

Inhaltsvalidität eines Testinstruments zur Erfassung deklarativen Wissens zur Quantenoptik

Philipp Bitzenbauer, Jan-Peter Meyn

Friedrich-Alexander-Universität Erlangen-Nürnberg
philipp.bitzenbauer@fau.de

Kurzfassung

In der empirischen Unterrichtsforschung wird im Kontext von Testentwicklung oft von der *Validierung* eines Testinstruments gesprochen. Die Debatte über das Testgütekriterium *Validität* führte zu einer Verschiebung der Auffassung von Validität. Heute ist weniger von der Validität als Eigenschaft eines Tests die Rede. Vielmehr steht eine valide Testwertinterpretation im Zentrum der Sicherung von Testqualität. Dass ein Testverfahren valide Testwertinterpretation erlaubt, muss argumentativ abgeleitet werden. Die erforderlichen Argumentationsstränge sind dabei nicht standardisierbar. Insbesondere für Forschungsbereiche, in denen es nur wenig belastbare theoretische Vorarbeiten gibt, birgt das Hürden. Hier setzt dieser Artikel an: anhand eines „best-practice“-Beispiels aus der Didaktik der Quantenphysik wird aufgezeigt, wie ein solcher argumentativer Prozess methodisch erfolgen kann. Zunächst wird der Bedarf einer Neuentwicklung eines Testinstruments für die summative Evaluation eines Unterrichtskonzepts zur Quantenoptik begründet. Die Testentwicklung wird beschrieben. Inwiefern eine valide Testwertinterpretation möglich ist, wird argumentativ abgeleitet: Die Ergebnisse von einer Studie zum Lauten Denkens, sowie einer Expertenbefragung werden mit Ergebnissen einer quantitativen Pilotstudie kombiniert.

1. Entwicklungsforschung zum Quantenphysikunterricht

Die Verbesserung des Physikunterrichts ist ordinäres Ziel physikdidaktischer Forschung, die in verschiedenen Bereichen erfolgt. Ein wichtiger Bereich ist dabei die unterrichtsbezogene Entwicklungsforschung [1].

Heute existiert eine fortwährende Tradition unterrichtsbezogener Entwicklungsforschung nicht nur zu Themen des einführenden Physikunterrichts, wie der Mechanik (z.B. [2]) oder der Elektrizitätslehre (z.B. [3]), sondern auch zu fortgeschrittenen Themen der modernen Physik, wie der Quantenphysik [4, 5, 6, 7].

Zur Entwicklungsforschung gehört die Evaluation. Anders als in gängigen Bereichen des Physikunterrichts gibt es in der Quantenphysik keine Standardtestinstrumente zum Erheben von Präkonzepten oder der Lernwirksamkeit von neuen Unterrichtskonzepten [8]. Drei Aspekte scheinen ursächlich:

1. Die entwickelten Unterrichtskonzepte zur Quantenphysik decken eine große inhaltliche Bandbreite ab und stehen überwiegend disjunkt nebeneinander [9].
2. Es herrscht in den (internationalen) Physikcurricula an Schulen kein Konsens über die Quantenphysikthemen, die im Unterricht

zu behandeln sind [10], wengleich jüngste Untersuchungen zumindest Indizien dafür liefern, dass aus Expertensicht ein Konsens über die Key-Items im Quantenphysikunterricht zu finden sein könnte [11].

3. Nicht in allen nationalen Schulcurricula ist Quantenphysik im Schulunterricht überhaupt festgelegt, sondern ist stattdessen für die universitäre Ausbildung vorgesehen.

Der letzte Punkt begründet, warum es zwar Testinstrumente zur Quantenphysik mit der Zielgruppe „Studierende“ gibt, aber nur wenige für Lernende an Schulen. Entwicklungsforschung zum Quantenphysikunterricht erfordert daher üblicherweise die Entwicklung von Testinstrumentarium, das den jeweiligen Forschungsanliegen gerecht wird. In diesem Artikel wird die Entwicklung eines Testinstruments zum Begriffswissen „Quantenoptik“ beschrieben. Dieses Instrument wurde zur summativen Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik entwickelt. Die nachfolgende Infobox gibt einen kurzen Einblick in das Konzept. Details findet man bei [4].

Das Erlanger Unterrichtskonzept zur Quantenphysik für die gymnasiale Oberstufe dient der Einführung in die Quantenphysik in vier Unterrichtsstunden. Das Experiment aus der bahnbrechenden Publikation „Experimental evidence for a photon anticorrelation effect on a beamsplitter“ von Grangier et al. (1986) liegt dem Konzept inhaltlich zu grunde. Anhand von technischen Aspekten und Ergebnissen quantenoptischer Realexperimente gelangen die Lernenden auf direktem Wege zu den Wesenszügen der Quantenphysik.

Abb. 1: Infobox zum Erlanger Unterrichtskonzept zur Quantenoptik. Dessen summative Evaluation erfolgt im mixed-methods-Design. Ein Teil der Untersuchung bezieht sich auf das deklarative Wissen zur Quantenoptik, welches mit dem hier vorgestellten Testinstrument erhoben werden kann.

2. Testinstrumente zur Quantenphysik

Die Forschung zu Lernendenvorstellungen zur Quantenphysik hat mittlerweile eine lange Tradition: Bereits Fischler und Lichtfeldt forderten in ihrer Arbeit von 1992 eine Abkehr von Analogien zur klassischen Physik in der Lehre der Quantenphysik [17]. Sie begründeten das mit der Evaluation eines Unterrichtsgangs zur Quantenphysik und der im Rahmen von dessen Evaluation erhobenen Schülervorstellungen.

Daran anknüpfende Arbeiten, wie [18, 19], griffen einzelne Ideen von Fischler und Lichtfeldt auf, aber es entwickelte sich kein einheitliches Testinstrumentarium. Heute existiert stattdessen eine große Reihe von Testinstrumenten zur Quantenphysik von unterschiedlichem Format, mit unterschiedlichen thematischen Schwerpunkten und vor allem hauptsächlich mit der Zielgruppe „Studierende“.

Test /Autor	Jahr	Fragenformat	Inhaltsschwerpunkt	Zielgruppe
Ireson [18]	2000	Statements mit fünfstufiger Ratingskala	Quantenphänomene und -modelle	Ursprünglich Studierende (auch für Schülerinnen und Schüler geeignet)
Quantum measurement test [20]	2001	Offene Fragen	Messprozess und Zeitentwicklung	Studierende
QMVI [21]	2002	Zweistufige Multiple-Choice-Items	Quantenphysikalischer Formalismus	Studierende
Müller und Wiesner [22]	2002	Statements mit fünfstufiger Ratingskala	Atomvorstellungen, Determinismus, Eigenschaftsbegriff, Heisenberg'sche Unbestimmtheitsrelation	Schülerinnen und Schüler
QPCS [23]	2009	Multiple-Choice-Items	Welle-Teilchen-Dualismus, de-Broglie-Wellenlänge, Doppelspalt-Experiment, Heisenberg'sche Unbestimmtheitsrelation, Wellenfunktion	Studierende (teilweise auch für Schülerinnen und Schüler geeignet)
QMAT [24]	2009	Offene Fragen	Quantenphysikalischer Formalismus	Studierende
QMCS [25]	2010	Multiple-Choice-Items	Quantenphysikalischer Formalismus	Studierende
QMS [26]	2012	Multiple-Choice-Items	Quantenphysikalischer Formalismus	Studierende
QMCA [27]	2015	Multiple-Choice-Items	Quantenphysikalischer Formalismus	Studierende
QME [28]	2019	Zweistufige Multiple-Choice-Items	Materiewellen, Messprozess, Atome und Elektronen	Studierende (teilweise auch für Schülerinnen und Schüler geeignet)
QMFPS [29]	2019	Multiple-Choice	Quantenphysikalischer Formalismus und seine Postulate	Studierende

Tab. 1: Überblick über Testinstrumente zur Quantenphysik.

Die Übersicht zeigt einen deutlichen Bedarf: Die entwickelten Testinstrumente sind zur Evaluation von Lehrkonzepten zur Quantenphysik an Schulen überwiegend ungeeignet. Der Vorstellungsfragebogen von Müller und Wiesner baut auf dem von Ireson auf und wurde im Rahmen der Evaluation des Münchner Unterrichtskonzepts zur Quantenmechanik eingesetzt. Zur Erhebung von Vorstellungen Lernender soll dieser auch im Rahmen der summativen Evaluation Erlanger Unterrichtskonzepts zur Quantenoptik eingesetzt werden.

Allerdings soll mit dem Erlanger Konzept auch ein alternatives Begriffsnetz vermittelt werden. Der Frage, ob Lernende innerhalb dieses Begriffsnetzes konsistent argumentieren können, wird mit Leitfadeninterviews nachgegangen. Notwendige Voraussetzung scheint aber eine grundsätzliche Kenntnis der einzelnen Begriffe der Quantenoptik zu sein. In einem neu entwickelten Fachwissenstest soll daher das deklarative Wissen zur Quantenoptik ökonomisch erhoben werden.

3. Entwicklung eines Fachwissenstests zur Quantenoptik

Eine Präzisierung des Wissensbereichs „Quantenoptik“ aus didaktischer Literatur ist nicht möglich, weil vergleichbare Erhebungen zu vergleichbaren Konzepten noch nicht durchgeführt werden.

Die theoretisch begründete Operationalisierung des Konstrukts „Quantenoptik“ ist demnach nicht möglich und muss sich notwendig eng an dem zu evaluierenden Konzept orientieren. In einem Themengebiet wie der Quantenoptik – oder auch anderen Themen der modernen Physik, die noch weniger untersucht sind - ist es deswegen schwieriger als in „klassischen“ Themenbereichen, eine Substruktur des Wissens der Lernenden zu ermitteln. Es existiert in der fachdidaktischen Forschung kein Standardverfahren dafür, wie man ein solches Gebiet erschließt. Wir wollen daher am Beispiel dieser Testentwicklung ein erfolgsversprechendes Verfahren vorstellen:

Zunächst wurde dem dem neu entwickelten Testinstrument ein Modell zugrunde gelegt, das die drei Teilaspekte „Theoretische Aspekte“, „Photonen“ und „Einzelphotonenexperimente“ enthält. Diese sollten im Testinstrument abgebildet sein. Damit werden physikalische Grundlagen („theoretische Aspekte“), Allgemeines zu Quantenobjekten am Beispiel des Photons („Photonen“) sowie technisch-experimentelle Überlegungen („Einzelphotonenexperimente“) abgefragt.

Die Zuordnung der zentralen Begriffe des Erlanger Unterrichtskonzepts zu diesen Teilaspekten wurde mit Hilfe eines *Blueprints*, vorgeschlagen von [30],

vorgenommen. Dabei handelt es sich um eine Themenmatrix, die Inhaltsbereiche und die in diesen Bereichen zu erreichenden Lernniveaus beinhaltet und gewichtet [31].

Weitere Aspekte der Testentwicklung waren:

- *Entscheidung für ein Aufgabenformat:* Aus ökonomischen Gründen fiel die Wahl auf ein geschlossenes Aufgabenformat in der Variante Single-Choice. Um den Rateeinfluss zu minimieren wurden, wurden zweistufige Items entwickelt [33]: In der ersten Stufe ist nur eine von drei Antwortmöglichkeiten richtig. In der zweiten Stufe sollen die Befragten auf einer fünfstufigen Ratingskala zusätzlich angeben, wie sicher sie sich bei der Beantwortung waren. Ein Punkt wird nur vergeben, wenn die Antwort in Stufe eins richtig ist und sich der Proband mindestens sicher war [32].
- *Finden geeigneter Distraktoren:* Die Qualität von Items im geschlossenen Format hängt von der Qualität der Distraktoren ab, also von den falschen Antwortalternativen. Diese müssen von Wissenden als falsch erkannt, aber von Unwissenden als möglicherweise wahr wahrgenommen werden. Bei der Entwicklung des Testinstruments zur Quantenoptik wurden daher zunächst 21 Items - verteilt auf die verschiedenen Inhaltsbereiche zur Quantenoptik - im offenen Format formuliert. Diese wurden an $N = 23$ Lehramtsstudierende gegeben. Auf diese Art wurde ein erster Testentwurf gewonnen, weil teilweise richtige, oder häufig aufgetauchte falsche Antworten der Studierenden als Distraktoren verwendet wurden.
- *Einheitliche Itemformulierung*

Frage 9. Man verwendet Interferometer in Einzelphotonenexperimenten, um...					
<input type="checkbox"/>	...zu zeigen, dass Licht nicht aus Photonen besteht.				
<input type="checkbox"/>	...zu zeigen, dass Photonen kein Weg zugeordnet werden kann.				
<input type="checkbox"/>	...zu zeigen, dass Photonen Quanteneigenschaften besitzen.				
<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>	<input type="checkbox"/>
Sehr sicher.	Sicher.	Unentschlossen.	Unsicher.	Geraten.	

Abb. 2: Beispiel-Item aus dem Testinstrument zur "Quantenoptik".

4. Pilotstudie zum Testinstrument „Quantenoptik“

Auf Grundlage einer Pilotierung des Testinstruments mit $N = 86$ angehenden Ingenieurstudierenden, wurden ausführliche Itemanalysen durchgeführt. Die Ergebnisse der Pilotstudie finden sich bei [32]. Die wichtigsten Ergebnisse lauten:

- Die interne Konsistenz des Testinstruments ist ausreichend ($\alpha = .78$).
- Signifikante Korrelationen mit den Kriterien „Zeugnisnoten Physik“ und „Interesse

an Quantenphysik“ sprechen für die Kriteriumsvalidität des Tests.

- Die dreifaktorielle Struktur des Testinstruments mit den Skalen „Theoretische Aspekte“, „Einzelfotonenexperimente“ und „Photonen“ lässt sich konfirmatorisch bestätigen ($\chi^2/df = .83$, $p < .001$, $CFI = 1.00$, $RMSEA = .00$, $SRMR = .063$). Die internen Konsistenzen der Subskalen ist zwar niedrig ($.52 \leq \alpha \leq .68$), für sehr kurze Skalen (je 3-4 Items) aber noch akzeptabel [34].

Aufgrund kleiner Stichproben wurde nur mit klassischer Testtheorie ausgewertet. Diese Aspekte (außer der Inhaltsvalidität) wurden bereits in [32] diskutiert. Die Inhaltsvalidität wird argumentativ gestützt. Die dazugehörigen Studien werden nachfolgend diskutiert.

5. Inhaltsvalidität des Testinstruments zur Quantenoptik

Unter *Inhaltsvalidität* versteht man Aussagen darüber, inwiefern das zu erhebende Konstrukt in Gänze durch ein Testinstrument abgedeckt wird. Die dem zugrundeliegende Anforderung an das Testinstrument wird in Anlehnung an [35] so formuliert: „Die Aufgaben umfassen relevante und repräsentative Inhalte und Anforderungen aus der Zieldomäne.“

Um die Validität argumentativ zu stützen wurde daher bereits entwicklungsbegleitend eine „Laute-Denken“-Studie durchgeführt, die zur Optimierung der Validität auf Itemebene beitragen soll. Abschließend wurde ferner eine Expertenbefragung durchgeführt, um die Passung der Items zum Erlanger Unterrichtskonzept, sowie zur Qualität der Distraktoren zu beurteilen [36]. Die folgenden Argumente stützen gemeinsam eine valide Testwertinterpretation für das hier berichtete Testinstrument zur Quantenoptik:

- Signifikante Korrelationen mit Außenkriterien (Kriteriumsvalidität)
- Bestätigung der dreifaktoriellen Skalastruktur mittels konfirmatorischer Faktorenanalyse (Konstruktvalidität)
- Laute-Denken-Studie (Validität auf Itemebene)
- Expertenbefragung (Inhaltsvalidität)

5.1 Lautes Denken

Die Methode des „Lauten Denkens“, manchmal auch „Think Aloud“ oder „Denke-Laut-Methode“, hat ihren Ursprung in der Denkpsychologie [37]. Um zu überprüfen, ob die intendierten fachlichen Zielsetzungen der jeweiligen Items von den Probandinnen und Probanden richtig verstanden werden, macht die Evaluierung eines entwickelten Testinstruments mit Hilfe der Methode des „Lauten Denkens“ Sinn [38, p. 88]. Dadurch werden nämlich die ablaufenden kognitiven

Prozesse von Befragten bei einer Testsituation zugänglich [39, p. S. 224]. Insbesondere bei gebundenen Antwortformaten (hier Single-Choice) erscheint eine Untersuchung der Testaufgaben auf der Ebene der Items sinnvoll, weil die Antwortmöglichkeiten mitunter nicht alle denkbaren Reaktionen der Probandinnen und Probanden abdecken, sich gegenseitig beeinflussen oder nicht der natürlich Antwort der Befragten entsprechen [40]; dies kann die Validität der Erhebungsergebnisse negativ beeinflussen [41].

In der Entwicklung des Fachwissenstest „Quantenoptik“ diente die Methode des „Lauten Denkens“ vorrangig zur Optimierung der Aufgabengüte und damit zur Verbesserung der Validität auf Item-Ebene. Dazu wurden $N = 8$ Schülerinnen bzw. Schüler der gymnasialen Oberstufe, die zuvor das Unterrichtskonzept erlebt haben, gebeten ihre Gedanken während der Auseinandersetzung mit den Test-Items laut zu verbalisieren. Die Durchführung erfolgte mit jedem Teilnehmer bzw. jeder Teilnehmerin separat und die Gespräche wurden aufgezeichnet und transkribiert. Für jedes Item und alle Probanden wurden die folgenden Kategorien (entnommen aus [39]) ausgewertet:

1. Verständlichkeit des Items
 - a. Wurde das Item wie intendiert umschrieben? (IU)
 - b. Wurde die ins Item eingebrachte Barriere wahrgenommen? (BW)
2. Kognitive Prozesse
 - a. Wurde der Entscheidungs- bzw. Antwortfindungsprozess als komplex wahrgenommen? (EK)
 - b. Wurde die Hürde als schwierigkeitszeugend wahrgenommen? (HS)
 - c. War ein eigener Entscheidungsprozess mit Begründung möglich? (EE)
3. Eignung Antwortformat und Itemschwierigkeit
 - a. Sind die Distraktoren authentisch und unterscheidbar? (DU)
 - b. Wie wird die Itemschwierigkeit empfunden? (IS)

Inwieweit diese Kategorien aus Sicht der jeweiligen Probanden für die einzelnen Items erfüllt sind, wurde mittels skalierender Inhaltsanalyse in Anlehnung an [3] ausgewertet:

- „trifft zu“ - zugeordneter Zahlenwert 1,0
- „trifft teils zu“ - zugeordneter Zahlenwert 0,5
- „trifft nicht zu“ - zugeordneter Zahlenwert 0,0

Die Beurteilung wurde von zwei unabhängigen Kodieren vorgenommen ($\kappa = .87$).

Die jeweiligen Itemmittelwerte sind auf zwei Nachkommastellen gerundet und farblich hinterlegt in der untenstehenden Tabelle angegeben. Die so erhaltene Topographie ist für jedes Item von links nach rechts zu lesen (je Item). Items, in deren Zeilen sich rote Zellen häufen, müssen entweder grundlegend überarbeitet oder eliminiert werden.

Item früher (jetzt)	IIU	BW	EK	HS	EE	DA	DU	IS
1 (1)	0,19	0,25	0,44	0,31	0,44	0,25	0,31	0,38
2 (raus)	0,14	0,25	0,50	0,50	0,07	0,14	0,00	0,44
4 (3)	0,19	0,19	0,13	0,13	0,31	0,13	0,06	0,31
5 (4)	0,06	0,06	0,19	0,19	0,13	0,00	0,00	0,57
6 (5)	0,00	0,43	0,29	0,43	0,14	0,00	0,06	0,43
7 (raus)	0,13	0,13	0,31	0,19	0,00	0,25	0,00	0,31
8 (6)	0,19	0,13	0,25	0,13	0,19	0,21	0,00	0,21
9 (raus)	0,19	0,19	0,31	0,25	0,38	0,13	0,25	0,56
10 (raus)	-	-	-	-	-	-	-	-
11 (raus)	0,00	0,19	0,44	0,25	0,13	0,14	0,00	0,57
12 (raus)	0,06	0,31	0,56	0,63	0,13	0,25	0,06	0,81
13 (7)	-	-	-	-	-	-	-	-
14 (8)	-	-	-	-	-	-	-	-
15 (10)	0,06	0,44	0,44	0,56	0,00	0,19	0,00	0,63
16 (ersetzt durch neue 11)	-	-	-	-	-	-	-	-
17 (raus)	0,29	0,36	0,13	0,31	0,25	0,06	0,00	0,25
18 (ersetzt durch neue 12)	-	-	-	-	-	-	-	-
19 (raus)	0,19	0,38	0,31	0,38	0,13	0,38	0,00	0,31
20 (raus)	0,07	0,42	0,50	0,25	0,36	0,25	0,00	0,64
21 (13)	0,00	0,43	0,81	0,88	0,00	0,31	0,06	0,75

Abb. 3: Ergebnisse der "Lautes-Denken"-Studie in der Übersicht. Die Bezeichnung der einzelnen Spalten entspricht der Benennung der Kategorien nach denen die Laute-Denken-Transkripte ausgewertet. Sie sind auf der vorherigen Seite skizziert.

Hierbei ist allerdings zu beachten: es sind auch solche Items aus dem Set eliminiert worden, die beispielsweise schlechte statistische Kennwerte besaßen. Die Items 16 und 18 wurden nicht einbezogen, weil sie bereits vor der „Laute-Denken“-Studie aus inhaltlichen Gründen eliminiert und erst danach durch neue Items ersetzt wurden. Außerdem wurde Item 10 bereits vor Beginn der „Laute Denken“-Erhebung ersatzlos gestrichen. Die Items 13 und 14 (jetzt 7 und 8) wurden bei der „Laute Denken“-Erhebung nicht berücksichtigt, weil die jeweiligen Item-Inhalte bei der Konzeptdurchführung in dieser Studie aus zeitökonomischen Gründen nicht behandelt und somit auch nicht gefragt werden konnten.

Die Kategorien IIU und BW fallen allesamt positiv aus. Kleinere Änderungen an den Itemstämmen in Rücksprache mit den Probandinnen und Probanden führten auf eine Optimierung der Verständlichkeit der Items.

Die intendierten kognitiven Prozesse bei den Befragten wurden im Wesentlichen beobachtet. Es fällt auf, dass der Entscheidungsprozess bei der Mehrzahl der Items als komplex empfunden wurde (EK), dass aber dennoch einige Items nur im mittleren oder gar negativen Bereich (Item 21) liegen. Dies ist nicht erwünscht, denn ein wenig oder gar nicht komplexer Entscheidungsprozess spricht dafür, dass der Proband bzw. die Probandin durch Raten oder Ausschlussprinzip zu einer (mitunter richtigen) Lösung kommen, sodass insbesondere diese Items in der Revision überarbeitet wurden.

Bei der überwiegenden Mehrheit der konzipierten Testaufgaben wurde von den Schülerinnen und Schülern die Antwortfindung fachlich korrekt begründet (Werte $EK \leq 0,56$ mit einem Ausreißer bei Frage 21; Werte $EE \leq 0,44$). Die eingebauten Hürden werden mit einer Ausnahme als mindestens einigermaßen Schwierigkeitserzeugend empfunden.

Besonders wichtig aus Sicht der Fragebogenkonstruktion sind die Kriterien „Distraktoren authentisch“ und „Distraktoren unterscheidbar“ (DA, DU). Dass diese allesamt niedrige Werte vorweisen, spricht dafür, dass die Teilnehmerinnen und Teilnehmer diese als gut bewerteten.

Durch die Studie zum „Laute Denken“ festgestellt werden, dass die Schülerinnen und Schüler bei fehlendem Fachwissen auch andere Lösungsstrategien nutzen als inhaltlich-argumentative, z.B. entschieden sie sich dann aufgrund bestimmter Formulierungen für die eine oder andere Antwortoption. Folglich wurden nach der Auswertung aller Protokolle alle erarbeiteten Items dahingehend geprüft, dass alle Distraktoren in ähnlicher Weise formuliert waren, sich semantisch und formal ähnelten und auf einem Abstraktionsniveau lagen.

5.2 Expertenbefragung

Die Absicherung der Frage nach der Relevanz und Repräsentativität der Testinhalte erfolgte abschließend mit Hilfe einer Befragung von $N = 8$ Wissenschaftlerinnen und Wissenschaftlern aus der Physikdidaktik nach den Kriterien von Jenßen, die bei [42] nachgelesen werden können. Die WissenschaftlerInnen wurden gebeten die Items des Testinstruments auf einer fünfstufigen Rating-Skala hinsichtlich verschiedener Aspekte zu bewerten. Die Ergebnisse wurden, wie für Rating-Skalen empfohlen, mittels Diverging Stacked Bar Charts visualisiert [43]. Eingegangen wurde dabei für jedes Item auf die folgenden Aspekte: *Authentizität der Distraktoren*, *Entscheidender Inhaltsaspekt des Unterrichtskonzepts*, *Item ist gelungen*, *Inhaltsaspekt relevant*.

Neben einer Bewertung der einzelnen Items, wurde außerdem der Test als Ganzes bewertet. Dazu wurde eine Skala bestehend aus vier Items – wieder mit 5-stufiger Ratingskala – entwickelt:

1. Die Items stellen relevante Inhalte des Unterrichtskonzepts dar.
2. Die Inhalte stehen in einem angemessenen Verhältnis zueinander, d.h. die Gewichtung der Inhaltsbereiche ist sinnvoll.
3. Das Testinstrument besitzt eine hohe Passung zum entwickelten Unterrichtskonzept und den damit vermittelten Inhalten.
4. Das Testinstrument deckt wichtige Inhaltsaspekte der Einzelphotonenexperimente ab.

Die interne Konsistenz der Skala liegt bei $\alpha = .86$.

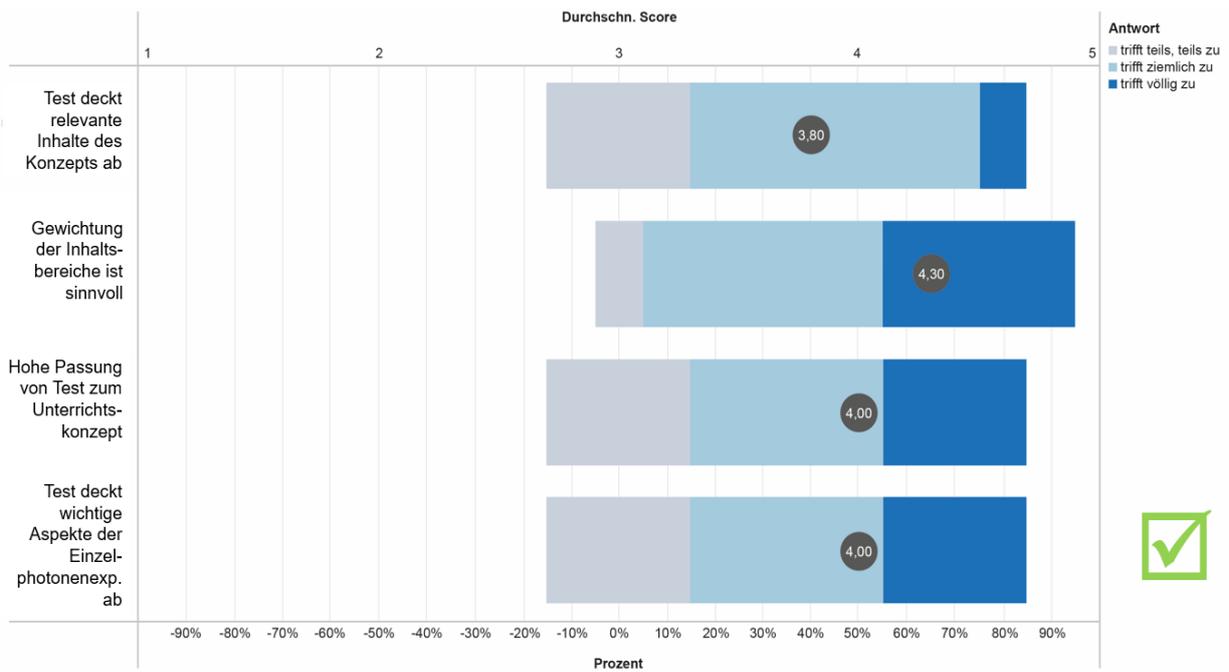


Abb. 4: Diverging-Stacked-Bar-Chart zu den Ergebnissen der Expertenbefragung zum Test als Ganzes.

Es zeigt sich, dass die Expertinnen und Experten eine hohe Einigkeit darüber haben, dass die im Test abgedeckten Inhalte das Konstrukt „Quantenoptik“ angemessen abbilden.

Ein ähnliches Bild zeigte sich auch auf Itemebene: es bestand ein großer Konsens darüber, dass die Items des Fachwissenstests relevante Inhalte zur Quantenoptik abfragen, wie beispielsweise die folgende Grafik zeigt:

Der Inhalt dieses Items ist relevant für den Unterricht der Quantenoptik

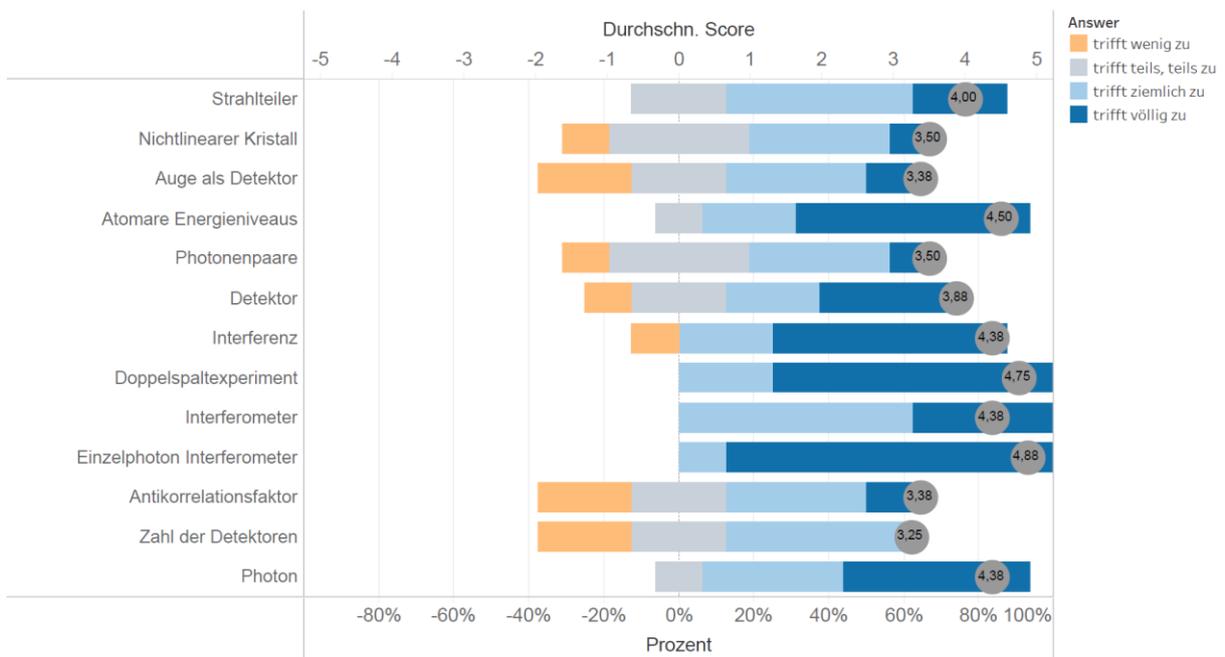


Abb. 5: Diverging-Stacked-Bar-Chart zu dem Aspekt "Inhaltsaspekt relevant" dargestellt für alle Items des Testinstruments in seiner finalen Version, die hier nach ihren abgedeckten Themen dargestellt sind.

6. Zusammenfassung und Ausblick

Wir zeigen in dieser Arbeit, wie die Absicherung einer validen Testwertinterpretation für die Entwicklung von Fachwissenstests ökonomisch aber trotzdem mit umfassenden Argumenten geschehen kann – insbesondere dann, wenn nicht auf umfassende theoretische Vorarbeiten zurückgegriffen werden kann.

Die Methode des lauten Denkens ist entwicklungsbegleitend zu denken und die Expertenbefragung sichert die Qualität zum Ende des Entwicklungsprozesses.

Zusammenfassend rechtfertigen die verschiedenen Argumente eine valide Testwertinterpretation für den Test zur Quantenoptik. Die Ergebnisse aus der „Laute-Denken“-Studie und der Expertenbefragung sichern eine inhaltliche und sprachliche Angemessenheit der Testitems, während die Ergebnisse aus der quantitativen Pilotstudie Argumente für Kriteriums- und Konstruktvalidität bereitgestellt.

Das Testinstrument erscheint daher dazu geeignet deklaratives Wissen zur Quantenoptik, wie es im Erlanger Unterrichtskonzept erhoben zu erfassen. In einer summativen Evaluation wird es in einem Mixed-Methods-Design gemeinsam mit dem Vorstellungfragebogen von Müller und Wiesner, sowie leitfadengestützten Interviews eingesetzt, um einen mehrperspektivischen Blick auf die ablaufenden Lernprozesse nachzeichnen zu können.

7. Literatur

- [1] H. Schecker und B. Ralle, „Naturwissenschaftsdidaktik und Lehrerbildung - Chancen und Risiken aktueller Entwicklungen,“ *Physik und Didaktik in Schule und Hochschule*, Bd. 8(2), pp. 73-83, 2009.
- [2] H. Wiesner, T. Wilhelm, A. Rachel, C. Waltner, V. Tobias und M. Hopf, *Mechanik I: Kraft- und Geschwindigkeitsänderung*, Köln: Aulis, 2011.
- [3] J.-P. Burde, *Konzeption und Evaluation eines Unterrichtskonzepts zu einfachen Stromkreisen auf Basis des Elektronengasmodells*, Berlin: Logos Verlag, 2018.
- [4] P. Bitzenbauer und J.-P. Meyn, „Quantenphysik g²reifbar machen,“ *Plus Lucis*, Bd. 3, pp. 17-21, 2019.
- [5] M. Malgieri, A. De Ambrosis und P. Onorato, „What is Light? From Optics to Quantum Physics Through the Sum over Paths Approach,“ *Conference Paper*, 2015.
- [6] R. Müller, *Quantenphysik in der Schule*, Logos Verlag, 2003.
- [7] A. Kohnle, I. Bozhinova, D. Browne, M. Everitt, P. Al Fomins, P. Kok, G. Kulaitis, M. Prokopas, D. Raine und E. Swinbank, „A new introductory quantum mechanics curriculum,“ *Eur. J. Phys.*, Bd. 35, 2014.
- [8] R. Müller und H. Schecker, „Schülervorstellungen zur Quanten- und Atomphysik,“ in *Schülervorstellungen und Physikunterricht*, H. Schecker, T. Wilhelm, M. Hopf und R. Duit, Hrsg., Heidelberg, Springer Spektrum, 2018, pp. 209-224.
- [9] R. Müller, „Die Quantenphysik im Spannungsfeld zwischen Fachlichkeit, empirischer Forschung und Schulpraxis,“ in *Authentizität und Lernen - das Fach in der Fachdidaktik. Tagungsband zur GDGP-Jahrestagung in Berlin 2015*, Regensburg, Universität Regensburg, 2016, pp. 13-24.
- [10] H. Staderman, E. van den Berg und M. Goedhart, „Analysis of secondary school quantum physics curricula of 15 different countries: Different perspectives on a challenging topic,“ *Phys. Rev. ST Phys. Educ. Res.* 15, 2019.
- [11] K. Krijtenburg-Lewerissa, H. J. Pol, A. Brinkman und W. R. van Joolingen, „Key topics for quantum mechanics at secondary schools: a Delphi study into expert opinions,“ *International Journal of Science Education*, 2018.
- [12] J. Küblbeck und R. Müller, *Die Wesenszüge der Quantenphysik: Modelle, Bilder, Experimente*, Aulis-Verlag Deubner, 2003.
- [13] P. Grangier, G. Roger und A. Aspect, „Experimental evidence for a photon anticorrelation effect on a beam splitter: A new light on single-photon interferences,“ *Europhys. Lett.*, Bd. 1, p. 173, 1986.
- [14] P. Bronner, A. Strunz, C. Silberhorn und J. Meyn, „Interactive screen experiments with single photons,“ *Eur. J. Phys.* (30), pp. 345-353, 2009.
- [15] M. Schreier und Ö. Odag, „Mixed Methods,“ in *Handbuch Qualitative Forschung in der Psychologie*, Wiesbaden, VS Verlag für Sozialwissenschaften | Springer Fachmedien, 2010, pp. 263-277.
- [16] U. Flick, „Triangulation,“ in *Handbuch Qualitative Forschung in der Psychologie*, Wiesbaden, VS Verlag für Sozialwissenschaften | Springer Fachmedien, 2010, pp. 278-289.
- [17] H. Fischler und M. Lichtfeldt, „Modern physics and students' conceptions,“ *International Journal of Science Education*, Bd. 14:2, pp. 181-190, 1992.
- [18] G. Ireson, „A multivariate analysis of undergraduate physics students' conceptions of quantum phenomena,“ *Eur. J. Phys.*, Bd. 20, pp. 193-199, 1999.
- [19] K. Mannila, I. T. Koponen und J. A. Niskanen, „Building a picture of students' conceptions of wave-particle-like properties of quantum

- entities," *Eur. J. Phys.*, Bd. 23, pp. 45-53, 2002.
- [20] C. Singh, „Student understanding of quantum mechanics," *Am. J. Phys.*, Bd. 69, pp. 885-895, 2001.
- [21] E. Cataloglu und R. W. Robinett, „Testing the development of student conceptual and visualization understanding in quantum mechanics through the undergraduate career," *Am. J. Phys.*, Bd. 70, pp. 238-251, 2002.
- [22] R. Müller und H. Wiesner, „Teaching quantum mechanics on an introductory level," *Am. J. Phys.*, Bd. 70, pp. 200-209, 2002.
- [23] S. Wutiprom, M. D. Sharma, I. D. Johnston, R. Chtaree und C. Soankwan, „Development and Use of a Conceptual Survey in Introductory Quantum Physics," *International Journal of Science Education*, Bd. 31, pp. 631-654, 2009.
- [24] S. Goldhaber, S. J. Pollock, M. Dubson, P. Beale und K. K. Perkins, „Transforming Upper-Division Quantum Mechanics: Learning Goals and Assessment," *AIP Conf. Proc. 1179*, pp. 145-148.
- [25] S. B. McKagan, K. K. Perkins und C. E. Wieman, „Design and validation of the quantum mechanics conceptual survey," *Phys. Rev. ST Phys. Educ.*, Bd. 6, p. 020121, 2010.
- [26] G. Zhu und C. Singh, „Surveying students' understanding of quantum mechanics in one spatial dimension," *Am. J. Phys.*, Bd. 85, p. 252, 2012.
- [27] H. R. Sadaghiani und S. J. Pollock, „Quantum mechanics concept assessment," *Phys. Rev. ST Phys. Educ.*, Bd. 11, p. 010110, 2015.
- [28] U. S. Uccio, A. Colantonio, S. Galano, I. Marzoli, F. Trani und I. Testa, „Design and validation of a two-tier questionnaire on basic aspects of quantum mechanics," *Phys. Rev. ST Phys. Educ. Res.*, Bd. 15, p. 010137, 2019.
- [29] E. Marshman und C. Singh, „Validation and administration of a conceptual survey on the formalism and postulates of quantum mechanics," *Phys. Rev. ST Phys. Educ. Res.*, Bd. 15, p. 020128, 2019.
- [30] T. L. Flateby, „A Guide for Writing and Improving Achievement Tests," University of South Florida, 2013.
- [31] R. Krebs, „Multiple-Choice Fragen? Ja, aber richtig," Medizinische Fakultät; Institut für Medizinische Lehre IML; Abteilung für Assessment- und Evaluation AAE, Bern, 2008.
- [32] P. Bitzenbauer und J.-P. Meyn, „Evaluation eines Unterrichtskonzepts zur Quantenoptik mit Einzelphotonenexperimenten - Ergebnisse einer Pilotstudie," in *Naturwissenschaftliche Kompetenzen in der Gesellschaft von morgen. Tagungsband zur GDGP-Jahrestagung in Wien 2019*, Universität Duisburg-Essen, 2020.
- [33] C. Erfmann und R. Berger, „Mehrstufige Multiple-Choice-Aufgaben zur Erfassung qualitativer Aspekte der elektromagnetischen Induktion," *Tagungsband zur DPG-Frühjahrstagung in Wuppertal*, 2015.
- [34] S. Blömeke, „Messung des fachbezogenen Wissens angehender Mathematiklehrkräfte," in *Professionelle Kompetenz angehender Lehrerinnen und Lehrer. Wissen, Überzeugungen und Lerngelegenheiten deutscher Mathematikstudierender und -referendare*, Münster, Waxmann Verlag, 2008, pp. 49-88.
- [35] M. Dickmann, Messung von Experimentierfähigkeiten. Validierungsstudien zur Qualität eines computerbasierten Testverfahrens, Berlin: Logos Verlag, 2016.
- [36] H. Theyßen, „Methodik von Vergleichsstudien zur Wirkung von Unterrichtsmedien," in *Methoden in der naturwissenschaftsdidaktischen Forschung*, D. Krüger, I. Parchmann und H. Schecker, Hrsg., Berlin Heidelberg, Springer-Verlag, 2014, pp. 67-79.
- [37] R. Weidler und A. C. Wagner, „Verbale Daten - Eine Einführung in die Grundlagen und Methoden der Erhebung und Auswertung," Weinheim und Basel, 1982, pp. 81-103.
- [38] C. Kulgemeyer, Physikalische Kommunikationskompetenz. Modellierung und Diagnostik, Berlin: Logos Verlag, 2010.
- [39] C. Meinhardt, Entwicklung und Validierung eines Testinstruments von (angehenden) Physiklehrkräften in physikdidaktischen Handlungsfeldern, Berlin: Logos Verlag, 2018.
- [40] C. Schnell, „Lautes Denken als qualitative Methode zur Untersuchung der Validität von Testitems," *Zeitschrift für ökonomische Bildung*, pp. 26-49, 2016.
- [41] J. Rost, Lehrbuch Testtheorie - Testkonstruktion, Verlag Hans Huber, 2004.
- [42] L. Jenßen, S. Dunekacke und S. Blömeke, „Qualitätssicherung in der Kompetenzforschung: Empfehlungen für den Nachweis von Validität in Testentwicklung und Veröffentlichungspraxis," *Zeitschrift für Pädagogik*, Bd. 61 (Beiheft), pp. 11-31, 20015.
- [43] N. B. Robbins und R. M. Heiberger, „Plotting Likert and Other Rating Scales"
- [44] J. R. Anderson, „ACT, a simple theory of complex cognition," *American Psychologist*, Bd. 51(4), pp. 355-365.