

Entwicklung von Testinstrumenten zu deklarativem Wissen und Vorstellungen zur Quantenphysik in der Mittelstufe

Carsten Albert^{*+}, Gesche Pospiech⁺

* Leibniz-Institut für Festkörper- und Werkstoffforschung Dresden
+ Technische Universität Dresden, Professur für Didaktik der Physik
c.albert@ifw-dresden.de, gesche.pospiech@tu-dresden.de

Kurzfassung

Quantentechnologien bieten zahlreiche Kontexte zur Vermittlung konzeptueller quantenphysikalischer Aspekte. Das ermöglicht nicht zuletzt einen früheren Einstieg in die Thematik. In diesem kontextuellen Rahmen wurde ein Spin-First-Lehrkonzept zur Vermittlung konzeptueller und phänomenologischer Aspekte moderner Quantenphysik in Klassenstufe 9 entwickelt.

Die summative Evaluation des Konzeptes untersucht kognitive und affektive Lernendenmerkmale sowie Lehrkräfteeinschätzungen. Zur Untersuchung der Lernwirksamkeit des Konzeptes wurde ein Fragebogen zum erworbenen deklarativen Wissen entwickelt. Zudem entstand ein Vorstellungfragebogen, der die Erhebung des Ausprägungsgrades quantenphysikalischer Vorstellungen, Denkweisen und Konzepte fokussiert. Dieser Beitrag gibt einen Einblick in die Entwicklung und diskutiert die Sicherstellung von Testgütekriterien der beiden geschlossenen Testinstrumente.

1. Quantenphysik in der Schule

Quantenphysik erhält, nicht zuletzt durch die derzeitigen großen Entwicklungssprünge im Bereich von Quantentechnologien der zweiten Generation, auch im schulischen Bereich eine wachsende Aufmerksamkeit (Müller, 2023). Zwar ist sie im internationalen Vergleich bereits verhältnismäßig lange im deutschen Schulunterricht etabliert, dabei allerdings praktisch ausschließlich als Thema in der gymnasialen Oberstufe (Stadermann et al., 2019).

Die dabei etablierte quasi-historische Vorgehensweise wird bereits seit längerer Zeit kritisiert (siehe z. B. Müller, 2005). Dagegen bieten Quantentechnologien zahlreiche Kontexte, die insbesondere auf moderne und konzeptuelle Aspekte der Quantenphysik fokussieren – ein Trend, der sich mittlerweile auch in den Bildungsstandards für die gymnasiale Oberstufe niederschlägt (Müller, 2023).

Zahlreiche Autoren sprechen sich in Hinblick auf diese Entwicklungen und Ergebnisse für einen Spin-First-Ansatz aus (Stadermann et al., 2019). Ein solcher Ansatz betont die zentrale Rolle der Superposition in der Quantenphysik und erarbeitet konzeptuelle Aspekte anhand exemplarischer Zweizustandssysteme (Stadermann et al., 2019).

Eine zentrale Argumentation für diesen Ansatz ist die Einfachheit der betrachteten Systeme, verbunden mit ersten Erkenntnissen, dass Schülerinnen und Schüler solche Systeme auch phänomenologisch gut erfassen und dadurch formal besser analysieren können (Müller, 2016; Micheli & Stefanel, 2022).

Auch im Bereich der Hochschullehre gibt es für einen solchen Ansatz bereits positive Erfahrungswerte (Sadaghiani, 2016).

2. Entwicklung eines Mittelstufen-Lehrkonzeptes

Ein Spin-First-Ansatz eröffnet auch die Möglichkeit, die Quantenphysik unter einem konzeptuellen und phänomenologischen Fokus bereits in der Sekundarstufe 1 einzuführen, da er auf Seiten der Lernenden die Voraussetzungen an Vorwissen und mathematischen Kompetenzen für einen entsprechenden Lehrgang absenkt. So ist unter anderem eine fachlich saubere qualitative Diskussion konzeptueller Aspekte der Quantenphysik möglich, die vollständig auf den Wellenaspekt zu verzichtet und so die damit verbundenen Schülervorstellungen und Lernschwierigkeiten (siehe z. B. Körhasan & Miller, 2020) umgeht.

Damit kann insbesondere eine junge Zielgruppe in den Fokus fachdidaktischer Forschung rücken, zu der es bisher nur erst Hinweise gibt, dass auch hier eine Vermittlung zentraler Inhalte moderner Physik möglich ist (vgl. Stadermann et al., 2019).

Im Rahmen eines Entwicklungsforschungsprojektes wurde ein Lehrkonzept für die Klassenstufe 9 konzipiert, das auf einem Spin-First-Ansatz aufbaut und hinsichtlich der physikalischen Grundprinzipien des Quantencomputers eine quantentechnologisch geprägte Kontextorientierung aufweist.

Das Lehrkonzept soll innerhalb dieses Kontextes einen verhältnismäßig frühen qualitativen Einblick in moderne Ideen und konzeptuelle Aspekte der Quantenphysik geben, etwa zu Prinzipien wie Superposition und Verschränkung.

Das Konzept wurde im Design-Based-Research-Ansatz mit den zugehörigen Lehrmaterialien in mehreren Schritten entwickelt und wird im Rahmen einer Feldstudie abschließend summativ evaluiert. In diesem Rahmen kommen vorhandene, aber auch neu entwickelte Testinstrumente zum Einsatz.

2.1. Entwicklungsschritte und Methodik

Zunächst wurden auf Grundlage bisheriger Erkenntnisse aus der fachdidaktischen Forschung Design-Prinzipien für den Kurs abgeleitet und ein erster Entwurf für ein Vermittlungskonzept erarbeitet. Dieser Konzeptentwurf wurde anschließend in einem ersten empirischen Schritt anhand von Akzeptanzbefragungen (leitfadenbasierte Einzelinterviews mit Interventionsphasen) als Laborstudie erprobt und angepasst. Für Details hierzu siehe Albert und Pospiech (2023). Es folgte eine Pilotierungsphase, in der

- 1) das Konzept unter realistischen Bedingungen im Feld erprobt und verbessert wurde, sowie
- 2) Erhebungsinstrumente für die summative Evaluation entwickelt wurden.

2.2. Einblick in das Lehrkonzept

Das Lehrkonzept adressiert in erster Linie Schülerinnen und Schüler der Klassenstufe 9. Ausgehend von verschiedenen empirischen Befunden und theoretischen Überlegungen wurden wie beschrieben, unter Beachtung der praktischen Rahmenbedingungen, Design-Prinzipien erarbeitet, die dem gesamten Lehrkonzept zugrunde liegen. Dazu zählen insbesondere:

- Aufbau auf Zweizustandssystemen
- Verzicht auf das Konzept von Wellen
- klare Begriffsbildung und Sprachsensibilität
- Phänomen- und Konzeptorientierung
- sprachliche und inhaltliche Abgrenzung zwischen Quantenphysik und klassischer Physik
- Diskussion des Zusammenhangs von Quantenphysik und klassischer Physik
- Gamification zum Einstieg in die Quantenphysik

Im Zentrum des Konzeptes stehen vier Wesenszüge (Namensgebung orientiert an Küblbeck und Müller (2007)): Superposition, Messprozess, statistischer Charakter und Verschränkung.

Zeitlich umfasst der entwickelte Kurs insgesamt etwa sechs bis sieben Doppelstunden. Details zur Sachstruktur des Konzeptes sind ebenfalls in Albert und Pospiech (2023) zu finden.

3. Entwicklung von Testinstrumenten

Die Entwicklung von Tests, die der Beantwortung von Forschungsfragen dienen, ist eine wichtige und komplexe Aufgabe fundierter fachdidaktischer Forschung (Schmiemann & Lücken, 2014). Voraussetzungen sind präzise Forschungsfragen sowie die Vorlagerung von Theoriearbeit und zahlreicher Entwicklungs- und Analyseschritte (vgl. z. B. Riese & Reinhold, 2014; Schmiemann & Lücken, 2014).

Insbesondere müssen solche Tests den Standards und Kriterien der Wissenschaftlichkeit entsprechen, um eine hohe Testqualität sicherzustellen. Dazu können etwa die Hauptgütekriterien für Tests betrachtet werden (Döring, 2023).

Der Beitrag beschäftigt sich im Folgenden mit der Entwicklung von Testinstrumenten, die im Rahmen der summativen Evaluation des beschriebenen Konzeptes eingesetzt werden, und unter Beachtung der angegebenen wissenschaftlichen Standards entwickelt wurden.

Dafür werden zunächst ausgewählte Forschungsfragen der summativen Evaluation vorgestellt, um darauf aufbauend die Entwicklung zweier Testinstrumente zur Beantwortung dieser Forschungsfragen zu skizzieren.

3.1. Forschungsfragen

Im Zuge der summativen Evaluation stellen sich unter anderem folgende zentrale Forschungsfragen in Hinblick auf kognitive Lernendenmerkmale:

- FF1: Wie lernwirksam ist das Konzept und welche Verständnisschwierigkeiten treten auf?
- FF2: Sind bei den Lernenden nach dem Unterricht quantenphysikalisch geprägte Vorstellungen und Denkweisen erkennbar?

3.2. Testinstrument zum deklarativen Wissen

Zur Beantwortung von FF1 kommt unter anderem ein geschlossener Fragebogen im Single-Choice-Format für Schülerinnen und Schüler zum Einsatz. Dieser Fragebogen wird bezüglich Zielstellung, Entwicklung und Analyse im Folgenden kurz vorgestellt.

3.2.1. Zielstellung

Durch den Fragebogen soll die Lernwirksamkeit des Konzeptes unter Praxisbedingungen evaluiert werden. Unter Lernwirksamkeit wird innerhalb dieser Studie das erworbene deklarative (domänenspezifische) Wissen zu Themen, grundlegenden Begriffen und Ideen der Quantenphysik im Rahmen des Lehrkonzeptes verstanden. Dabei umfasst deklaratives Wissen ein Spektrum aus Sach-, Fakten- und Begriffswissen (Lenske et al., 2015).

Evaluationsergebnisse, die sich auf Grundlage dieses Fragebogens ergeben, können auch Hinweise auf Lernschwierigkeiten geben. Jedoch geben auch andere, hier nicht vorgestellte Erhebungen, Einblick in diesen Teil der Forschungsfrage (insbesondere Lehrerfragebogen und Schülerinterviews).

Ziel des Fragebogens zum deklarativen Wissen ist es somit insbesondere nicht, quantenphysikalische Vorstellungen, Transfer- und Argumentationsfähigkeiten oder ein generelles Verständnis zur Quantenphysik zu erheben (vgl. Bitzenbauer, 2020).

3.2.2. Entwicklungsschritte

Der Fragebogen wurde grundlegend nach den Empfehlungen von Jonkisz et al. (2011) und Bühner (2010) konstruiert. Zudem orientiert er sich, insbesondere auch in der in der Entwicklungsmethodik, in großen Teilen an Bitzenbauer (2020).

Methodologisch wurde der Fragebogen linear entsprechend der nachfolgend dargelegten Vorgehensweise entwickelt:

- 1) Konstrukt und Curriculumsbezug
- 2) offene Fragen/Lerntagebücher
- 3) Entwicklung einer vorläufigen Testversion
- 4) Lautes Denken
- 5) Überarbeitung
- 6) Pilotierung und psychometrische Prüfung
- 7) Überarbeitung
- 8) Expertenvalidierung

Der Ansatz zum Konstrukt sowie Einblicke in die Laut-Denken-Studie und abschließende psychometrische Kennwerte werden im Folgenden gegeben.

3.2.3. Konstrukt

Entsprechend der Zielstellung bezweckt der Fragebogen die Erhebung von Sach-, Fakten- und Begriffswissen zentraler Begriffe und Ideen aus dem Unterrichtskonzept. Die Festlegung dieser Begriffe erfolgt auf Grundlage eines Blueprints nach Flateby (2013). Konkret wurde hierfür der den Lehrkräften vorgegebene Stoffverteilungsplan zugrunde gelegt, in dem die als zentral geltenden Begriffe und Aspekte vermerkt sind. Durch die Zuordnung der Items im finalen Format des Fragebogens zu den Begriffen und Ideen des Blueprints wird curriculare Validität sichergestellt.

3.2.4. Lautes Denken

Auf Grundlage des Konstruktes und Blueprints entstand ein erster Testentwurf, dessen Distraktoren aus Schülerantworten in Lerntagebüchern (eingesetzt in der ersten Pilotierungsklasse) abgeleitet wurden. Zur Sicherstellung inhaltlicher Validität insbesondere auf Itemebene wurde eine an Bitzenbauer (2020) orientierte Laut-Denken-Studie mit $N = 8$ Schülerinnen und Schülern durchgeführt, welche das Lehrkonzept zuvor im Rahmen der Pilotierung kennengelernt und durchlaufen haben.

Unter der Methode „Lautes Denken“ wird „[...] das gleichzeitige laute Aussprechen von Gedanken bei der Bearbeitung einer Aufgabe [...]“ (Knoblich & Öllinger, 2006, zitiert nach Schnell, 2016) verstanden. Im vorliegenden Fall wurde der erste Entwurf des Fragebogens zum deklarativen Wissen von den Teilnehmenden bearbeitet. Auf Grundlage dieser Form des Interviews sollte geklärt werden, ob

- a) die Items domänenkonform interpretiert werden,
- b) die kognitiven Hürden in den Items wahrgenommen werden,
- c) die Items sprachlich verständlich formuliert sind,
- d) die entwickelten Distraktoren als hinreichend plausibel und unterscheidbar wahrgenommen werden, sowie
- e) jeweils eine begründete Entscheidung für eine Antwort möglich ist.

Die Auswertung der Interviews erfolgte durch eine skalierende Inhaltsanalyse nach Mayring (2010) mit einem deduktiven Kategoriensystem sowie einer

daraus abgeleiteten gerichteten Exploration des Interviewmaterials zur Ableitung von Anpassungen an den Items.

Zu jedem der genannten Aspekte bzw. Fragestellungen a) – e) kam folgendes dreistufiges ordinales Kategoriensystem, genauer definiert in einem Kodiermanual, zum Einsatz:

- trifft zu (kodiert mit 0; grün)
- trifft teilweise zu (kodiert mit 0,5; gelb)
- trifft nicht zu (kodiert mit 1; rot)

Nach Kodierung aller Interviews und Items ergaben sich für die 16 finalen Items die in Tabelle 1 dargestellten arithmetischen Mittelwerte zu den jeweiligen Fragestellungen a) – e).

Tab. 1: Ergebnisse der Laut-Denken-Studie.

Item	Inhalt	Hürde	Verständlichkeit	Distraktoren	Begründung
1	0,00	0,19	0,25	0,38	0,19
2	0,00	0,38	0,00	0,56	0,00
3	0,00	0,31	0,00	0,63	0,13
4	0,13	0,31	0,00	0,25	0,31
5	0,44	0,13	0,00	0,38	0,38
6	0,13	0,06	0,00	0,19	0,25
7	0,13	0,00	0,00	0,19	0,06
8	0,00	0,25	0,06	0,25	0,06
9	0,31	0,06	0,06	0,31	0,06
10	0,13	0,19	0,00	0,44	0,00
11	0,25	0,00	0,13	0,19	0,06
12	0,13	0,00	0,00	0,31	0,06
13	0,06	0,56	0,00	0,56	0,06
14	0,00	0,00	0,19	0,38	0,06
15	0,00	0,44	0,00	0,69	0,19
16	0,00	0,25	0,25	0,31	0,06

12,5% der Daten wurden von einem unabhängigen Rater mit $\kappa = 0,71$ zweikodiert.

Diese Mittelwerte dienen zunächst der übersichtlichen Zusammenfassung und Dokumentation der Kodiererergebnisse und geben so für die untersuchten Aspekte Hinweise und Nachweise für die Validität auf Itemebene. Auf Grundlage dieser Kodierung fand zur Weiterentwicklung im nächsten Schritt eine tiefere qualitative Exploration der Interviewaussagen und eine weitere Runde offener Fragen bei zusätzlichen Schülerinnen und Schülern aus den Pilotierungsklassen statt. Auf Grundlage dieser Ergebnisse und Antworten wurde der Fragebogen, insbesondere Formulierungen und Distraktoren, weiter angepasst.

3.2.5. Finales Format und psychometrische Eigenschaft

Die finale Ausgestaltung des Fragebogens besteht aus 16 Items im Single-Choice-Format. Jedes Item umfasst einen Attraktor sowie zwei Distraktoren. Hinterlegt ist bezüglich der Wahl der korrekten (1 Punkt) oder falschen (0 Punkte) Antwort eine dichotome Skala. Der Testwert entspricht der Gesamtpunktzahl. Auf Grundlage von Erhebungen in insgesamt fünf finalen Klassen/Kursen der Pilotierung wurden psychometrische Kennzahlen zu diesen Items

bestimmt. Diese Werte basieren auf $N = 84$ vollständig ausgefüllten Fragebögen.

14 der 16 finalen Items liegen innerhalb der von Döring (2023) vorgeschlagenen Grenzen für Itemschwierigkeit und/oder Trennschärfe. Die Items 9 und 15 wurden trotz leichter Unter- oder Überschreitung dieser Grenzen beibehalten, da sie vor allem auf inhaltlich zentrale Aspekte rekurren (curriculare Validität) und zugleich der Vermeidung eines Bodeneffekts dienen können.

Darüber hinaus sind die Items zweistufig gestaltet; bei jedem Item wird zur Ratekorrektur unmittelbar die Antwortsicherheit der Teilnehmenden mit erhoben (vgl. Bitzenbauer, 2020).

Die interne Konsistenz des Fragebogens ergibt sich insgesamt zu $\alpha = 0,80$. Tabelle 2 zeigt psychometrische Kennwerte für alle Items des finalen Fragebogens unter Einbeziehung einer Item-Ratekorrektur.

Tab. 2: Psychometrische Kennzahlen. Strichlinien markieren bevorzugte Bereiche nach Döring (2023) (Itemschwierigkeit: [0,20;0,80], Itemtrennschärfe: [0,30;1,00]).

Item	Schwierigkeit	Trennschärfe
1	0,71	0,60
2	0,61	0,51
3	0,54	0,38
4	0,58	0,50
5	0,57	0,43
6	0,51	0,31
7	0,72	0,40
8	0,22	0,31
9	0,29	0,26
10	0,59	0,30
11	0,67	0,60
12	0,23	0,30
13	0,72	0,52
14	0,68	0,35
15	0,86	0,19
16	0,48	0,41
Durchschnitt	0,56	0,40

3.2.6. Testgütekriterien

Tabelle 3 zeigt eine Zusammenstellung zum Nachweis der Testgütekriterien des Fragebogens zum deklarativen Wissen. Korrelationskoeffizienten zur Sicherstellung der Kriteriumsvalidität basieren auf Antworten und Angaben der gleichen Stichprobe.

Die Expertenvalidierung ist zum Zeitpunkt der Einreichung des Beitrages noch nicht abgeschlossen. Es zeichnet sich jedoch auch hier, basierend auf den vorläufigen Daten, eine Bestätigung der Validität auf dieser Ebene ab.

Tab. 3: Zusammenfassung der Testgütekriterien.

Kriterium	Fragebogen zum deklarativen Wissen
Objektivität	geschlossene Form ✓
Reliabilität	$\alpha = 0,8$ ✓
Validität	Inhaltvalidität global (✓) <i>Lautes Denken + Expertenvalidierung</i>
	Inhaltvalidität auf Itemebene ✓ <i>Lautes Denken</i>
	Kriteriumsvalidität ✓ <i>Punktzahl - Note:</i> $r = 0,59, p < 0,001$ <i>Punktzahl - Interesse an Quantenphysik:</i> $r = 0,44, p < 0,001$
	curriculare Validität ✓ <i>Orientierung am Stoffverteilungsplan</i>

3.3. Testinstrument zu quantenphysikalischen Vorstellungen und Denkweisen

Zur Beantwortung von FF2 kommt unter anderem ein geschlossener Fragebogen für Schülerinnen und Schüler mit fünfstufiger Ratingskala zum Einsatz. Dieser Fragebogen wird bezüglich Zielstellung, Ansatz, Entwicklung und Validierung im Folgenden kurz vorgestellt.

3.3.1. Zielstellung

Mithilfe des Fragebogens sollen Einblicke in den Umfang quantenphysikalisch geprägter Vorstellungen und in das konzeptuelle Verständnis der Schülerinnen und Schüler nach Abschluss des Unterrichts gewonnen werden. Die Ergebnisse dieses Fragebogens werden durch begleitende leitfadensbasierte Interviews im Rahmen der Hauptstudie trianguliert.

3.3.2. Ansatz des Fragebogens

Der Vorstellungsfragebogen orientiert sich an Müller (2003) und Bitzenbauer (2020) und besitzt in seiner finalen Form zwölf Items, die jeweils eine physikalische Aussage treffen. Diese Aussagen sind jeweils entweder geprägt von einer quantenphysikalischen oder einer klassischen Denkweise. Anhand der Ratingskala kann jeder Aussage auf fünf Stufen zugestimmt, eine neutrale Haltung eingenommen, oder sie abgelehnt werden.

Jedes Item hat, entsprechend der sich aus dem Lehrkonzept ergebenden Relevanz, eine Wichtung hinterlegt (einfach oder doppelt gewichtet). Zudem besitzt jedes Item ein hinterlegtes Vorzeichen, entsprechend seiner inhaltlichen Ausrichtung:

- positiv (+) entspricht einer quantenphysikalisch adäquaten Denkweise
- negativ (−) entspricht nicht einer quantenphysikalisch adäquaten Denkweise

Damit ergibt sich für jedes Item i ein Gewichtungsfaktor $g_i \in \{-2, -1, +1, +2\}$.

Bezeichnet man für jedes Item i die von den Teilnehmenden gewählte Bewertung mit v_i ($v_i \in \{1, \dots, 5\}$), dann lässt sich für jeden vollständig ausgefüllten Fragebogen ein Gesamtvorstellungsindex entsprechend

$$C = \frac{\sum_{i=1}^n g_i \cdot 50(v_i - 3)}{\sum_{i=1}^n |g_i|} \quad \{1\}$$

mit $C \in [-100, +100]$ berechnen.

Dabei laufen für $C = -100$ alle Antworten der quantenphysikalisch adäquaten Denkweise völlig zu wider, 0 bezeichnet eine indifferente Haltung, und für $C = 100$ entsprechen alle Antworten vollkommen einer adäquaten quantenphysikalischen Denkweise.

3.3.3. Entwicklungsschritte

Grundlage für den Fragebogen waren die Vorstellungsfragebögen von Müller (2003) und Bitzenbauer (2020).

Darauf aufbauend wurde der Fragebogen methodologisch linear entsprechend der folgenden Vorgehensweise entwickelt:

- 1) Konzeptmodell festlegen
- 2) Entwicklung einer vorläufigen Testversion
- 3) Validierungsinterviews I
- 4) Überarbeitung
- 5) Validierungsinterviews II
- 6) Überarbeitung
- 7) Expertenvalidierung

Ein Kurzüberblick zum Konzeptmodell sowie zur zweiten Runde der Validierungsinterviews werden im Folgenden gegeben.

3.3.4. Konzeptmodell

Die Items wurden inhaltlich auf das Curriculum des Lehrkonzeptes abgestimmt. Der Vorstellungsfragebogen ist damit domänenspezifisch im Kontext des Lehrkonzeptes. Insbesondere werden Vorstellungen zu den dort vorgestellten Wesenszügen, sowie zum Grundprinzip des Indeterminismus erfasst. Darüber hinaus sind Aspekte zur Dekohärenz eingeschlossen. Abbildung 1 zeigt einen Überblick zu den entsprechenden Konzepten.

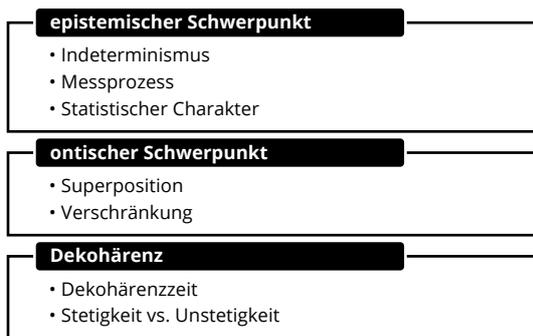


Abb. 1: Konzepte im Vorstellungsfragebogen.

3.3.5. Validierungsinterviews

Zur Validierung des Fragebogens wurden, ergänzend zum Vorgehen von Müller (2003) und Bitzenbauer (2020), leitfadensbasierte Einzelinterviews geführt.

Jedem Item des Vorstellungsfragebogens ist in diesen Interviews eine entsprechende Frage im Interviewleitfaden zugeordnet. Die Antwort der Teilnehmenden im Interview wurde fünfstufig ordinalskaliert kodiert (adäquate ... inadäquate Vorstellungsmuster). Damit kann die Kodierung der Interviewantwort mit der Antwort der Teilnehmenden zur entsprechenden Frage im Fragebogen numerisch in Beziehung gesetzt werden. Unter Beibehaltung der Gewichtungsfaktoren aus dem Fragebogen kann damit ein Vorstellungsindex aus den Interviews $C_{\text{Interview}}$ ermittelt werden, der probandenweise mit dem Ergebnis des Fragebogens $C_{\text{Fragebogen}}$ verglichen werden kann.

Es fanden zwei solcher Interviewrunden statt. Die finale Interviewrunde umfasste $N = 10$ Schülerinnen und Schüler aus einem Pilotunterricht, welche zunächst den Fragebogen ausgefüllt und unmittelbar im Anschluss am Validierungsinterview teilgenommen haben.

Die Interviews wurden vollständig durch einen unabhängigen Rater zweitcodiert, es ergab sich $\kappa = 0,76$. Tabelle 4 zeigt die Ergebnisse.

Tab. 4: Ergebnisse bei den Validierungsinterviews II.

Probanden-Nr.	$C_{\text{Fragebogen}}$	$C_{\text{Interview}}$
1	76	68
2	68	58
3	34	21
4	42	42
5	53	37
6	29	21
7	89	82
8	42	66
9	50	87
10	-3	-16

Ein graphischer Vergleich ist im Streudiagramm in Abbildung 2 zu sehen.

Man erkennt, dass ein höherer Vorstellungsindex basierend auf Fragebogenantworten ebenfalls mit einem höheren Vorstellungsindex basierend auf Interviewantworten einhergeht. Damit kann die Annahme getroffen werden, dass der Vorstellungsfragebogen grundsätzliche Tendenzen im Gesamtvorstellungsindex ähnlich zuverlässig wie das Interview nachweisen kann.

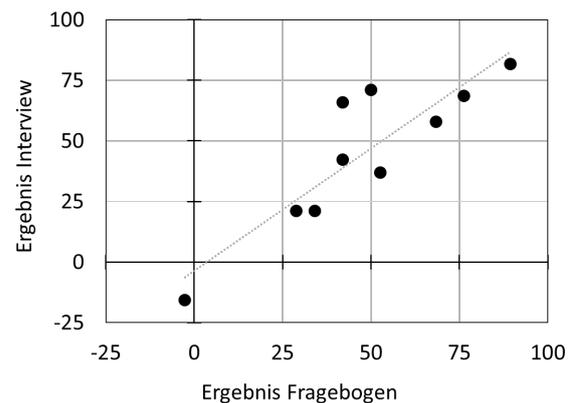


Abb. 2: Streudiagramm zu den probandenbezogenen Gesamtvorstellungsindizes aus Vorstellungsfragebogen und Validierungsinterviews II.

4. Zusammenfassung und Ausblick

Zur Evaluierung des Lernerfolgs durch das entwickelte Lehrkonzept und -material zur Quantenphysik in Klassenstufe 9 kommen unter anderem zwei Testinstrumente zum Einsatz, deren Entwicklung im Rahmen dieses Beitrages skizziert wurde.

Auf Grundlage des entwickelten Curriculums sowie zahlreicher Schritte zur Entwicklung und Analyse der Testinstrumente konnte deren grundsätzliche Eignung für den Einsatz im Rahmen einer summativen Evaluierung nachgewiesen werden.

Im Vergleich zum Interview geht beim Vorstellungsfragebogen durch die Reduktion von Vorstellungsmustern und Denkweisen auf ein numerisches Maß ein Großteil der Informationen und der damit verbundenen Interpretationsmöglichkeiten verloren. Darum werden ergänzend zum Vorstellungsfragebogen in

der Hauptstudie leitfadensbasierte Interviews zur Triangulation stichprobenartig durchgeführt.

Die Hauptstudie zur summativen Evaluation findet im Schuljahr 2023/24 in Zusammenarbeit mit zahlreichen Schulen in Deutschland sowie einiger internationaler Partnerschulen statt.

Zur Sicherstellung externer Validität der Studie findet diese als Feldstudie unter realen Bedingungen statt. Lehrkräfte unterrichten im Zuge dieser Studie das Lehrkonzept wissenschaftlich begleitet, aber selbstständig. Zur Sicherstellung interner Validität durchliefen alle teilnehmenden Lehrkräfte eine vorgelagerte Lehrkräftefortbildung.

Eine mittlere dreistellige Zahl von Schülerinnen und Schülern wird so nach dem Lehrkonzept unterrichtet. Die Evaluation im Mixed-Method-Design untersucht abschließend die beschriebenen kognitiven Lernendenmerkmale, sowie affektive Effekte des Lehrkonzeptes und Erfahrungen der beteiligten Lehrkräfte.

5. Literatur

- Albert, C., & Pospiech, G. (2023). Quantenphysik in Klasse 9: Ergebnisse einer Akzeptanzbefragung für ein Spin-First-Unterrichtskonzept. *PhyDid B - Didaktik Der Physik - Beiträge Zur DPG-Frühjahrstagung*, 1(1). Abgerufen von <https://ojs.dpg-physik.de/index.php/phydid-b/article/view/1370>
- Bitzenbauer, P. (2020). *Quantenoptik an Schulen. Studie im Mixed-Methods Design zur Evaluation des Erlanger Unterrichtskonzepts zur Quantenoptik*. Logos Verlag.
- Bühner, M. (2010). *Einführung in die Test- und Fragebogenkonstruktion* (PS Psychologie, 2., aktualisierte und erw. Aufl., [Nachdr.]. München: Pearson Studium.
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (6., vollst. überarb., akt. u. erw. Auflage 2023). Berlin: Springer Berlin; Springer.
- Flateby, T. L. (2013). *A Guide for Writing and Improving Achievement Tests*, University of South Florida. Verfügbar unter: https://www.academia.edu/27786842/A_Guide_for_Writing_and_Improving_Achievement_Tests
- Jonkisz, E., Moosbrugger, H. & Brandt, H. (2011). Planung und Entwicklung von Tests und Fragebogen. In H. Moosbrugger & A. Kelava (Hrsg.), *Testtheorie und Fragebogenkonstruktion* (Springer-Lehrbuch, 2. Aufl. 2012, S. 27–74). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-20072-4_3
- Knoblich, G. & Öllinger, M. (2006). Die Methode des lauten Denkens. In J. Funke & P. A. Frensch (Hrsg.), *Handbuch Allgemeine Psychologie - Kognition* (Handwörterbuch der Psychologie, S. 691–696). Göttingen: Hogrefe.
- Körhasan, N. D. & Miller, K. (2020). Students' mental models of wave-particle duality. *Canadian Journal of Physics*, 98(3), 266–273. <https://doi.org/10.1139/cjp-2019-0019>
- Küblbeck, J. & Müller, R. (2007). *Die Wesenszüge der Quantenphysik. Modelle, Bilder, Experimente* (3., unveränd. Aufl.). Aulis-Verl.
- Lenke, G., Thillmann, H., Wirth, J., Dicke, T. & Leutner, D. (2015). Pädagogisch-psychologisches Professionswissen von Lehrkräften: Evaluation des ProwiN-Tests. *Zeitschrift für Erziehungswissenschaft*, 18(2), 225–245. <https://doi.org/10.1007/s11618-015-0627-5>
- Mayring, P. (2010). *Qualitative Inhaltsanalyse. Grundlagen und Techniken* (11., aktual. und überarb. Aufl.). Beltz (Pädagogik).
- Michelini, M. & Stefanel, A. (2022). *Research based studies on learning quantum physics*, University of Udine.
- Müller, R. (2005). *Qualitative Quantenphysik. Eine Handreichung für die Sekundarstufe I*, IPN Kiel.
- Müller, R. (2016). Quanteninformation. Ein neues Paradigma für den Quantenphysik-Unterricht? *Praxis der Naturwissenschaften - Physik in der Schule*, 65 (1), S. 4.
- Müller, R. & Heusler, S. (2023). Schulische Zugänge mit den Wesenszügen der Quantenphysik und didaktische Potenziale. *Naturwissenschaften im Unterricht Physik*, 198, S. 2.
- Riese, J. & Reinhold, P. (2014). Entwicklung eines Leistungstests für fachdidaktisches Wissen. In I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (Springer eBook Collection, S. 257–267). Berlin, Heidelberg: Springer Spektrum.
- Sadaghiani, H. R. (2016). Spin First vs. Position First instructional approaches to teaching introductory quantum mechanics. In *2016 Physics Education Research Conference Proceedings* (S. 292–295). American Association of Physics Teachers.
- Schmiemann, P. & Lücken, M. (2014). Validität – Misst mein Test, was er soll? In I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (Springer eBook Collection, S. 107–118). Berlin, Heidelberg: Springer Spektrum.
- Schnell, C. (2016). „Lautes Denken“ als qualitative Methode zur Untersuchung der Validität von Testitems. Erkenntnisse einer Studie zur Diagnose des ökonomischen Fachwissens von Schülerinnen und Schülern der Sekundarstufe I. *Zeitschrift für ökonomische Bildung*, 2016(5), 26–49. <https://doi.org/10.7808/0502>
- Stadermann, H. K. E.; van den Berg, E. & Goedhart, M. J. (2019). Analysis of secondary school quantum physics curricula of 15 different countries: Different perspectives on a challenging topic. *Phys. Rev. Phys. Educ. Res.*, 15 (1), S. 10130. <https://doi.org/10.1103/PhysRevPhysEducRes.15.010130>