des Schülerlabors zu berücksichtigen und die Wirksamkeit des Schülerlabors zu messen.

Das Vorgehen basiert auf dem konstruktivistischen Modell der Didaktischen Rekonstruktion, das drei wechselwirkende Teilaufgaben beinhaltet: Fachliche Klärung, Erfassung der Perspektive der Lernenden und didaktische Strukturierung [3]. Lernen wird darin als aktive Modifizierung oder Rekonstruktion

vorunterrichtlicher Vorstellungen oder erworbener Konzepte angesehen. Wissenschaftliche Perspektiven und Schülervorstellungen werden als persönliche Konstrukte der jeweiligen Personengruppen gleichwertig behandelt. Während der didaktischen Strukturierung können aus dem Vergleich der Ansichten Leitlinien für den Unterricht entwickelt werden, die Korrespondenzen zwischen dem Fachwissen und individuellen Lernbedingungen beinhalten [4]. Die Kenntnis der Perspektiven der Lernenden wird somit als Voraussetzung für die Entwicklung von effektiven Lernumgebungen angesehen, was eine zeitnahe Erhebung der Schülervorstellungen innerhalb der Zielgruppe erforderlich macht.

Für die Untersuchung von Schülervorstellungen gibt es verschiedene Erhebungsmethoden, die je nach Forschungslage zum Einsatz kommen können. Diskursiv-dialogische Interviews ermöglichen durch offene Fragen mit zunehmend gezielteren Nachfragen das Erfassen subjektiver Sinnbezüge, Sicht-, und Handlungsweisen [5]. Ebenso können Concept Maps als Diagnoseinstrument von Schülervorstellungen verwendet werden [6]. Diese Methode erlaubt es zum Beispiel bei Vorgabe bestimmter Begriffe, anhand ihrer spezifischen Anordnung und Verbindung durch die Lernenden, auf vorliegende Sinnzusammenhänge zu schließen. Mithilfe von diesen und anderen qualitativen Forschungsmethoden, die oft mit Bildmaterial oder Experimenten unterstützt werden, lassen sich alternative Schülervorstellungen zu einem neuen Themengebiet anhand einer kleinen Fallzahl erfassen (z. B. [7]). Repräsentative Ergeb-

nisse bezüglich der Ausprägung einzelner Vorstellungen oder ihrer Entwicklung lassen sich damit dagegen nicht ohne großen Aufwand bewerkstelligen. Die anhand von Einzelgesprächen erforschten Vorstellungen bieten jedoch eine sichere Grundlage für die Erstellung von standardisierten, verfahrensökonomischen Verfahren zur Erhebung von Schülervorstellungen. Dazu zählen Multiple-Choice Tests, die in kurzer Zeit Vorstellungen bezüglich verschiedener Inhalte erfassen können. Allerdings besteht selbst bei authentischen Distraktoren, die anhand von Schülerantworten formuliert wurden, die Wahrscheinlichkeit eine richtige Antwort anzukreuzen, ohne das entsprechende Konzept zu besitzen [8]. 1986 machte Treagust den Vorschlag, durch zweistufige Multiple-Choice Items mit dieser Unsicherheit umzugehen [9]. In diesen Tests werden Schülerinnen und Schüler zunächst aufgefordert, eine inhaltliche Frage zu beantworten (Stufe 1) und im nächsten Schritt ihre Antwort zu begründen (Stufe 2). Zur Auswahl stehen dabei Aussagen, die mit wissenschaftlichen Konzepten oder alternativen Vorstellungen im Einklang stehen. Erst mit der passenden Begründung wird dabei ein Item auch als richtig oder falsch gewertet. Im Folgenden wird neben der Entwicklung und Gestaltung des vorgestellten Testinstruments auch die Erhebungs- und Auswertemethode näher beleuchtet.

2. Testentwicklung

Die Erforschung von Schülervorstellungen zur Optik hat eine Vielzahl an Ergebnissen hervorgebracht, die regelmäßig durch aktuelle Studien erweitert werden (z. B. zu weißem Licht und Farben [10]). Angehende Lehrkräfte können somit auf strukturierte Zusammenstellungen verbreiteter Ansichten bezüglich elementarer physikalischer Konzepte zurückgreifen [11].

Die Entwicklung des zweistufigen Testinstruments für Schülervorstellungen zur Anfangsoptik (TIA₂) verlief deshalb in den folgenden vier Phasen:

- Zusammenstellung elementarer physikalischer Konzepte sowie bekannter Schülervorstellungen zur Anfangsoptik
- Auswahl von Situationen mit je nach Vorstellung alternativen Antwortmöglichkeiten
- Gestaltung von Abbildungen sowie Formulierung von Fragestellungen, alternativen Antworten und vorstellungsbezogenen Begründungen
- Neugestaltung der Abbildungen sowie Anpassung der Formulierungen nach Expertenbefragung und Itemanalysen während der Pilotierung

In der ersten Phase wurde eine literaturbasierte Vorauswahl an Konzepten und Vorstellungen getroffen, die unter dem Oberbegriff Anfangsoptik zusammengefasst werden können und noch stark an Alltagserfahrungen anknüpfen (s. Tab. 1). Die Konzepte zu Abbildungsvorgängen mit Linsen oder zur Lichtbre-

chung wurden nicht einbezogen und können mit fachorientierten Tests erhoben werden (z. B. der einstufige Rasch-skalierte Multiple-Choice Fachwissenstest zur geometrischen Optik [12]).

Bei der Auswahl und Anordnung der Objekte zur Gestaltung der Situationen wurden in der zweiten Phase Ansätze aus einer Reihe von bereits konzipierten Tests in Betracht gezogen [13]–[15]. Die Autoren dieser Tests folgen jedoch unterschiedlichen Gestaltungsmustern und Formulierungsrichtlinien. Mitunter wurden ähnliche Situationen verwendet, die entweder durch offene Fragen, ein- oder zweistufige Multiple-Choice Fragen sowie durch Konstruktion verschiedener Strahlengänge eine Diagnose von Schülervorstellungen ermöglichen.

Um eine Standardisierung zu gewährleisten, wurden in der dritten Phase eigene Abbildungen nach einheitlichen Vorgaben gestaltet und mit demselben zweistufigen Testformat ausgestattet. Die 16 im Test verwendeten Abbildungen samt zugehörigen Formulierungen weichen deshalb im Endergebnis von den zur Testentwicklung einbezogenen Vorbildern ab. Die Hälfte der Items entstand in Eigenentwicklung.

Eine erste Version wurde in der vierten Phase von vier Experten kritisch begutachtet. Die eingebundenen Didaktiker verfügen über Expertenwissen zur Optik, zu Schülervorstellungen und Testkonstruktion. Sowohl inhaltliche als auch methodische Schwierigkeiten konnten dank der Expertenbefragung aufgedeckt werden. Nach der ersten Überarbeitung der Abbildungen und Formulierungen wurde der Test in einer 8. Klasse und mit Studierenden des Lehramts im ersten Fachsemester pilotiert. Als Ergebnis konnten zwei Items mit überdurchschnittlich hohen Anteilen ohne inhaltliche Deckung der Antwort mit der Begründung ermittelt und umgestaltet werden. Die dritte und finale Version wurde schließlich für eine Erhebung der Schülervorstellungen in verschiedenen Lerngruppen eingesetzt und die Ergebnisse zur Evaluation des Tests genutzt.

3. Testgestaltung

Bei der Entwicklung des TIA₂ wurde der Versuch unternommen, den Testcharakter der Aktivität einzudämmen. Im Vordergrund spielt sich deshalb eine Geschichte von zwei Protagonisten ab, die in verschiedenen Situationen mit optischen Phänomenen konfrontiert sind. Die zweistufigen Items folgen einem einheitlichen Gestaltungsmuster. Jede Seite des Testheftes enthält folgende Elemente (s. Abb. 1):

- Eine **Fortschrittsanzeige** macht transparent, an welcher Stelle der Bearbeitung man sich befindet.
- Ein **Prozesshinweis** soll die Motivation der Testpersonen aufrechterhalten.
- Eine **Beschreibung der Situation** liefert mit wenigen und kurzen Sätzen relevante Informationen über die Lichtverhältnisse, Personen sowie Anordnungen der Objekte im Raum.

Inhaltsbereich	Physikalisches Konzept	Alternative Vorstellung
Ausbreitung (ohne Wechselwirkung)	Kontinuierlich, unendlich, ungerichtet	Unverzöglich, endlich, gerichtet
	Unabhängig von den Lichtverhältnissen	Abhängig von den Lichtverhältnissen
Licht	Streuung und Reflexion	Licht auf dem Objekt oder auf der Quelle
	Absorption und Emission	Alles-oder-Nichts-Prinzip
	Primäre und sekundäre Lichtquellen	Licht nur aus selbstleuchtenden Objekten
Sehvorgang	Lichtweg Quelle-Objekt-Empfänger	Licht auf dem Objekt oder als Lichtbad im Raum
	Passives Empfangen	Aktives Abtasten
	Licht als Notwendigkeit	Anpassung an die Dunkelheit
Schatten	Fehlendes Licht	Graufärbung
	Schattenbild als Projektion	Schattenbild entspricht der Körperform
	Dreidimensionaler Schattenraum	Zweidimensionale Schattenfläche
Spiegel	Lichtweg Objekt-Spiegel-Empfänger	Aktives Erfassen der Objekte durch den Spiegel
	Spiegelbild hinter dem Spiegel	Spiegelbild auf dem Spiegel
	Ort des Spiegelbildes abhängig vom Ort des Objekts	Ort des Spiegelbildes abhängig vom Ort des Beobachters
	Vorne-Hinten-Umkehrung	Rechts-Links-Umkehrung
Farben	Lichtfarbe als Spektralfarbe oder ihre Zusammensetzung	Farbloses Licht oder Farbmischung von Lichtfarben entspricht Farbmischung von Körperfarben
	Ausfiltern bestimmter Farben (Absorption und Emission)	Ganzheitliches Einfärben des Lichts oder des Objekts (Alles-oder-Nichts-Prinzip)
	Abhängigkeit des Farbreizes von der Körperfarbe und der Lichtfarbe	Lichtfarbe überwiegt Körperfarbe

Tab.1: Inhaltliche Übersicht der im Test vorkommenden Konzepte und Vorstellungen der Anfangsoptik.

- Eine schematische **Abbildung der Situation** illustriert die beschriebene Szene.
- Eine **Fragestellung** fordert zur Abgabe einer Vorhersage auf.
- Drei möglichst erschöpfende **Antwortmöglichkeiten** garantieren eine ungezwungene Meinungsäußerung.
- Eine Frage nach der „besten“ **Begründung der Antwort** deutet auf einen Spielraum in den Formulierungen an.
- Vier **Begründungsmöglichkeiten** stellen eine Auswahl an etwa gleich langen Aussagen dar, die in einer ähnlich einfachen Sprache formuliert sind.

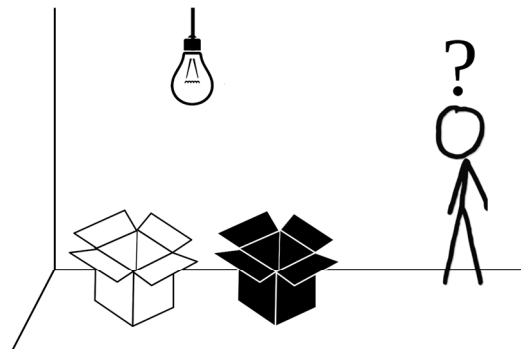
Die Abbildungen sollen Vorstellungen anregen und enthalten deswegen keine Modellelemente wie Lichtstrahlen oder optische Erscheinungen wie Schatten- und Spiegelbilder. Bei ihrer Gestaltung wurde zwischen *tagsüber* (Objekte mit schwarzen Konturen auf weißem Hintergrund) und *nachts* (Objekte mit weißen Konturen auf schwarzem Hintergrund) unterschieden. Letzteres deutet auf wenig bis gar kein Licht im Raum, während sonst eine in diffuses Licht gehüllte Situation ohne klare Schattengrenzen vorliegt. Zusätzliche Punktlichtquellen werden als *Lampen* bezeichnet und Lichtbündel werden durch *Taschenlampen* erzeugt. Als Streuobjekte oder Hindernisse werden *Tische*, *Eimer* oder *Kartons* eingesetzt. Als Empfänger dienen entweder die Augen der Protagonisten oder es wird von vorläufigen Empfängern ausgegangen, wie dem *Boden* oder der *Wand*. *Spiegel* werden als beschriftete Rahmen angedeutet und Bewegungsrichtungen mit Pfeilen angegeben.

Im Unterricht aufkommende Fachbegriffe wie *reflektieren*, *strahlen* oder *streuen* werden bewusst vermieden. Eine Aussage entspricht dabei einer physikalisch vertretbaren Auffassung und drei weitere können auf verbreitete alternative Schülervorstellungen zurückgeführt werden (s. Tab. 1).

Frage 7 von 16

Es kommen gleich Fragen zu Schatten.

Tian will tagsüber in seinem Zimmer das Licht anmachen. Unter der Lampe stehen ein weißer und ein schwarzer Karton.



Von wo wird Licht bei Tian ankommen?

- nur von der Lampe
 von der Lampe und vom weißen Karton
 von der Lampe und von beiden Kartons

Welche Aussage erklärt deine Antwort am besten?

- Schwarze Dinge schlucken das ganze Licht und weiße nicht.
 Angeleuchtete Dinge werfen einen Teil des Lichts zurück.
 Durch die Lampe geht Strom, der sie zum Leuchten bringt.
 Auf beiden Kartons liegt Licht und macht sie sichtbar.

Abb. 1: Item zu primären und sekundären Lichtquellen.

4. Testmethode

Zur Erhebung der Schülervorstellungen im angewandten Testverfahren werden folgende Schritte durchlaufen:

- Wahrnehmung der Situation durch die Testperson anhand einer Beschreibung und der zugehörigen Abbildung
- Von der Wahrnehmung geprägte Antwort zur Lösung der Situation (Stufe 1)
- Auswahl einer physikalischen oder alternativen Begründung für die Antwort (Stufe 2)
- Überprüfung der Art der Antwort und der Begründung sowie ihrer inhaltlichen Deckung

Die Feststellung bei einer Testperson entweder eines physikalischen Konzepts oder einer alternativen Vorstellung erfolgt ausschließlich bei inhaltlicher Deckung der Antwort mit der Begründung. Das Ankreuzverhalten aus der ersten und zweiten Stufe wird damit in Kombination betrachtet und nicht gesondert ausgewertet.

Während der Auswertung wurden deshalb für jedes zweistufige Item folgende Ausgänge kodiert: Die Testperson wählt eine ...

- ... physikalische Begründung, die sich mit der Antwort deckt.
- ... physikalische Begründung, die sich **nicht** mit der Antwort deckt.
- ... alternative Begründung, die sich mit der Antwort deckt.
- ... alternative Begründung, die sich **nicht** mit der Antwort deckt.

Die Art der Begründung steht bei der gewählten Kodierung im Vordergrund, da die Antwort auf die Frage zunächst davon abhängt, wie die dargestellte Situation wahrgenommen und verstanden wurde. Dagegen ist die Begründung in der Regel allgemein formuliert und leitet sich von physikalischen Konzepten oder verbreiteten alternativen Vorstellungen ab. Das Vorhandensein der jeweiligen Vorstellung wird allerdings erst durch die Entscheidung für die entsprechende Antwort bestätigt.

Diese Form der Auswertung stellt einen hohen Anspruch an die Konstruktion der Items dar, da nur folgerichtige Kreuze eine Vorstellungseinschätzung liefern (A&C). Als Alternative könnten richtige Antworten und Begründungen jeweils mit Punkten bewertet und summiert werden. Die Bewertung könnte dabei gleich oder unterschiedlich gewichtet sein, was das Endergebnis allerdings von der Wahl der Gewichtungsfaktoren abhängig machen würde.

5. Ergebnisse der Hauptzielgruppe

Als Hauptzielgruppe für den TIA₂ wird die Klassenstufe 7 betrachtet. Die mit dem Test erhobenen Schülervorstellungen betreffen Inhalte, die üblicherweise in dieser Klassenstufe, am Anfang des Physikunterrichts behandelt werden. Nichtsdestot-

rotz macht die Schülervorstellungsforschung deutlich, dass Schülerinnen und Schüler auch nach dem regulären Unterricht in Interviews oder Tests überwiegend mit alternativen Schülervorstellungen und nicht mit physikalischen Konzepten argumentieren. Die entworfene Intervention in Form eines Schülerlabors, dessen Wirksamkeit gemessen werden soll, setzt bei diesem Zustand an. Zur Bildung einer Normstichprobe wurde der Test deshalb in sieben Lerngruppen der Klassenstufe 7 im Anschluss an eine erste Unterrichtseinheit zur Optik eingesetzt (N = 165). Es nahmen fünf Gymnasialklassen und eine Realschulklasse einer Schule aus dem hessischen Hochtaunuskreis und eine Lerngruppe des Goethe-Schülerlabors aus einem Frankfurter Gymnasium teil.

Der Anteil gültiger Angaben, bei denen sowohl Antwort als auch Begründung genau einmal angekreuzt wurden, liegt bei 94%. Das heißt die Testpersonen hielten sich nicht immer an die Vorgabe, jeweils nur ein Kreuz pro Auswahlmöglichkeit zu setzen, oder haben teilweise keine Angaben gemacht. Bei den gültigen Angaben liegt die relative Häufigkeit ohne inhaltliche Deckung bei 11% (s. Tab. 2). In solchen Fällen lässt sich keine Vorstellung bezüglich eines Sachverhalts zuordnen. Das lässt sich sowohl auf Verständnisschwierigkeiten als auch auf Flüchtigkeitsfehler zurückführen.

Klassenstufe 7 (N = 165)	A	B	C	D
M	4.72	.51	8.58	1.21
SD	1.98	.76	2.25	1.36
H	778	84	1415	199
h	31%	3%	57%	8%

Tab.2: Mittelwert (M), Standardabweichung (SD), absolute (H) und relative Häufigkeit (h) der Ausgänge gültiger Angaben in der Klassenstufe 7.

Das Testergebnis der Stichprobe, bezogen auf diagnostizierte physikalische Konzepte (A), ist erwartungsgemäß niedrig. Die Spannweite physikalisch richtig gelöster Items erstreckt sich von 0 bis 10 bei einem möglichen Maximum von 16 (s. Abb. 2). Der erhobene Anteil alternativer Vorstellungen (C) fällt entsprechend höher aus.

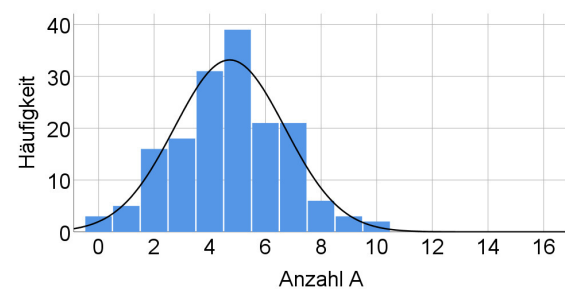


Abb.2: Häufigkeitsverteilung des Ausgangs A.

Es nahmen 82 Schülerinnen und 78 Schüler an dieser Erhebung teil, fünf Testpersonen machten zum Geschlecht keine Angabe (s. Tab. 3). Die Häufigkeit physikalisch gültiger Konzepte fällt bei den männli-

chen Testpersonen höher aus als bei den weiblichen (s. Abb. 3). Der Unterschied weist jedoch in dieser Stichprobe keine statistische Signifikanz auf, $t(158) = 1.022, p = .309$.

Geschlecht (N = 160)		A	B	C	D
weiblich (n = 82)	M	4.63	.41	9.12	1.05
	SD	1.80	.68	1.95	1.28
	H	380	34	748	86
	h	30%	3%	60%	7%
männlich (n = 78)	M	4.95	.58	7.90	1.35
	SD	2.09	.78	2.39	1.44
	H	386	45	616	105
	h	34%	4%	53%	9%

Tab.3: Ergebnisübersicht nach Geschlecht.

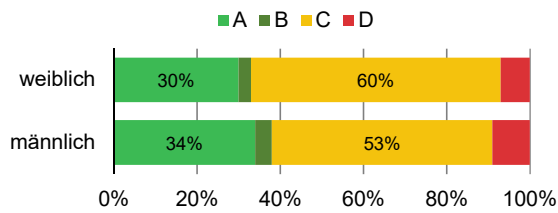


Abb.3: Relative Häufigkeiten der Ausgänge nach Geschlecht.

Allerdings ist die Häufigkeit alternativer Vorstellungen bei den weiblichen Testpersonen signifikant höher als bei den männlichen, $t(158) = -3.558, p < .001, d = -.57$. Gleichzeitig ist die inhaltliche Deckung der Antworten mit der Begründung bei weiblichen Testpersonen (A&C, Median = 14) höher als bei den männlichen (Median = 13), Mann-Whitney-U-Test: $U = 2555, p = .026, r = .18$.¹ Die Ergebnisse deuten darauf, dass Schülerinnen den zweistufigen Test zwar konzentrierter bearbeiten, aber dadurch offenbaren sich eher alternative Vorstellungen als physikalische Konzepte.

Die mit dem Test angesprochenen Vorstellungen und Konzepte lassen sich den vier Inhaltsbereichen Licht, Schatten, Spiegel und Farben zuordnen (s. Tab. 1). Anhand des Diagramms in Abbildung 4 lässt sich erkennen, dass die Items zu Schatten häufiger aus physikalischer Sicht richtig gelöst werden als zu Licht oder Spiegel. Die zwei Items zu Farben zeigen die niedrigste Lösungswahrscheinlichkeit, da das Thema im Anfangsunterricht der Optik selten behandelt wird und in diesem Bereich alternative Schülervorstellungen am weitesten verbreitet sind.

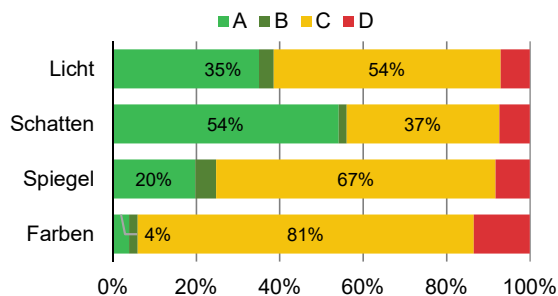


Abb.4: Relative Häufigkeiten der Ausgänge nach Inhalt.

Entsprechend zeigt lediglich eine Subskala aus sechs Items zu Licht eine akzeptable interne Konsistenz mit Cronbachs Alpha = .69 und guten Trennschärfen zwischen .35 und .50 (Auswertung von n = 245 gültigen Fällen aus der Gesamtstichprobe N = 279). Die restlichen Items sprechen insgesamt betrachtet, und auch innerhalb eines Inhaltsbereichs, sehr unterschiedliche Vorstellungen und Konzepte an, sodass man testtheoretisch nicht mehr von der Messung von genau einem latenten Konstrukt ausgehen kann. Im Folgenden werden deshalb Ergebnisse bezüglich bestimmter Schülervorstellungen innerhalb der Inhaltsbereiche berichtet (s. Tab. 4). In der Auswahl befinden sich pro Inhaltsbereich jeweils zwei physikalische Konzepte mit den niedrigsten relativen Häufigkeiten und die alternativen Vorstellungen, die stattdessen diagnostiziert werden konnten.

Inhalt	Beschreibung des Konzepts (A) oder der alternativen Vorstellungen (C1-C3)	h
Licht	A Von selbstleuchtenden und beleuchteten Objekten geht Licht aus.	14%
	C1 Nur von selbstleuchtenden Objekten geht Licht aus.	17%
	C2 Von schwarzen beleuchteten Objekten geht überhaupt kein Licht aus.	36%
	C3 Auf beleuchteten Objekten liegt Licht.	11%
	A Ohne Licht kann man nichts sehen.	18%
	C1 Man kann sich daran gewöhnen, ohne Licht zu sehen.	8%
	C2 Katzen können sich daran gewöhnen, ohne Licht zu sehen.	67%
	C3 Um etwas im Dunkeln zu sehen, muss man lediglich in die richtige Richtung schauen.	2%
	Schatten	A Zum Schatten gehört der ganze Raum hinter dem Objekt.
C1 Im Raum hinter dem Objekt befindet sich mehr Licht als im Schattenbild.		1%
C2 Zum Schatten gehört nur das Schattenbild.		32%
C3 Schatten färbt den Raum hinter dem Objekt unabhängig von den Lichtverhältnissen grau ein.		16%
A Schatten entsteht hinter einem Objekt an den Stellen, an denen kein Licht ankommt.		56%
C1 Die Helligkeit des Schattens ist abhängig von dem Größenunterschied zwischen Objekt und Schattenbild.		11%
C2 Schatten färbt die Fläche hinter einem Objekt unabhängig von den Lichtverhältnissen grau ein.		7%
C3 Die Helligkeit des Schattens ist abhängig von seiner Größe.		15%
Spiegel		A Wenn man den Abstand zum Spiegel vergrößert, ist der sichtbare Bildausschnitt des eigenen Spiegelbildes gleich, weil der Körper und das Spiegelbild sich gleichzeitig vom Spiegel entfernen.
	C1 ... größer, weil der Spiegel so mehr vom Körper einfangen kann.	72%
	C2 ... kleiner, weil die Dinge im Spiegel von weiter weg kleiner erscheinen.	4%
	C3 ... gleich, da es auf dem Spiegel liegt.	6%

Farben	A	Das Spiegelbild entsteht hinter dem Spiegel und ist unabhängig von der Position des Beobachters.	8%
	C1	Der Ort des Spiegelbildes ist abhängig von der Position des Beobachters.	40%
	C2	Das Spiegelbild liegt direkt hinter dem Objekt mit dem Beobachter auf einer Linie.	8%
	C3	Das Spiegelbild entsteht auf dem Spiegel und ändert seine Position nicht.	27%
	A	Hinter einem Farbfiler gibt es weniger Licht als vorher, da bestimmte Lichtfarben nicht durchgelassen werden.	4%
	C1	Farbfiler färben das gesamte Licht in die entsprechende Farbe um.	4%
	C2	Farbfiler färben das Licht, wie die Farben im Malkasten.	79%
	C3	Farbfiler haben verschiedenen Farben, aber farbloses Licht geht durch sie nicht durch.	6%
	A	Objekte strahlen ihre Körperfarbe nur dann ab, wenn sie mit Licht bestrahlt werden, das diese Farbe enthält.	4%
	C1	Nur die Farbe des Lichts bestimmt über die Farbe des Objekts.	25%
	C2	Die Farbe des Objekts ist unabhängig von der Farbe des Lichts, mit dem es angestrahlt wird.	25%
	C3	Die Farben des Lichts und der Objekte können sich gegenseitig auslöschen.	22%

Tab.4: Relative Häufigkeiten physikalischer Konzepte und alternativer Vorstellungen bezogen auf die Angaben mit inhaltlicher Deckung.

Aus den vorliegenden Ergebnissen der Hauptzielgruppe können unter anderem folgende Leitlinien für eine Intervention abgeleitet werden:

Für eine Veränderung der Schülervorstellungen zum Sehvorgang wird die Rolle der Streuung des Lichts an allen sichtbaren Oberflächen im Lichtweg Quelle-Objekt-Empfänger unterstrichen (physikalische Komponente). Außerdem werden Lichtreize auf der Netzhaut als Notwendigkeit für alle Lebewesen herausgestellt (biologische Komponente).

Während das zweidimensionale Schattenbild mehrheitlich als ein Bereich aufgefasst wird, indem weniger Licht ankommt als um das Schattenbild herum, ist das Konzept des dreidimensionalen Schattenraumes nicht fest verankert und wird mithilfe von 3D-Darstellungen der Lichtwege gestärkt.

Phänomenologische Beobachtungen am Spiegel werden mit Modelldarstellungen der Lichtwege Objekt-Spiegel-Empfänger vorbereitet. Insbesondere werden Fälle erarbeitet, bei denen Objekt und Empfänger gemeinsam oder voneinander unabhängig vor dem Spiegel bewegt werden.

Die aus dem Kunstunterricht bekannten Primärfarben des Deckfarbkastens nach DIN 5023 (Gelb, Magentarot und Cyanblau) mischen sich nicht zu Rot, Grün und Blau, sondern zu Orange, Grün und Violett [16]. Als Einstieg in das Thema Farben werden deshalb Farbmischungen mit den Farben eines Tintenstrahldruckers behandelt, die einer Farblehre mit physikalischer anstelle von künstlerisch-ästhetischer Herangehensweise folgen. Daran an-

knüpfend wird die Wirkung von mehreren Filtern in den Grundfarben der subtraktiven Farbmischung (Gelb, Magenta und Cyan) auf weißes Licht im Zusammenhang mit dem Konzept der Absorption bestimmter Lichtfarben durch Körperfarben diskutiert (physikalische Komponente). Anschließend werden Farbdarstellungen nicht nur im weißen, sondern auch im Licht verschiedener Farben betrachtet. Die Wahrnehmung der vorhandenen Farb-reize wird im Sinne der additiven Farbmischung interpretiert (biologische Komponente).

6. Evaluation

Neben inhaltlichen Einschätzungen und statistischen Methoden zur Überprüfung der internen Konsistenz lässt sich die Validität eines empirischen Tests mithilfe eines äußeren Kriteriums der Testpersonen bewerten. Üblicherweise werden bei Leistungstests Schulnoten herangezogen. Allerdings basieren gute Noten häufig auf der Lösungskompetenz schematischer Routineaufgaben, während sich in einem abgewandelten Aufgabenkontext wiederum Alltagsvorstellungen offenbaren [17]. Um trotzdem einschätzen zu können, inwieweit der entwickelte Test in der Lage ist, erwartbar unterschiedliche Ausprägungen der Schülervorstellungen abzubilden, wurden neben der Hauptzielgruppe noch andere Lerngruppen getestet. Das sind zum einen Schülerinnen und Schüler der Klassenstufe 6, die noch keinen Physikunterricht hatten und eher auf Alltagsvorstellungen zurückgreifen, und zum anderen Schülerinnen und Schüler der Klassenstufe 8, deren physikalische Vorstellungen ausgeprägter sein sollten (s. Tab. 5).

Klassenstufe (N = 214)		A	B	C	D
6 (n = 26)	M	3.82	.69	10.31	.85
	SD	2.13	.84	1.89	.78
	H	102	18	268	22
	h	25%	4%	65%	5%
7 (n = 165)	M	4.72	.51	8.58	1.21
	SD	1.98	.76	2.25	1.36
	H	778	84	1415	199
	h	31%	3%	57%	8%
8 (n = 23)	M	5.52	.61	7.91	.78
	SD	2.13	.84	2.15	.95
	H	127	14	182	18
	h	37%	4%	53%	5%

Tab.5: Ergebnisübersicht nach Klassenstufe.

Zunächst lässt sich feststellen, dass sich die Anzahl der Items mit inhaltlicher Deckung der Antworten mit der Begründung zwischen den Testgruppen nicht signifikant unterscheidet (Median = 14), Kruskal-Wallis-Test: Chi-Quadrat(2) = 2.741, $p = .254$. Die relative Häufigkeit der Ausgänge A steigt jedoch mit der Klassenstufe an, während sie für Ausgang C entsprechend abnimmt (s. Abb. 5), Kruskal-Wallis-Test für A: Chi-Quadrat(2) = 9.592, $p = .008$ und für C: Chi-Quadrat(2) = 18.269, $p < .001$. Anschließend durchgeführte Post-hoc Tests mit von der Bonferro-

ni-Korrektur angepassten Signifikanzwerten zeigen, dass sich für Ausgang A lediglich die Klassenstufen 6 und 8 signifikant unterscheiden ($z = -3.051$, $p = .007$, $r = .44$). Für Ausgang C unterscheiden sich sowohl die Klassenstufen 6 und 8 ($z = 3.929$, $p < .001$, $r = .56$) als auch 7 und 6 ($z = 3.799$, $p < .001$, $r = .27$) signifikant.

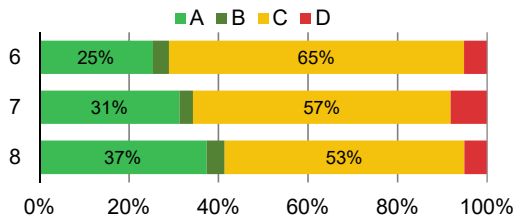


Abb.5: Relative Häufigkeiten der Ausgänge nach Klassenstufe.

Schließlich wurden die Testergebnisse der Schülerinnen und Schüler der Klassenstufen 6 bis 8 mit den Ergebnissen von Studierenden des Lehramts Physik im ersten Fachsemester verglichen (s. Tab. 6). Die relative Häufigkeit der Ausgänge A und C ist bei Studierenden in der Summe etwas höher, aber der Unterschied ist nicht signifikant (Median = 14), Mann-Whitney-U-Test: $U = 5922$, $p = .066$. Die Ausgänge A und C haben dagegen jeweils einen signifikanten Unterschied mit einem großen Effekt (s. Abb. 6), für A: $t(277) = 13.434$, $p < .001$, $d = 1.61$ und für C: $t(277) = -10.677$, $p < .001$, $d = -1.28$.

Bildungseinrichtung (N = 279)		A	B	C	D
Schule (n = 214)	M	4.71	.54	8.71	1.12
	SD	2.04	.78	2.28	1.27
	H	1007	116	1865	239
	h	31%	4%	58%	7%
Universität (n = 65)	M	8.80	.75	5.32	.74
	SD	2.48	.83	2.11	.96
	H	572	49	346	48
	h	56%	5%	34%	5%

Tab.6: Ergebnisübersicht nach Bildungseinrichtung.

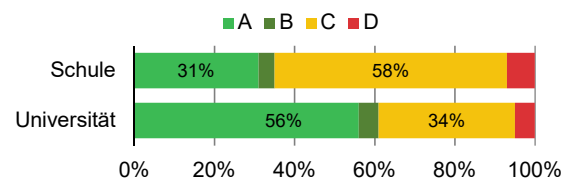


Abb.6: Relative Häufigkeiten der Ausgänge nach Bildungseinrichtung.

Obwohl die Ergebnisse der Studierenden deutlich besser sind als die der Schülerinnen und Schüler, können 40% nur die Hälfte oder weniger Items richtig beantworten und begründen. Insbesondere in den Inhaltsbereichen Spiegel und Farben bleiben auch nach der Schule alternative Vorstellungen erhalten. Des Weiteren zeigen die Häufigkeitsverteilungen, dass das Testinstrument für beide Gruppen geeignet ist, da keine wesentlichen Boden- oder Deckeneffekte auftreten (s. Abb. 7).

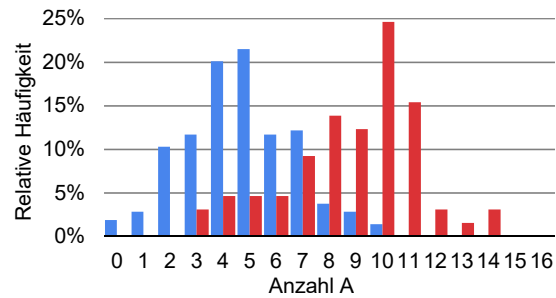


Abb.7: Relative Häufigkeitsverteilungen des Ausganges A für Schülerinnen und Schüler (blau) und Studierende (rot).

7. Zusammenfassung

Der TIA₂ ist als Instrument zur Erhebung von Schülervorstellungen zu verschiedenen Inhaltsbereichen der Anfangsoptik entwickelt worden. Multiple-Choice Tests gelten als in der Durchführung und Auswertung objektive Erhebungsverfahren. Das Auftreten etwaiger Messfehler wird durch den Einsatz von zweistufigen Multiple-Choice Items im Zusammenhang mit dem strengen Kriterium inhaltlicher Deckung der Antwort mit der Begründung eingegrenzt, was die Reliabilität und Validität der Messung verbessern soll.

Nichtsdestotrotz muss man damit rechnen, dass bei vollständig ausgefüllten Testheften etwas 10% der Angaben nicht zu einer fundierten Einschätzung der Schülervorstellungen verwendet werden können. Im Gegenzug verteilen sich die restlichen 90% zu aussagekräftigen Ergebnissen. Diese Frageform senkt allerdings die maximale Anzahl an Items, die mit gleichbleibender Konzentration bearbeitet werden können. Kommt noch die Absicht hinzu, unterschiedliche Konzepte und Vorstellungen in endlicher Zeit zu erheben, dann geht das zulasten der Homogenität der Items und schränkt die interne Konsistenz der Messung ein. Es wurde dennoch der Versuch unternommen, inhaltlich valide Items zu verschiedenen Sachverhalten mithilfe von Literatur und Experten zu erstellen. Die geringe Korrelation der Items untereinander und unterschiedliche Ergebnisse bezüglich verschiedener Inhaltsbereiche lässt sich in diesem Sinne auch als ein Hinweis auf eine Diskriminanzvalidität deuten. Für genauere Untersuchungen bietet es sich deshalb an, nicht die Mittelwerte allein zu interpretieren, sondern die Ergebnisse nach einzelnen Vorstellungen differenziert zu betrachten. Erst dadurch wird die Erfassung der Perspektive der Lernenden im Sinne des Modells der Didaktischen Rekonstruktion möglich. Die vorgestellten Ergebnisse unterstützen die Entwicklung einer an die Zielgruppe angepassten Intervention.

Insgesamt betrachtet stimmen die erhobenen Testergebnisse im Sinne einer Kriteriumsvalidität mit den angenommenen Unterschieden zwischen den Testgruppen verschiedener Alters- bzw. Wissensstufen überein. Bei etwa gleichbleibendem Anteil nicht zuordenbarer Angaben wird ein Anstieg physikalisch gültiger Konzepte deutlich, während alternative

Vorstellungen abnehmen. Weitere Vergleiche zeigen jedoch, dass es durchaus vorkommen kann, dass signifikante Unterschiede nur für bestimmte Ausgänge gemessen werden können. So zeigen beispielsweise die getesteten Schülerinnen und Schüler der Klassenstufe 7 zwar weniger alternative Vorstellungen als in der Klassenstufe 6, aber das muss nicht unbedingt mit der Zunahme physikalischer Vorstellungen zusammenhängen, sondern könnte auch mit der abnehmenden inhaltlichen Deckung der Antworten mit den Begründungen erklärt werden.

In Bezug auf die Messung der Wirksamkeit einer Intervention zur Veränderung von Schülervorstellungen wäre eine Erweiterung des Testinstruments zur Bildung von Subskalen denkbar, die gleiche Vorstellungen in verschiedenen Kontexten messen.

8. Literatur

- [1] R. Duit, „Zur Rolle der konstruktivistischen Sichtweise in der naturwissenschaftsdidaktischen Lehr- und Lernforschung.“, *Zeitschrift für Pädagogik*, Bd. 41, Nr. 6, S. 905–923, 1995.
- [2] A. Teichrew und R. Erb, „Videobasierte Analyse des Lernens mit dynamischen Modellen“, in *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe*, C. Maurer, Hrsg. Universität Regensburg: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018, 2019, S. 464.
- [3] U. Kattmann, R. Duit, H. Gropengießer, und M. Komorek, „Das Modell der Didaktischen Rekonstruktion - Ein Rahmen für naturwissenschaftsdidaktische Forschung und Entwicklung.“, *Zeitschrift für Didaktik der Naturwissenschaften*, Bd. 3, Nr. 3, S. 3–18, 1997.
- [4] S. Reinfried, C. Mathis, und U. Kattmann, „Das Modell der Didaktischen Rekonstruktion. Eine innovative Methode zur fachdidaktischen Erforschung und Entwicklung von Unterricht“, *Beiträge zur Lehrerinnen- und Lehrerbildung*, Bd. 27, Nr. 3, S. 404–414, 2009.
- [5] G. Mey und K. Mruck, „Qualitative Interviews“, in *Qualitative Marktforschung in Theorie und Praxis: Grundlagen, Methoden und Anwendungen*, G. Naderer und E. Balzer, Hrsg. Gabler, 2007, S. 249–278.
- [6] H. Fischler und J. Peuckert, Hrsg., *Concept mapping in fachdidaktischen Forschungsprojekten der Physik und Chemie*. Berlin: Logos-Verl, 2000.
- [7] T. Wilhelm und L. Glatz, „Schülervorstellungen zur Ausdehnung bei Erwärmung“, in *Naturwissenschaftliche Bildung als Grundlage für berufliche und gesellschaftliche Teilhabe*, C. Maurer, Hrsg. Universität Regensburg: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Kiel 2018., 2019, S. 691.
- [8] D. K. Gurel, A. Eryilmaz, und L. C. McDermott, „A Review and Comparison of Diagnostic Instruments to Identify Students’ Misconceptions in Science“, *EURASIA Journal of Mathematics, Science & Technology Education*, Bd. 11, Nr. 5, S. 989–1008, Okt. 2015.
- [9] D. Treagust, „Evaluating students’ misconceptions by means of diagnostic multiple choice items“, *Research in Science Education*, Bd. 16, Nr. 1, S. 199–207, Dez. 1986.
- [10] C. Haagen-Schützenhöfer, „Students’ conceptions on white light and implications for teaching and learning about colour“, *Physics Education*, Bd. 52, Nr. 4, S. 044003, Juli 2017.
- [11] H. Schecker, T. Wilhelm, M. Hopf, und R. Duit, Hrsg., *Schülervorstellungen und Physikunterricht: Ein Lehrbuch für Studium, Referendariat und Unterrichtspraxis*. Springer Spektrum, 2018.
- [12] J. Weber, J. Winkelmann, R. Erb, F. Wenzel, M. Ullrich, und H. Holger, „Ein Fachwissenschaftstest zur geometrischen Optik“, in *Implementati-on fachdidaktischer Innovation im Spiegel von Forschung und Praxis*, C. Maurer, Hrsg. Universität Regensburg: Gesellschaft für Didaktik der Chemie und Physik, Jahrestagung in Zürich 2016, 2017, S. 107.
- [13] C.-C. Chen und M.-L. Lin, „Developing a two-tier diagnostic instrument to assess high school students’ understanding The formation of image by plane mirror“, *Proceedings of the National Science Council, Republic of China. Part D, Mathematics, science and technology education*, Bd. 12, Nr. 3, S. 106–121, 2002.
- [14] H.-E. Chu und D. F. Treagust, „Secondary Students’ Stable and Unstable Optics Conceptions Using Contextualized Questions“, *Journal of Science Education and Technology*, Bd. 23, Nr. 2, S. 238–251, 2014.
- [15] Z. Lakhdar, I. Culaba, V. Lakshminarayanan, J. Maquiling, A. Mazzolini, und D. Sokoloff, *Active learning in optics and photonics: training manual*. UNESCO, 2006.
- [16] „Der Pelikan Farbmischer – Farben am Bildschirm mischen – Pelikan Lehrer-Info“, *Pelikan Deckfarbkasten K12®*, 2019. [Online]. Verfügbar unter: https://www.pelikan.com/pulse/Pulsar/de_DE.CMS.displayCMS.210018./online-farbkasten-farben-online-mischen. [Zugegriffen: 17-Mai-2019].
- [17] H. Schecker und E. Klieme, „Mehr Denken, weniger Rechnen: Konsequenzen aus der internationalen Vergleichsstudie TIMSS für den Physikunterricht“, *Physikalische Blätter*, Bd. 57, Nr. 7–8, S. 113–117, 2001.

¹ Nichtparametrische Tests wurden bei nicht normalverteilten Daten oder kleinen Stichprobengrößen gerechnet.