

die hochschullehre – Jahrgang 7-2021 (32)

Herausgebende des Journals: Ivo van den Berk, Jonas Leschke, Marianne Merkt, Peter Salden, Antonia Scholkmann, Angelika Thielsch

Dieser Beitrag ist Teil des Themenheftes „Paderborner Beiträge 2021“ (herausgegeben von Diana Bücken und Nerea Vöing).

Beitrag in der Rubrik Praxisforschung

DOI: 10.3278/HSL2132W

ISSN: 2199–8825 wbv.de/die-hochschullehre



Videografiertes Experimentieren als Ergänzung zum Versuchsprotokoll – Studentische Wahrnehmung eines performativen Prüfungsverfahrens

DAVID WOITKOWSKI

Zusammenfassung

Experimentierpraktika, in denen Planung, Durchführung, Auswertung und Dokumentation physikalischer Experimente geübt werden sollen, stellen einen zentralen Lehrveranstaltungstyp im Physik-Studium dar. In neuerer Zeit wird das traditionelle Veranstaltungsformat mitunter wegen einer schlechten Passung zwischen Lernzielen und tatsächlichen Lehr-, Lern- und Prüfungshandlungen kritisiert. In schulbezogenen Experimentierpraktika für Lehramtsstudierende der Physik wurde in den vergangenen Semestern eine *videografierte Experimentierprüfung* als performatives Prüfungsformat in Ergänzung zum traditionellen Versuchsprotokoll etabliert. Der Beitrag evaluiert und kontrastiert die studentische Sicht auf diese Prüfungsformate im Hinblick auf Aspekte wie Prüfungsvorbereitung, Schwierigkeit, Klarheit des Bewertungsmaßstabs, Feedbackqualität und Praxisnähe. Trotz der deutlichen strukturellen Unterschiede schätzen die befragten Studierenden die beiden Formate in den meisten Kategorien sehr ähnlich ein – die videografierte Experimentierprüfung wird jedoch als wesentlich praxisnäher wahrgenommen. In einer Gesamtschau aus theoretischer Auseinandersetzung und studentischer Bewertung zeigt sich, dass nur eine Kombination beider Prüfungsformate eine umfassende und valide Prüfungsform für Experimentierpraktika darstellen kann.

Schlüsselwörter: Experimentierpraktikum; Performanztest; Videografie; Versuchsprotokoll; Prüfungsformat

1 Ausgangslage

Experimentierpraktika stellen in Kombination mit der klassischen Vorlesungs-Übungs-Kombination eine traditionelle Lehrveranstaltungsform im Physik- und Physik-Lehramtsstudium dar (Sacher et al. 2015; Westphal 1938). Die üblicherweise damit angestrebten Lernziele umfassen das Zusammenbringen physikalischer Theorie und Praxis, den Erwerb experimenteller Fähigkeiten, motivationale Ziele und solche der Persönlichkeitsentwicklung sowie die Überprüfung von Wissen (Welzel et al. 1998). Konkret im Lehramtsstudium kommt dazu, dass das Experiment ein zentrales, strukturbildendes Element des Physikunterrichts darstellt (Tesch & Duit 2004) und das schulorientierte Experimentieren in den ländergemeinsamen Anforderungen explizit als Inhalt der Lehramtsausbildung genannt wird (KMK 2008, S. 41).

In der aktuellen Diskussion wird dieses Lehrformat allerdings auch betrachtet. Bereits Haller (1999) macht eine Diskrepanz zwischen Organisation, Zielen und Lernhandlungen im physikalischen Praktikum aus. In der Reaktion gab es in den vergangenen zwei Dekaden eine Vielzahl von Versuchen der konzeptuellen Um- oder Neugestaltung einzelner Experimentierumgebungen (Hucke 1999; Hüther 2005; Nagel 2009; Sander 2000; Zastrow 2001) oder des Experimentierpraktikums als Ganzes (Bauer & Sacher 2018; Neumann 2004; Theyßen 1999), wobei hier verschiedene Aspekte wie eine Anforderungsprogression, Feedback- und Betreuungsformate, fachliche und methodische Rekonstruktionen im Vordergrund standen.

Gegenstand der vorliegenden Studie ist das verbreitete Prüfungsformat des Versuchsprotokolls, einer traditionell geformten, stark mathematisierten Textform, die in allen o. g. Um- und Neugestaltungen als ‚gegeben‘ hingenommen wurde und zu der bisher kaum oder keine Alternativen diskutiert oder implementiert wurden. An der Universität Paderborn wurde dies in den schulbezogenen Experimentierpraktika für Lehramtsstudierende durch eine videografierte Experimentierprüfung ergänzt und teilweise ersetzt. Dabei handelt es sich um einen Performanztest, der aufgrund seiner als höher angenommenen Authentizität und Praxisnähe im hochschuldidaktischen Kontext häufig favorisiert wird (Biggs & Tang 2011; Schaper, Hilkenmeier & Bender 2013; Schröder 2015). Die Studie geht der studentischen Wahrnehmung dieser beiden Prüfungsformen nach und fragt nach der wahrgenommenen Schwierigkeit, Klarheit der Anforderungen und Praxisnähe im Vergleich. Sie soll damit einen Beitrag zur Überprüfung der Etablierung dieses Prüfungsformates leisten.

2 Lehrveranstaltung: Schulphysik

Gegenstand der hier berichteten Lehrinnovation sind Lehrveranstaltungen in Lehramtsstudiengängen Physik an der Universität Paderborn für zwei Schulformen. In beiden Studiengängen sind Veranstaltungen zum schulbezogenen Experimentieren in sehr unterschiedlichem Umfang und an unterschiedlicher Stelle verortet. Diese Veranstaltungen werden nur von Studierenden des jeweiligen Lehramtsstudiengangs belegt.

Die Prüfungsordnung für das Lehramt an Haupt-/Real-/Sekundar- und Gesamtschulen mit dem Unterrichtsfach Physik (Präsidium der Universität Paderborn 2016a) sieht für das 2. bis 4. Semester drei Module *Schulphysik I bis III*, bestehend aus einem Experimentierpraktikum (3 SWS, 105 h Workload) und einer begleitenden Übung (1 SWS, 15 h Workload) vor. Während die physikalischen Inhalte thematisch den jeweils ein Semester davor verorteten Modulen *Experimentalphysik I bis III* folgen, sollen die Studierenden in der Schulphysik jeweils ein „vertieftes Verständnis“ der jeweiligen Fachinhalte, die Fähigkeit, diese Konzepte „zu vernetzen und bei der Bearbeitung einfacher experimenteller Problemstellungen anzuwenden“, sowie „Erfahrungen bei der Durchführung, Auswertung und Interpretation qualitativer wie quantitativer Praktikums- und Schulphysikexperimente“ (Präsidium der Universität Paderborn 2016a, S. 14) erwerben. Diese Ziele stehen in Übereinstimmung mit den verbreiteten Zielvorstellungen physikalischer Praktika bei Welzel et al. (1998).

Im Studiengang für das Lehramt an Gymnasien und Gesamtschulen wird erst im Master ein Experimentierpraktikum *Experimente der Schulphysik* (2 SWS, 60 h Workload) angeboten (Präsidium der Universität Paderborn 2017), das auf den physikalischen Praktika im Bachelorstudium aufbaut (Präsidium der Universität Paderborn 2016b). Die Lernziele gleichen hier denen der *Schulphysik I bis III*, beziehen sich aber zusätzlich auf die Inhalte der Sekundarstufe II.

Die Prüfungsform für diese Module ist jeweils ein Abschlussportfolio, in das laut Prüfungsordnung die Vorbereitung, Durchführung und Auswertung von Experimenten eingehen soll. Aktuell wird dies durch bewertete Antestate als „kleine“ Prüfungsform sowie zwei Versuchsproto-

kolle und zwei videografierte Experimentierprüfungen als „große“ Prüfungsformen realisiert.¹ Gegenwärtig ist dabei weder theoretisch klar, ob und inwieweit die Versuchsprotokolle und videografierten Experimentierprüfungen als „große“ Prüfungsformen vergleichbare Anforderungen an die Studierenden stellen noch wie sie von den Studierenden in Bezug auf typische Qualitätsmerkmale von Prüfungen – wie Klarheit der Anforderungen, Feedbackqualität, Nachvollziehbarkeit der Bewertung und Schwierigkeit – eingeschätzt werden.

Die Prüfungen werden von den jeweiligen Lehrenden abgenommen. Dabei handelt es sich um wissenschaftliche Mitarbeiter:innen der Arbeitsgruppe Didaktik der Physik. Zwischen den Lehrenden der verschiedenen Veranstaltungen und verschiedener Semester werden Aufgabenstellungen und Prüfungsanforderungen regelmäßig ausgetauscht und abgeglichen, um einen gemeinsamen und für die Studierenden transparenten Bewertungsstandard herzustellen.

3 Prüfungsformen in Experimentierpraktika

Die beiden „großen“ Prüfungsformen in der aktuellen Form dieser Experimentierpraktika weisen unterschiedliche Charakteristika auf, die es theoretisch plausibel erscheinen lassen, dass sie sowohl andere Fähigkeiten und Fertigkeiten prüfen als auch von Studierenden unterschiedlich wahrgenommen und bewertet werden. Im Folgenden werden zunächst die beiden Prüfungsformen ausführlicher dargestellt und anschließend in die Kompetenzpyramide von Miller (1990) sowie in Bezug auf typische Qualitätskriterien als theoretischen Rahmen eingeordnet. Daraus ergeben sich dann Forschungsfragen und Hypothesen für die vorliegende Studie.

3.1 Versuchsprotokolle

Klassische Versuchsprotokolle sind eine stark strukturierte und mathematisierte, traditionell überrkommene Textform, in der Versuchsziel, theoretische Annahmen, Aufbau und Durchführung sowie Ergebnisse eines Experiments dargestellt werden (Cordes 2017; Krabbe 2015; Riewerts 2016). Die Studierenden machen sich während der Versuchsdurchführung Notizen zu Messergebnissen oder weiteren Sachverhalten (Laborbuch). Das eigentliche, bewertete Versuchsprotokoll ist dann ein ca. 5–10 Seiten umfassender Text, welcher Fragestellung, Theorie, Versuchsaufbau, Auswertung und Interpretation des Versuchs auf Hochschulniveau darstellt.

In den hier diskutierten Lehrveranstaltungen wird das Protokoll korrigiert und zurückgegeben und kann einmalig überarbeitet werden. Die Bewertung wird anhand einer Checkliste mit Bewertungskriterien transparent gemacht. Wie in vielen traditionellen Experimentierpraktika ist die eigentliche Durchführung des Experiments (anders als von der Prüfungsordnung gefordert) nicht bzw. nur in Form einer nachträglichen Darstellung Gegenstand der Prüfung.

In der Praxis der vergangenen Jahre haben sich verschiedene Probleme mit Versuchsprotokollen als einzige „große“ Prüfungsform herauskristallisiert. Dazu gehört, dass das Protokoll lediglich eine schriftliche Darstellung post-hoc ist, also nicht die Durchführung des Experiments selbst abbildet. Dazu kommt, dass die Textform den Studierenden aus der Schule (zumindest in der geforderten Strenge) nicht bekannt ist, sodass in der Übung zur *Schulphysik I* wie auch in den Physikalischen Grundpraktika im Gymnasial-/Gesamtschul-Lehramt nennenswerte Zeit auf das Einüben dieser Textform verwendet werden muss, die die Studierenden dann (außer in einer sehr reduzierten Variante) wahrscheinlich nicht für ihren späteren Lehrberuf benötigen. Unabhängig von der Praxisrelevanz gilt das Schreiben eines Versuchsprotokolls aber als wesentliche zu erwerbende Fähigkeit in vielen naturwissenschaftlichen Studiengängen (Riewerts 2016).

Dennoch ist die Qualität der Protokolle häufig auch nach der Überarbeitung nicht zufriedenstellend. Die Studierenden haben sichtlich Probleme mit der stark restringierten Textform (Born-

¹ Die Ausgestaltung und Gewichtung der verschiedenen Teilleistungen ist in der Studienordnung nicht festgelegt und variiert zwischen den jeweiligen Lehrenden.

schein 2016): Textgliederung, Mathematisierung, Darstellen eines kohärenten Gedankenganges, Konzentration auf das Wesentliche. Die Studierenden äußern im Gespräch häufig mangelnde Motivation, zumal das Abfassen von Versuchsprotokollen nicht als praxisrelevant für den späteren Lehrberuf angesehen wird.

3.2 Videografierte Experimentierprüfung

Bei der videografierten Experimentierprüfung handelt es sich ursprünglich um einen fachdidaktischen Performanztest, der die Experimentierhandlungen der Studierenden in einer realitätsnahen Laborsituation sichtbar machen soll. Das von Szabone Varnai und Reinhold (2017) dargestellte Erhebungsverfahren wurde in den hier dargestellten Lehrveranstaltungen als Prüfungsverfahren adaptiert.

Bei der videografierten Experimentierprüfung führen Studierende vor laufender Kamera ein Experiment vom Aufbau bis zur Auswertung durch. Die Kamera ist dabei von oben auf den Experimentiertisch gerichtet, sodass die Experimentierhandlung, nicht jedoch die/der Experimentator:in aufgezeichnet wird. Zwischen den einzelnen Phasen erläutern und begründen die Studierenden ihre jeweiligen Experimentierhandlungen, zentrale Ideen, theoretische Hintergründe und Vorgehensweisen. Sie bekommen dazu einen Leitfaden, dessen Inhalt und Struktur denen klassischer Protokolle gleichen. Zusätzlich werden die experimentellen Ergebnisse (i. d. R.) in Form einer Exceltabelle erfasst und mit bewertet. In den hier diskutierten Lehrveranstaltungen wird die Bewertung wie beim Versuchsprotokoll mit einer Checkliste transparent gemacht, die die Studierenden im Vorfeld kennen. Struktur und Inhalt der Checkliste ähneln dabei stark denen für Versuchsprotokolle.

Anders als das klassische Versuchsprotokoll wird hier lediglich mündlich, nicht aber schriftlich (allerdings durchaus fachsprachlich und fachbezogen strukturiert) erklärt. Die schriftliche Ausarbeitung wird also (anders als von der Prüfungsordnung gefordert) nicht bewertet. Weiterhin wurde das in der Experimentierprüfung durchzuführende Experiment bereits im Semesterverlauf (ggf. leicht abgewandelt) durchgeführt. Das dabei abgefasste Laborbuch darf von den Studierenden in der Prüfung als Hilfsmittel verwendet werden, was den Studierenden (zumindest im Prinzip) eine längerfristige Vorbereitung auf die Prüfung ermöglicht. In der Praxis zeigt sich, dass einige Studierende hier größere Schwierigkeit mit der freien Rede unter Benutzung von Fachsprache haben, zum Teil zeigen sich sogar Anzeichen von Prüfungsangst.

Die Methode wird den Studierenden gegenüber als Experimentieren in Analogie zu einem Demonstrationsexperiment im Physikunterricht begründet – auch, um die Motivation durch ein höheres Maß an Praxisrelevanz zu erhöhen. Ob und inwieweit dies gelingt, ist bisher unklar.

3.3 Vergleich der Prüfungsanforderungen

Für einen theoretischen Vergleich der beiden vorgestellten Prüfungsformen kann zunächst auf die Unterscheidung der Kompetenzpyramide nach Miller (1990) zurückgegriffen werden (Abb. 1). Die höchste Stufe wäre hier die Beobachtung freien beruflichen Handelns, hier also des Experimentierens im Unterricht. Verglichen damit könnte die videografierte Experimentierprüfung als eine Performanz in einer simulierten Situation beschrieben werden, da hier zwar eine vergleichbare Handlung gezeigt, eine Reihe von Kontextbedingungen (Schüler:innen, Klassenraumsituation) aber unterdrückt bzw. kontrolliert werden. Kulgemeyer, Riese, Borowski, Schreiber und Vogelsang (2018, S. 47) bezeichnen solche „Verfahren, bei denen Unterrichtshandlungen unter standardisierten Rahmenbedingungen simuliert werden“ als Performanztest.

Im Vergleich dazu würde das Verfassen eines Versuchsprotokolls teils einen Test deklarativen Wissens (z. B. in der Darstellung der theoretischen Grundlagen) oder einen Kompetenztest (z. B. in der Auswertung und Fehlerdiskussion) darstellen und somit als Prüfungsform geringerer Authentizität gelten (Kulgemeyer et al. 2018). Auf der anderen Seite sehen Sacher, Probst und Reinhold (2015) das Versuchsprotokoll in einem Experimentierpraktikum für Fach-Studierende der Physik als Vorstufe der Veröffentlichung eines experimentellen Befundes in einer Fachzeitschrift.

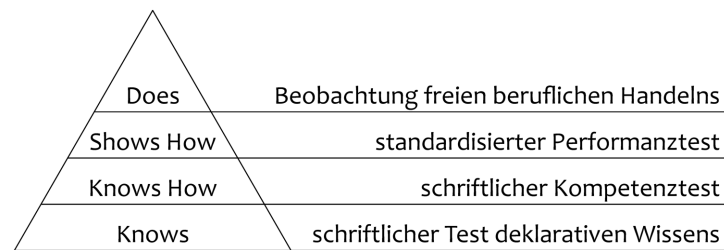


Abbildung 1: Kompetenzpyramide nach Miller (1990) mit zugehörigen Testformaten (Übersetzung nach Kulgemeyer et al. 2018)

Ebenso könnte das Versuchsprotokoll als ausführlichere Variante einer auch im Physikunterricht einzuübenden Darstellungsform (vgl. Krabbe 2015) gelten. Die Frage der Authentizität ist also nicht ganz eindeutig zu entscheiden.

Ein weiterer Unterschied zwischen den Prüfungsformaten ist in einer Abwägung zwischen Reliabilität und Validität zu sehen. Während Testverfahren weiter unten in der Kompetenzpyramide hoch reliabel sein können, sind die oberen Testverfahren authentischer und damit valider, gleichzeitig aufgrund der Variabilität der authentischen Situation aber weniger reliabel (Wass et al. 2001). Kulgemeyer et al. (2018) versuchen, diesen Reliabilitätsverlust durch Standardisierung zwar zu mindern, dies gelingt aber nur begrenzt und nur unter Beschneidung der getesteten Situation auf einen sehr eingeschränkt abgebildeten Realitätsbereich – eine Vorgehensweise, die in analoger Weise auch auf die videografierte Experimentierprüfung angewandt wird.

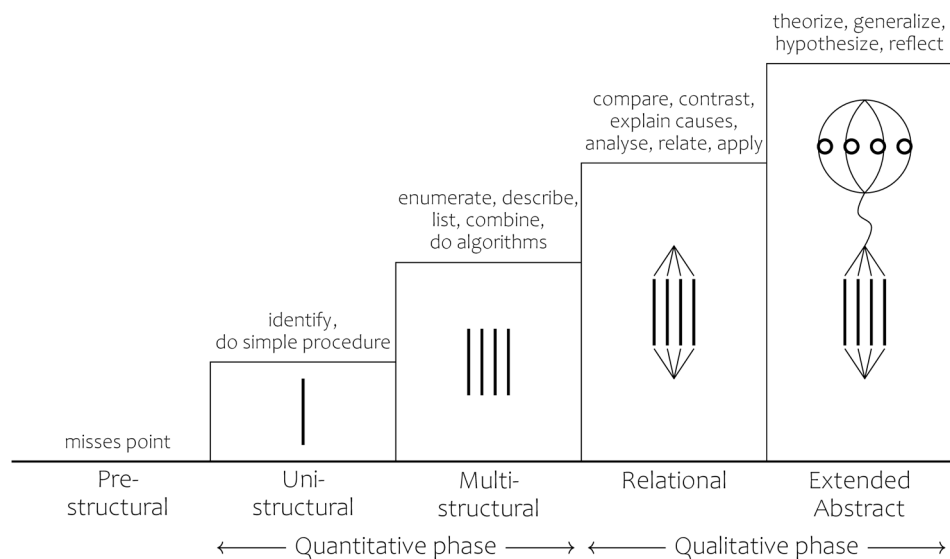


Abbildung 2: Taxonomie von Anforderungen nach Biggs und Tang (2011, S. 91)

Die Einordnung in die Kompetenzpyramide macht zunächst noch keine Aussage über die Schwierigkeit der gestellten Anforderungen – auch ein schriftlicher Test kann sehr schwierig sein. Für eine erste kriteriale Schwierigkeitseinschätzung kann eine Taxonomiestufung wie z. B. die von Biggs und Tang (2011) herangezogen werden (Abb. 2), die sich stärker als die bekannte Taxonomie von Bloom (Anderson & Krathwohl 2001) auf Prüfungsformen der Hochschule konzentriert, ansonsten aber vergleichbare Einschätzungen liefert. Hier erscheint eine Verortung der Prüfungsanforderungen teils auf der multistrukturellen Ebene (Studierende müssen vorhandenes Wissen zu einer Darstellung kombinieren und bei der Auswertung bekannte Algorithmen auf den konkreten Versuch anwenden) und auf der relationalen Ebene (Studierende erklären Zusammenhänge aufgrund vorhanden Wissens, sie analysieren z. B. Fehler und unerwartete Messdaten und wenden erlernte Vorgehensweisen und Argumentationsweisen an) für beide Prüfungsformen gerechtfertigt.

Beim Vergleich so unterschiedlicher Prüfungsformate sind darüber hinaus aber noch weitere Schwierigkeitsunterschiede zu erwarten, die durch eine Taxonomiestufung nicht angemessen erfasst werden: Studierende könnten z. B. die schriftliche Form, das exakte Formulieren, als größere Herausforderung ansehen oder sie sehen in der Ad-hoc-Formulierung unter dem Zeitdruck der videografierten Experimentierprüfung eine größere Schwierigkeit, die sich teilweise eventuell in prüfungsangstähnlichen Zuständen äußern könnte.

Dazu kommt, dass die Studierenden das Experiment der videografierten Experimentierprüfung im Semester schon einmal durchgeführt haben, sodass es sich hier (perfekte Vorbereitung vorausgesetzt) auch um eine rein reproduktive Leistung handeln könnte. Beim Versuchsprotokoll besteht umgekehrt die Möglichkeit, geschickt (und ggf. unentdeckt) Protokollteile von anderen Gruppen zu übernehmen. Beide Effekte stellen die Aussagekraft der Taxonomiestufung für diesen Vergleich zumindest teilweise infrage.

3.4 Qualitätskriterien von Prüfungen

Wunderlich und Szczyrba (2011, S. 14) nennen 4 Qualitätskriterien von Prüfungen: Sie sollen *valide*, *reliabel*, *fair* und *ökonomisch* sein (vgl. auch Dubs 2011, S. 4), wobei das letzte eher für die Lehrenden- als für die Studierendensicht relevant erscheint.

Valide ist eine Prüfungsform, wenn sie tatsächlich die Kompetenzen abprüft, die laut Modulbeschreibung gelernt werden sollen. Da diese sowohl das Experimentieren als Prozess als auch die anschließende Darstellung und Auswertung der Ergebnisse umfassen (Präsidium der Universität Paderborn 2016a; Präsidium der Universität Paderborn 2017), erscheint eine valide Prüfung in diesem Sinne nur durch eine Kombination beider Formen möglich. Aus Sicht der Studierenden kommt neben den Lernzielen der Prüfungsordnungen aber auch das langfristige Ziel der Ausbildung, nämlich der Lehrberuf als Validitätskriterium, zum Tragen – ein Aspekt der häufig „Praxisnähe“ genannt wird. Wie erläutert, wird die videografierte Experimentierprüfung den Studierenden zumindest unter diesem Aspekt vorgestellt.

Prüfungsformen sind dann reliabel, wenn sie reproduzierbar, also ohne zufällige Schwankungen, bewertet werden. Studierende werden hier vor allem die spezifische Schwierigkeit einer Prüfungsform (die auch in die Validität eingehen könnte) sowie die Qualität, Nachvollziehbarkeit und Transparenz des Feedbacks wahrnehmen (Wunderlich & Szczyrba 2011, S. 14). Die Schwierigkeiten des Versuchsprotokolls liegen sicherlich eher in den Herausforderungen einer mathematisierten Formulierung eines konsistenten Gedankenganges; bei der videografierten Experimentierprüfung können eher das freie Sprechen und eventuell eine durch die Prüfungsform begünstigte Prüfungsangst genannt werden. Mithilfe der Checklisten wird versucht, die Qualität des Feedbacks in beiden Prüfungsformen auf hohem Niveau zu halten.

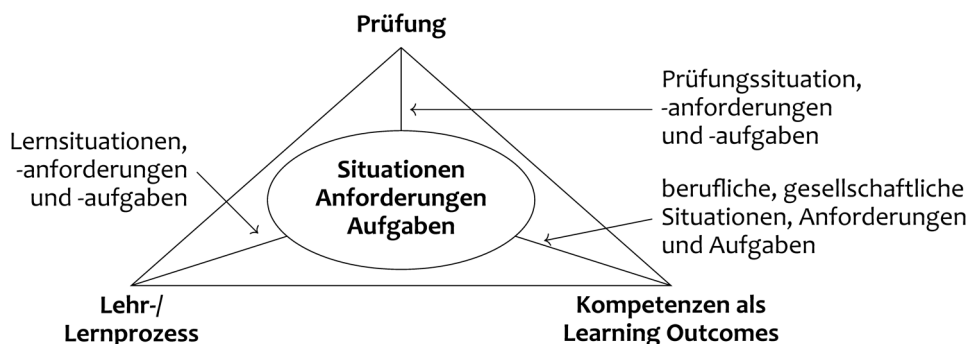


Abbildung 3: Constructive Alignment (Quelle: Wildt & Wildt 2011, S. 9)

Die Fairness beziehen Wunderlich und Szczyrba (2011, S. 14) mit Bezug auf die Studierendensicht stark darauf, ob die Studierenden sich im Laufe der Veranstaltung gut auf die Prüfung vorbereiten konnten (was wiederum auf die Validität der Prüfung zurückfällt) und ob der Bewertungsmaßstab im Vorhinein klar kommuniziert wurde. Inwiefern sich die Studierenden tatsächlich gut auf die

Prüfung vorbereitet fühlen, ist bei beiden Formen schwer abzuschätzen, die Gestaltung der Lehrveranstaltungen sollte im Prinzip aber dazu beitragen. Der Bewertungsmaßstab wird (wiederum in Form der Checklisten) bereits vor der Prüfung veröffentlicht und sollte insofern ausreichend klar sein.

Sowohl die Validität als auch die Fairness einer Prüfung können auch im Kontext eines *Constructive Alignment* (Biggs & Tang 2011), also der Abstimmung zwischen Lernhandlung, Prüfungsform und Lernziel (Abb. 3), betrachtet werden. Während bei der videografierten Experimentierprüfung eine Experimentier- und Erklärhandlung geprüft wird, die im Rahmen des Praktikumsbetriebs und der Antestate bereits geübt wurde, ist diese Passung bei den Versuchsprotokollen weit weniger gut. Die Passung zu den erwarteten Kompetenzen wiederum kann unterschiedlich bewertet werden, je nachdem, ob man eher die Versuchsdurchführung im Unterricht oder die strukturierte Versuchsauswertung und -dokumentation als Maßstab heranzieht.

4 Fragestellung und Hypothesen

Im Folgenden sollen die beiden „großen“ Prüfungsformen in der Schulphysik aus Perspektive der Studierenden verglichen werden. Kategorien dieses Vergleichs ergeben sich aus den oben angestellten theoretischen Vergleichen. Als Forschungsfrage wird formuliert:

Wie nehmen Studierende die beiden Prüfungsformen „Versuchsprotokoll“ und „videografierte Experimentierprüfung“ in Bezug auf (a) eigene Vorbereitung, (b) Schwierigkeit, (c) Klarheit des Bewertungsmaßstabs, (d) Feedbackqualität und (e) Praxisnähe wahr?

Zu den fünf Teilfragen (a) bis (e) können wie folgt Hypothesen formuliert werden:

- a) Die Studierenden müssen sich generell in jedem Fall auf das Antestat vor dem Versuch vorbereiten, da dieses vor Durchführung des Experiments bestanden werden muss. Auf das Verfassen eines Versuchsprotokolls werden sie zu Beginn der jeweiligen Praktika ausführlich vorbereitet. Die Vorbereitung auf die Experimentierprüfung geschieht stärker selbstgesteuert durch das Führen eines Laborbuches. Insgesamt kann – auch vor dem Hintergrund des besseren Constructive Alignment – die Vorbereitung auf die Experimentierprüfung als besser eingeschätzt werden.
- b) Die Verortung auf den Taxonomiestufen von Biggs spricht für eine vergleichbare Schwierigkeit. Auf die subjektive Schwierigkeitswahrnehmung sollte sich aber auch das notenmäßige Resultat der jeweiligen Prüfung auswirken. Da die Noten für Versuchsprotokolle in den vergangenen Semestern tendenziell schlechter ausfielen als für die videografierten Experimentierprüfungen, wäre hier eine höhere subjektive Schwierigkeit der Protokolle anzunehmen. Aufgrund der anderen Prüfungssituation kann aber bei den Experimentierprüfungen eher Prüfungsangst auftreten.
- c) Aufgrund der häufig schwachen Qualität der überarbeiteten Versuchsprotokolle, kann dort von einer geringen Klarheit des Bewertungsmaßstabes ausgegangen werden. Da an die videografierte Experimentierprüfung ein vergleichbarer Bewertungsmaßstab angelegt wird, ist hier eine ähnlich geringe Klarheit zu vermuten.
- d) Versuchsprotokolle werden im Text korrigiert bzw. kommentiert und die Bewertung mit einer Checkliste zusammengefasst, bei der Experimentierprüfung wird i. d. R. aber nicht das Video kommentiert, sondern nur eine Checkliste verwendet, sodass hier ein geringerer Bezug zur tatsächlich gezeigten Prüfungshandlung besteht. Daher ist bei den Versuchsprotokollen von einer höheren Feedbackqualität auszugehen.
- e) Aufgrund der Argumentation (auch den Studierenden gegenüber), die videografierte Experimentierprüfung stelle eine Simulation eines Demonstrationsexperimentes aus dem Physikunterricht unter kontrollierten Bedingungen dar, ist hier mit einer wesentlich höher wahrgenommenen Praxisnähe zu rechnen.

5 Methode

5.1 Testinstrument

Der eingesetzte Online-Fragebogen ist aus testökonomischen Gründen bewusst kurz gehalten. Zu Beginn der Befragung geben die Teilnehmenden jeweils an, in welchem Experimentierpraktikum (Modulname und Semester) sie zuletzt mindestens eine videografierte Experimentierprüfung absolviert und mindestens ein Versuchsprotokoll verfasst haben.

Der Kern des Fragebogens besteht aus 14 geschlossenen und 2 offenen Items (Tab. 1). Hier wurden zu jeder Teilforschungsfrage je 2 oder mehr 4-stufige Zustimmungssitems als Indikatoren formuliert. Die Items zur Schwierigkeit enthalten zusätzlich einen niedrigschwelligen Indikator für Prüfungsangst (Item S2) sowie für Fachsprache (Item S4) und Mathematik (Item S5) als spezifische schwierigkeiterzeugende Faktoren. In den offenen Items werden eine geringstrukturierte Gesamteinschätzung (Item O) sowie eine begründete Note für die Prüfungsform (Items N1, N2) abgefragt. Die Items sind in identischer Formulierung, aber randomisierter Reihenfolge je einmal für jede Prüfungsform aufgeführt, erst zur Experimentierprüfung, dann zum Versuchsprotokoll.

Tabelle 1: Itemüberblick und -zuordnung

ID	Geschlossene Items	Teilfrage
V1	Ich fühle mich auf die Experimentierprüfung gut vorbereitet.	(a) eigene Vorbereitung
V2	Ich brauche mehr Übungsgelegenheiten für die Experimentierprüfung.	
S1	Die Experimentierprüfung ist für mich sehr schwierig.	(b) Schwierigkeit
S2	Beim Gedanken an die Experimentierprüfung wird mir unwohl.	
S3	Bei der Experimentierprüfung fällt es mir leicht, eine gute Note zu bekommen.	
S4	Bei der Experimentierprüfung fällt mir die Benutzung richtiger Fachsprache schwer.	
S5	Der richtige Einsatz von Mathematik/Formeln fällt mir bei der Experimentierprüfung schwer.	
B1	Ich weiß, was in der Experimentierprüfung von mir erwartet wird.	(c) Klarheit des Bewertungsmaßstabs
B2	Ich kann die Note für die Experimentierprüfung häufig nicht nachvollziehen.	
B3	Mir ist klar, nach welchen Kriterien die Experimentierprüfung bewertet wird.	
F1	Das Feedback zur Experimentierprüfung hilft mir, beim nächsten Mal eine bessere Leistung zu zeigen.	(d) Feedback
F2	Ich bekomme nach der Experimentierprüfung hilfreiches Feedback.	
P1	Die Experimentierprüfung ist relevant für meinen späteren Beruf.	(e) Praxisnähe
P2	Die Experimentierprüfung hat nichts mit echtem Physikunterricht zu tun.	
	Offene Items	
O1	Folgendes gefällt mir an der Experimentierprüfung besonders gut oder schlecht.	
N1 N2	Bewerten Sie bitte die Prüfungsform „videografierte Experimentierprüfung“ mit einer Schulnote und begründen Sie diese Note.	
	Direkter Vergleich	
	Geben Sie jeweils an, welche der beiden Prüfungsformen insgesamt ...	
D1	... schwieriger ist.	(b) Schwierigkeit
D2	... praxisnäher ist.	(e) Praxisnähe

(Fortsetzung Tabelle 1)

ID	Geschlossene Items	Teilfrage
D3	... nachvollziehbarer bewertet wird.	(c) Klarheit des Bewertungsmaßstabs (d) Feedback
D4	... mehr Spaß macht.	
O2	Folgende Unterschiede zwischen den Prüfungsformen fehlen hier noch.	

Anmerkung: Die Items V1 bis N2 sind jeweils einmal für die videografierte Experimentierprüfung und dann für das Versuchsprotokoll formuliert. Daran schließen sich die Items D1 bis O2 an.

Abschließend werden die Proband:innen um eine vergleichende Einschätzung gebeten, welche der Prüfungsformen insgesamt schwieriger, praxisnäher, nachvollziehbarer bewertet sei und mehr Spaß mache (5-stufig² von Protokoll bis Experimentierprüfung). Außerdem wird auch hier noch einmal in einem offenen Item nach weiteren Unterschieden zwischen den Prüfungsformen gefragt.

Um die Noten der Studierenden als Vergleichsgröße (z. B. mit der subjektiv wahrgenommenen Schwierigkeit) heranziehen zu können, können die Studierenden auf der letzten Seite des Online-Fragebogens wählen, ob sie (a) unter Nennung ihres Namens die Einwilligung geben, die Noten von den jeweiligen Lehrenden zu erfragen (die Namen werden danach aus dem Datensatz gelöscht), oder (b) anonym bleiben und dann ihre Noten selbst angeben.

5.2 Stichprobe

Es handelt sich um eine Gelegenheitsstichprobe, die Studierenden der vergangenen Durchgänge der *Schulphysik I, II und III* sowie der *Experimente der Schulphysik* werden Ende der Frühjahrs-Semesterferien 2020 per Mail zur Teilnahme an einer Online-Befragung aufgefordert. Dabei wurden 34 Studierende angeschrieben, die in den vergangenen Semestern an Experimentierpraktika teilgenommen hatten, in denen beide untersuchten Prüfungsformen eingesetzt worden waren (7 aus der LV *Experimente der Schulphysik* und 27 aus *Schulphysik I, II, III*). Die Teilnahme an der Studie war freiwillig, geht nicht in die Bewertung des Moduls ein und ist kein Prüfungs- oder Bewertungskriterium anderer Art.

Von den angeschriebenen Studierenden haben nach drei Wochen 19 Personen an der Befragung teilgenommen (Tab. 2). Das Ausfüllen des Instruments dauerte im Mittel 18 min (SD = 15 min). Aus Gründen, die hier nicht weiter diskutiert werden sollen, nehmen die Studierenden tendenziell ein Jahr später an den *Schulphysik*-Praktika teil, als nach Studienordnung vorgesehen ist, einzelne Studierende geben auch *sehr* hohe Fachsemester-Zahlen an (eine Person in der *Schulphysik II* gibt ein Fachsemester von 15 an). Die Noten für die Experimentierprüfungen und Versuchsprotokolle sind tendenziell in den *Experimenten der Schulphysik* besser als in der *Schulphysik II* und dort wiederum besser als in der *Schulphysik III*. Die Unterschiede zwischen den Lehrveranstaltungen sind in einer ANOVA für beide Prüfungsformen jeweils signifikant (Experimentierprüfung: $F(2, 26) = 5.0$, $p = 0.014$; Protokolle: $F(2, 20) = 6.8$, $p = 0.05$). Ob es sich hier um einen Kohorten-, Studiengangs- oder Lehrenden-Effekt handelt, ist aufgrund der geringen Stichprobengröße nicht feststellbar.

2 Während bei den Zustimmungsisems (wie in vielen ähnlichen Skalen) eine Entscheidung zwischen „trifft zu“ und „trifft nicht zu“ erzwungen werden sollte, sollte bei den Vergleichsisems gezielt eine Möglichkeit „beide Prüfungsformen gleich“ gegeben werden, falls die Studierenden tatsächlich keinen Unterschied wahrnehmen sollten.

Tabelle 2: Stichprobenüberblick

LV	Stichprobe				Gesamte LV	
	N	FS	Protokolle	Experimentierprüfung	Protokolle	Experimentierprüfung
Schulphysik II	6	6.3 ± 4.5 (Bachelor)	1.9 ± 0.8 (10)	2.0 ± 0.9 (10)	1.58 ± 0.71 (12)	2.30 ± 0.96 (14)
Schulphysik III	7	5.7 ± 1.5 (Bachelor)	3.6 ± 1.3 (12)	2.7 ± 0.9 (12)	3.57 ± 1.15 (31)	2.94 ± 0.88 (29)
Experimente d. Schulphysik	6	1.8 ± 1.0 (Master)	1.5 (1)	1.5 ± 0.5 (7)	1.52 ± 0.60 (5)	1.51 ± 0.83 (7)
Gesamt	19		2.7 ± 1.4 (23)	2.2 ± 0.9 (29)		

Anmerkung: Für jede Lehrveranstaltung sind das durchschnittliche Fachsemester, die durchschnittlichen Noten für die Protokolle und Experimentierprüfungen in der Form $M \pm SD$ angegeben. Da nicht alle Proband:innen Noten angegeben haben, sind dort die Anzahl der Angaben in Klammern ergänzt. Es konnten jeweils maximal 2 Noten für Protokolle und 2 Noten für Experimentierprüfungen angegeben werden. Zum Vergleich wurde der Notenspiegel der gesamten Lehrveranstaltung angegeben, die Abweichungen sind jeweils nicht signifikant.

Eine Repräsentativität der Stichprobe in Bezug auf Merkmale wie physikalisches Vorwissen zu Studien-/Veranstaltungsbeginn, Sprachkenntnisse/Herkunft o. ä. ist (auch aufgrund der geringen Probandenzahl) nicht zu klären. Der Anteil an Studierenden mit Migrationshintergrund oder ohne deutsche Muttersprache ist unter Lehramts-Studierenden aber insgesamt relativ gering.

Dadurch, dass bevorzugt diejenigen an der Befragung teilnehmen, die mit der jeweiligen Prüfung besonders gut zurechtgekommen sind oder besonders frustriert waren, kann es zu einer Positiv- oder Negativauswahl kommen. Dies wurde überprüft, indem auch der anonymisierte Gesamt-Notenspiegel der vergangenen Semester in den drei Lehrveranstaltungen bei den Lehrenden abgefragt wurde. Den Überblick zeigt ebenfalls Tabelle 2. Eine signifikante Abweichung zwischen Stichprobe und so erhobener Grundgesamtheit kann mit dem Wilcoxon-Mann-Whitney-Test für alle drei Lehrveranstaltungen verneint werden. Es handelt sich also um keine Positiv- oder Negativauswahl.

6 Ergebnisse

Die Ergebnisse werden zunächst nach Teilfragen getrennt diskutiert, anschließend folgt die Behandlung der Fragen zum direkten Vergleich der Prüfungsformen.

Gruppenunterschiede bzw. Unterschiede zwischen den Prüfungsformen werden mit dem Wilcoxon-Mann-Whitney-Test (WMW-Test) auf Signifikanz geprüft, da aufgrund der geringen Probanden- und Itemzahl keine Normalverteilung angenommen werden kann (Hollander & Wolfe 1973). Bei signifikanten Unterschieden wird Cohens d als Effektstärkemaß angegeben. Dabei gelten Effekte mit $d > 0,2$ als klein, $d > 0,5$ als mittelgroß und $d > 0,8$ als groß (Tiemann & Körbs 2014, S. 291). Zur weiteren Interpretation der numerischen Ergebnisse werden die Freitextantworten herangezogen.

6.1 Einschätzung nach Teilfragen

Die Antworten auf die vierstufigen Items V1 bis P2 in Tabelle 1 werden in zwei Formen berichtet. Einerseits werden allgemeine Tendenzen angegeben, indem diejenigen Proband:innen ausgezählt werden, die einem Item zustimmen oder eher zustimmen, bzw. diejenigen Proband:innen, die nicht zustimmen oder eher nicht zustimmen. Zur weiteren numerischen Analyse werden die Items mit 1 = Trifft gar nicht zu bis 4 = Trifft völlig zu codiert. Abbildung 4 zeigt die daraus gebildeten Mittelwerte getrennt nach Prüfungsform. Da die zu einer Teilfrage gehörigen Items in den

meisten Fällen nicht signifikant miteinander korrelieren (was auch an der geringen Stichprobengröße liegen kann), werden diese Items hier nicht zu Skalen zusammengesetzt, sondern einzeln berichtet.

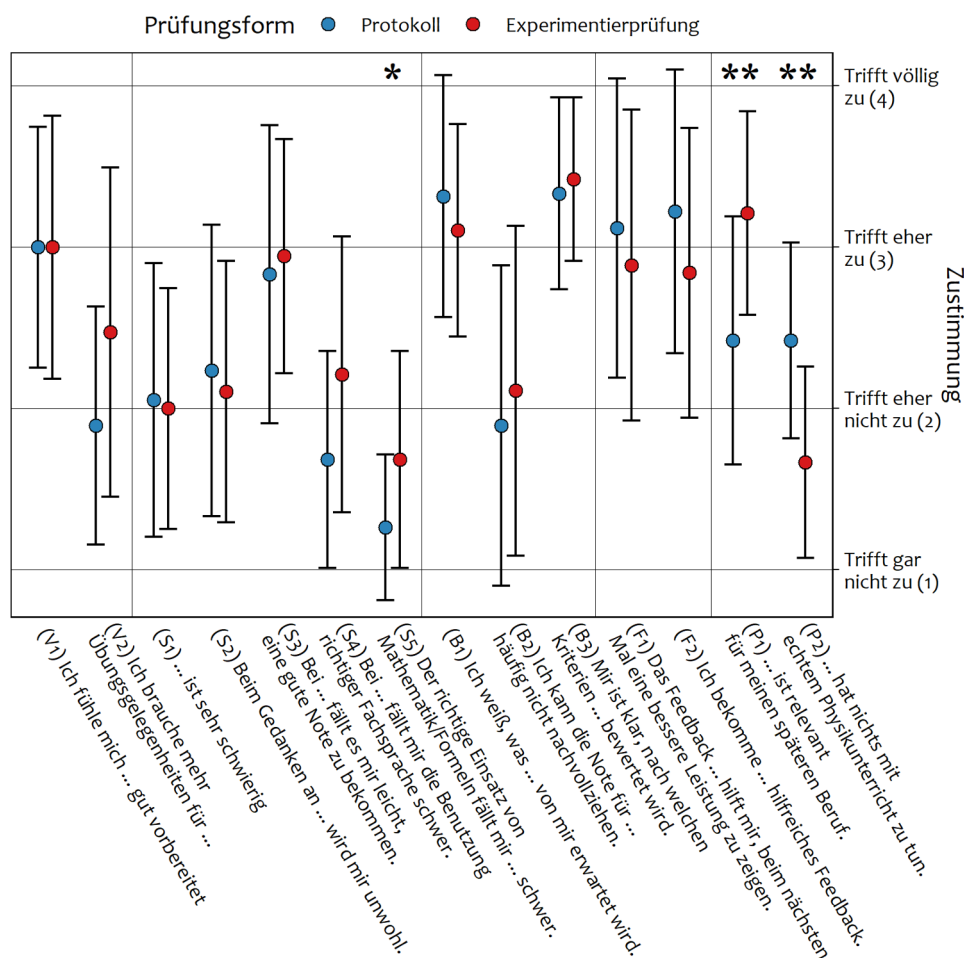


Abbildung 4: Zustimmung der Studierenden zu den Items aus Tabelle 1

Anmerkung: Angegeben sind jeweils das arithmetische Mittel und die Standardabweichung. Signifikanz im WMW-Test ist mit * ($p < 0.05$) und ** ($p < 0.01$) angegeben.

6.1.1 Teilfrage (a): Vorbereitung

Die Studierenden geben unabhängig von der Prüfungsform an, sich eher gut auf die Prüfung vorbereitet zu fühlen (V1). Für die videografierte Experimentierprüfung stimmen diesem Item lediglich 6 Probanden eher nicht zu, beim Versuchsprotokoll nur 5.

Die Frage, ob mehr Übungsgelegenheiten benötigt würden (V2), wird für die Versuchsprotokolle eher verneint (9 eher nicht, 9 gar nicht), bei den videografierten Experimentierprüfungen zeigt sich hier eine nicht signifikante Tendenz zu höherer Zustimmung mit vergleichsweise hoher Varianz. Zehn Proband:innen halten mehr Übungsgelegenheiten für ganz oder eher notwendig.

In den Freitext-Antworten wünschen sie sich vor allem mehr Gelegenheiten, das laute Ausprechen der Gedanken beim Experimentieren zu üben, sowie eine theoretische Auseinandersetzung damit, was beim schulischen Experimentieren eigentlich sinnvollerweise verbalisiert werden sollte. Insgesamt zeigt sich hier eher eine Unsicherheit mit dem freien Erklären physikalischer Inhalte. Diejenigen 9 Proband:innen, die zusätzliche Übungsgelegenheiten eher oder gar nicht für nötig halten, äußern diese Probleme nicht.

Das theoretisch angenommene *Constructive Alignment*, also die bessere Vorbereitung auf die videografierte Experimentierprüfung aufgrund deren Ähnlichkeit mit dem typischen Praktikums-geschehen, wird in den Daten also nicht sichtbar.

6.1.2 Teilfrage (b): Subjektive Schwierigkeit

Die Items S1 bis S5 lassen sich zu einer Likert-Skala zusammenführen (S3 invertiert), die für die videografierte Experimentierprüfung ein standardisiertes Cronbachs $\alpha = 0.83$ und für die Versuchsprotokolle von $\alpha = 0.77$ aufweist. Wie Abbildung 5 zeigt, werden beide Prüfungsformen auf dieser Gesamtskala als überraschend leicht eingeschätzt. Der Unterschied zwischen den Prüfungsformen ist im WMW-Test nicht signifikant. Die Schwierigkeitseinschätzungen korrelieren aber auch nicht wesentlich miteinander ($r = 0.34$, $p = 0.15$).

Betrachtet man die einzelnen Items (Abb. 4), so halten 14 von 19 Proband:innen die Experimentierprüfung und ebenso 14 Proband:innen das Schreiben von Versuchsprotokollen für eher oder gar nicht schwierig (S1). Der Frage, ob ihnen unwohl wird (S2), stimmen 14 bzw. 12 Proband:innen eher oder gar nicht zu. 13 bzw. 11 Studierende halten es für leicht, eine gute Note zu bekommen (S3). Auch im arithmetischen Mittel weisen die ersten drei Items S1, S2, S3 fast keine Unterschiede zwischen den Prüfungsformen auf.

In den Freitextantworten wird für beide Prüfungsformen auf den relativ geringen Prüfungsdruck hingewiesen, bei der videografierten Experimentierprüfung wird dies durch die „entspannte Atmosphäre“, bei den Versuchsprotokollen durch die Möglichkeit zur Zweitabgabe begründet. Bei beiden Prüfungsformen berichten Studierende aber auch von Zeitproblemen, weil die in der Experimentierprüfung zur Verfügung stehende Zeit eben nicht ganz ausreicht und die Versuchsprotokolle für die zum Schreiben zur Verfügung stehende Zeit zu lang seien.

Als einziges Item aus dieser Gruppe korreliert S3 mit den angegebenen Noten (Experimentierprüfung: $r = -0.82$, $p < 0.001$; Protokolle: $r = -0.86$, $p < 0.001$). Die Studierenden scheinen also durchaus in der Lage zu sein, zwischen der Schwierigkeit, eine gute Note zu bekommen, und anderen Herausforderungen der jeweiligen Prüfungsform zu unterscheiden.

Die Benutzung richtiger Fachsprache erscheint den Studierenden tendenziell bei der videografierten Experimentierprüfung als schwieriger. Dem Item S4 stimmen 7 Proband:innen für die videografierte Experimentierprüfung, aber nur 2 beim Versuchsprotokoll ganz oder eher zu. Der Mittelwertunterschied ist im WMW-Test nicht signifikant.

Die Studierenden weisen hier in den Freitextantworten mehrfach darauf hin, dass sowohl das Erklären physikalischer Sachverhalte als auch das Verfassen von Fachtexten von ihnen als relevantes Lernziel wahrgenommen werden.

Ein überraschender Unterschied findet sich in den Antworten auf die Frage nach den Schwierigkeiten im Umgang mit Formeln/Mathematik (S5). Beim Versuchsprotokoll stimmt kein:e Proband:in diesem Item zu (5 eher nicht, 14 gar nicht). Bei der videografierten Experimentierprüfung stimmen immerhin 2 Proband:innen eher zu (9 eher nicht, 8 gar nicht). Der Unterschied ist mäßig signifikant ($W = 242.5$, $p = 0.04$) mit mittelgroßer Effektstärke ($d = 0.74$). Die Studierenden haben mit der Mathematik in diesen Prüfungsformen also deutlich weniger Schwierigkeiten als antizipiert.

Bemerkenswerterweise geht keine einzige Freitextantwort auf die mathematischen Anforderungen in den beiden Prüfungsformen ein. Dies überrascht umso mehr, als dass dies zumindest für das Versuchsprotokoll sowohl aus Sicht der Lehrenden als auch der Literatur als zentrales, auch schwierigkeiterzeugende Merkmal dieser Prüfungsform gelten kann (Bornschein 2016; Riewerts 2016). Die größeren Schwierigkeiten, die Proband:innen hier bei der videografierten Experimentierprüfung berichten, gehen einher mit dem o. g. Befund, dass die Versprachlichung ganz allgemein als zu wenig geübt angegeben wird.

Während den Studierenden die unterschiedliche Art der Anforderungen in den beiden Prüfungsformen durchaus bewusst zu sein scheint, schlägt sich das insgesamt so gut wie nicht auf die wahrgenommene Schwierigkeit nieder.

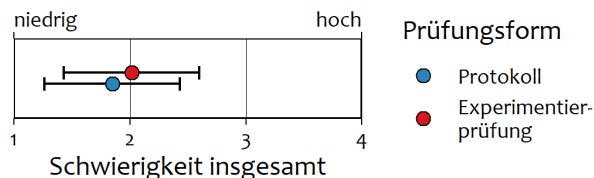


Abbildung 5: Aggregierte Schwierigkeit mit Mittelwert und Standardabweichung für beide Prüfungsformen

6.1.3 Teilfrage (c): Klarheit des Bewertungsmaßstabs

Die Klarheit des Bewertungsmaßstabs (Items B1 bis B3) wird insgesamt relativ hoch eingeschätzt. Bei jeder Prüfungsform geben lediglich 3 Proband:innen an, eher nicht zu wissen, was von ihnen erwartet würde (B1), und nur 1 Person sind die Kriterien für die Bewertung von Versuchsprotokollen eher nicht klar (B3; videografierte Experimentierprüfung: keine:r). Bemerkenswert ist die hohe Varianz bei der Nachvollziehbarkeit der Note (B2). Bei beiden Prüfungsformen können 4 Proband:innen die Note eher nicht nachvollziehen, wohingegen 5 (Experimentierprüfung) bzw. 8 (Versuchsprotokolle) dem Item gar nicht zustimmen, die Note also gut nachvollziehen können. Keiner der Mittelwertunterschiede zwischen den Prüfungsformen ist signifikant.

In den Volltext-Antworten wird auf diesen Aspekt nur wenig eingegangen. Eine Äußerung bemängelt, beim Protokoll nicht zu wissen, wann alle relevanten Punkte des theoretischen Hintergrundes ausreichend dargestellt seien. Eine andere Person spricht wahrgenommene Bewertungsunterschiede zwischen verschiedenen Korrektor:innen an, dies sei aber mit der Herausgabe eines Musterprotokolls besser geworden. Für die videografierte Experimentierprüfung wird lediglich angemerkt, dass die Aufgabenstellung zwischen der Übungssituation im Semester und der Prüfung stärker als angekündigt verändert würde.

Anders als hypothetisch formuliert, schätzen die Studierenden den Bewertungsmaßstab insgesamt also als überraschend klar ein – Kritik entzündet sich eher an der Anwendung dieses Maßstabs durch die jeweiligen Korrektor:innen.

6.1.4 Teilfrage (d): Feedbackqualität

Die Qualität des Feedbacks wird in beiden Items (bei relativ hoher Varianz) als relativ gut und nützlich eingeschätzt. Für das Feedback zur Experimentierprüfung bejahen 13 Personen eher oder völlig Item F1 und 12 Personen Item F2. Für die Versuchsprotokolle sind dies ebenfalls 13 (F1) bzw. 13 (F2). Der Unterschied zwischen den Prüfungsformen ist nicht signifikant. Die beiden Items korrelieren für beide Prüfungsformen hoch signifikant miteinander (Experimentierprüfung: $r = 0.66$, $p = 0.002$; Protokoll: $r = 0.74$, $p < 0.001$).

Die Freitextantworten kritisieren nicht so sehr die Form des Feedbacks, sondern eher den Umgang damit in der Lehrveranstaltung. So fehlt es in der Wahrnehmung der Studierenden häufig an einer Nachbesprechung, die sich an den zuvor ausgegebenen Kriterien auf der Checkliste orientiert. Mehrere Personen bemängeln insbesondere bei der videografierten Experimentierprüfung lediglich eine Note ohne Begründung mitgeteilt bekommen zu haben, diese Personen haben die Items F1 und F2 auch jeweils ganz oder eher abgelehnt. Das Potenzial einer strukturierten Rückmeldung wird hier also bei Weitem nicht ausgeschöpft.

6.1.5 Teilfrage (e): Praxisnähe

Die deutlichsten Unterschiede zwischen den Prüfungsformen zeigen die Ergebnisse (Abb. 4) in den beiden Items zur Praxisnähe. Die Antworten zum Versuchsprotokoll zeigen hier eine klare Tendenz zur Mitte. P1 stimmen 7 Proband:innen ganz oder eher zu, 12 lehnen die Relevanz für den späteren Beruf eher oder ganz ab. Das umgekehrt gepolte Item P2 lehnen 10 Proband:innen völlig oder eher ab, 9 stimmen eher zu. Die Zustimmung zu beiden Items korreliert hoch signifikant ($r = 0.64$, $p = 0.003$).

Die videografierte Experimentierprüfung wird deutlich positiver bewertet: Die Berufsrelevanz (P1) lehnen lediglich 2 Proband:innen eher ab. Nur 1 Proband:in sieht eher keinen Zusammenhang mit echtem Physikunterricht (P2). Die Zustimmung zu diesen Items korreliert hier nicht signifikant ($r = -0.26$, $p = 0.3$).

Die Zustimmung zu beiden Items unterscheidet sich hoch signifikant zwischen den Prüfungsformen mit sehr hoher Effektstärke (P1: $W = 280$, $p = 0.002$, $d = 1.23$; P2: $W = 73$, $p = 0.001$, $d = 1.26$).

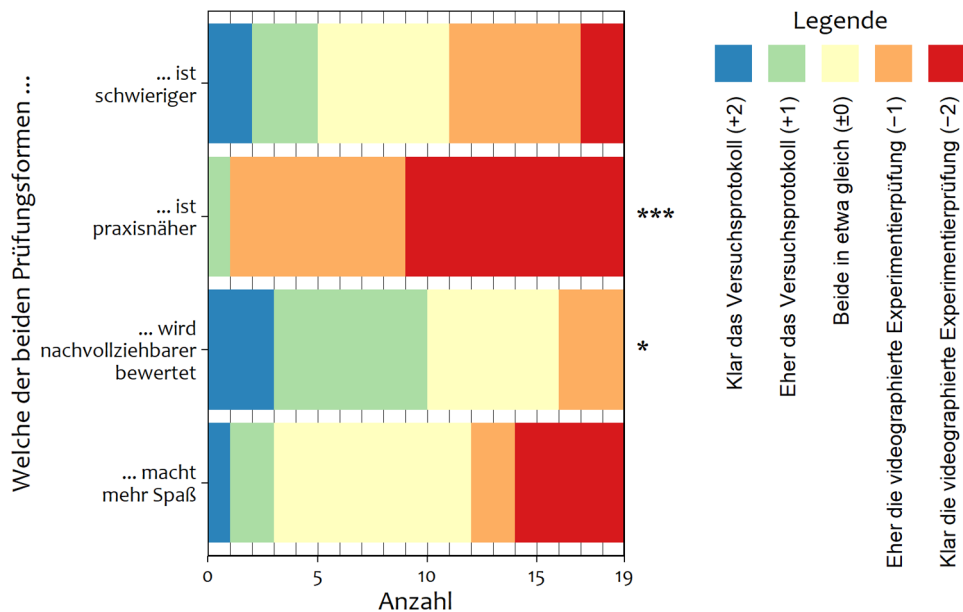


Abbildung 6: Direkter Vergleich der Prüfungsformen mit 5-stufigen Items

Anmerkung: Signifikante Abweichungen von einer mittigen Verteilung sind mit * ($p < 0.05$), ** ($p < 0.01$) und *** ($p < 0.001$) angegeben.

Die Freitext-Antworten adressieren die Praxisnähe nur selten, dann aber sehr deutlich: „Die Prüfungsform [videografierte Experimentierprüfung] beschäftigt sich mit der Durchführung und dem Erklären des präsentierten Experiments. Dies ist ein elementarer Teil des Physikunterrichts.“; „Das Verbalisieren der eigenen Handlungen [...] ist also eine von Lehrern zu erwartende Kompetenz, die im Rahmen des Studiums ansonsten nicht erworben werden würde. Außerdem halte ich es in einer Lehrveranstaltung, in der es um das Experimentieren geht, für sehr wichtig, dass auch die eigene Experimentierkompetenz und nicht etwa das Verfassen von Texten geprüft wird.“ Lediglich eine Rückmeldung zieht den Vergleich von universitären Versuchsprotokollen und schulischen Versuchsauswertungen, wie er in den theoretischen Begründungen für diese Prüfungsform (s. o.) zuweilen vorkommt.

Insgesamt gilt die videografierte Experimentierprüfung also (wie von den Lehrenden intendiert und hypothetisch angenommen) als wesentlich praxisnäher.

6.2 Direkter Vergleich

Nach den Items, die sich erst auf die eine, dann auf die andere Prüfungsform bezogen, wurden die Studierenden in fünf Items (D1 bis D4 und O2, vgl. Tabelle 1) um einen direkten Vergleich gebeten. Die Antworten zeigt Abbildung 6.

Um zu prüfen, ob die Studierenden-Rückmeldungen signifikant in Richtung einer Prüfungsform tendieren, wurden die Antworten numerisch von -2 bis $+2$ codiert ($-2 = \text{Klar die Experimentierprüfung}$, $+2 = \text{Klar das Versuchsprotokoll}$). Mit dem WMW-Test wird der Mittelwert dieser Antworten auf eine signifikante Abweichung von 0 geprüft. Fällt der Test signifikant aus, gehen wir davon aus, dass die Studierenden eine Prüfungsform z. B. für schwieriger halten als die andere.

Die Einschätzung, welche Prüfungsform schwieriger sei, ist hier sehr symmetrisch über beide Formen verteilt, was zum Vergleich der getrennt vorgenommenen Einschätzungen oben passt. Berechnet man Korrelationen zwischen der aggregierten Schwierigkeit (Abb. 5) mit dem direkten Vergleich, so fällt diese beim Versuchsprotokoll höchst signifikant aus ($r = 0.77$, $p < 0.001$), allerdings nicht mit der der videografierten Experimentierprüfung ($r = -0.09$, $p = 0.7$). Dabei ist die Varianz der beiden aggregierten Schwierigkeiten gleich groß (jeweils $SD = 0.58$). In den direkten Vergleich geht also vor allem die Wahrnehmung der Schwierigkeit der Versuchsprotokolle ein. Es

könnte vermutet werden, dass es sich hier um einen Kontrast-Effekt der „neuen“ gegenüber der „alten“ oder etablierten Prüfungsform handelt.

Wie schon bei den Einzeleinschätzungen oben, zeigt sich hier ein höchstsignifikanter Unterschied bei der Frage nach der Paxisnähe der Prüfungsformen. Lediglich ein:e Proband:in hält das Versuchsprotokoll im Vergleich für praxisnäher. Diese Person äußert in der Freitext-Antwort zum Versuchsprotokoll: „Versuchsprotokolle für Versuche sind schlicht und einfach ein wiss. Standard der eingehalten werden sollte. Jeder Physiker oder Techniker sollte wissen, wie man Laborberichte erstellt, da sie als Grundlage für jede Forschung dienen.“ Dies legt eine Orientierung dieser/dieses Probandin/Probanden an einer anderen „Praxis“, nämlich Wissenschaft/Technik und nicht Unterricht, nahe.

Unterschiede in der Nachvollziehbarkeit der Bewertung waren in den Einzeleinschätzungen nicht signifikant geworden. Im direkten Vergleich zeigt sich aber eine schwach signifikante Einschätzung zu Gunsten des Versuchsprotokolls. Die Personen, die hier „klar das Versuchsprotokoll“ angekreuzt haben, haben sich bei den Freitext-Antworten zur Experimentierprüfung dazu geäußert, dass die Rückmeldung dort häufig nicht über die Mitteilung der Note hinausginge und sich nicht an den Bewertungskriterien orientiere. Die bereits oben festgestellte Tendenz prägt sich also im direkten Vergleich lediglich stärker aus.

Die Frage, welche Prüfungsform „mehr Spaß“ mache, resultierte in keiner signifikanten Tendenz zu einer Prüfungsform. Es kann angenommen werden, dass bei dieser eher emotionalen und wenig kriterialen Bewertung eine ganze Reihe persönlicher Vorlieben eine Rolle spielt. In jedem Fall kann hier eine deutliche Bevorzugung einer der beiden Prüfungsformen durch die Studierenden aber nicht beobachtet werden.

Anschließend an diese vier geschlossenen Items wurde noch in einem offenen Item (O2) nach weiteren Unterschieden gefragt. Zwei Personen benannten hier den Zeitaspekt – während sich der Prüfungsstress bei der videografierten Experimentierprüfung auf ein Zeitfenster von 90 Minuten beschränkt, stehen zur Abfassung eines Versuchsprotokolls mehrere Tage zur Verfügung. Einerseits bedeutet das eine höhere Arbeitsbelastung, andererseits aber auch die Möglichkeit, mit Kommiliton:innen oder Lehrenden Rücksprache zu halten.

6.3 Gesamteinschätzung in Schulnoten

Nachdem die Studierenden zu jeder Prüfungsform befragt wurden, wurden sie jeweils um eine Gesamteinschätzung der Qualität der jeweiligen Prüfungsform als Schulnote inklusive Begründung gebeten. Dies bietet den Studierenden noch einmal die Möglichkeit, die verschiedenen Merkmale der jeweiligen Prüfungsform zu gewichten und zusammenfassend zu bewerten. Einen Überblick über die vergebenen Noten zeigt Abbildung 7.

Die Mittelwerte der Schulnoten betragen für die videografierte Experimentierprüfung 2.05 ($SD = 0.71$) und für die Versuchsprotokolle 2.47 ($SD = 0.90$). Der Unterschied ist im WMW-Test nicht signifikant ($W = 224.5$, $p = 0.169$). Die Noten beider Prüfungsformen korrelieren aber auch nicht signifikant ($r = 0.39$, $p = 0.09$).

Relevant erscheint nun, welche Kriterien die Befragten jeweils zur Begründung der Note heranziehen. Insgesamt wurde die Note für die videografierte Experimentierprüfung 14-mal und für das Versuchsprotokoll 12-mal begründet. Die Begründungen wurden induktiv kategorisiert, wobei jede Begründung einer oder mehreren Kategorien zugeordnet werden konnte. Einen gekürzten Überblick zeigt Tabelle 3, dabei sind Kategorien, die nur einmal vorkommen, jeweils zusammengefasst.

Im Vergleich zeigt sich, dass die videografierte Experimentierprüfung vor allem deshalb gut oder sehr gut bewertet wird, weil sie relevante Kompetenzen des Lehrberufs abbildet, wohingegen die guten und sehr guten Noten für das Versuchsprotokoll eher mit dessen Eigenschaft als wissenschaftlicher Standard und als Übungsgelegenheit für das wissenschaftliche Schreiben begründet werden. Schlechtere Noten werden bei der videografierten Experimentierprüfung vor allem mit der ungewohnten Prüfungssituation, für die es zudem zu wenige Übungsgelegenheiten gebe, be-

gründet. Bei dem Versuchsprotokoll sticht hier eher die geringe Relevanz für den Lehrberuf heraus.

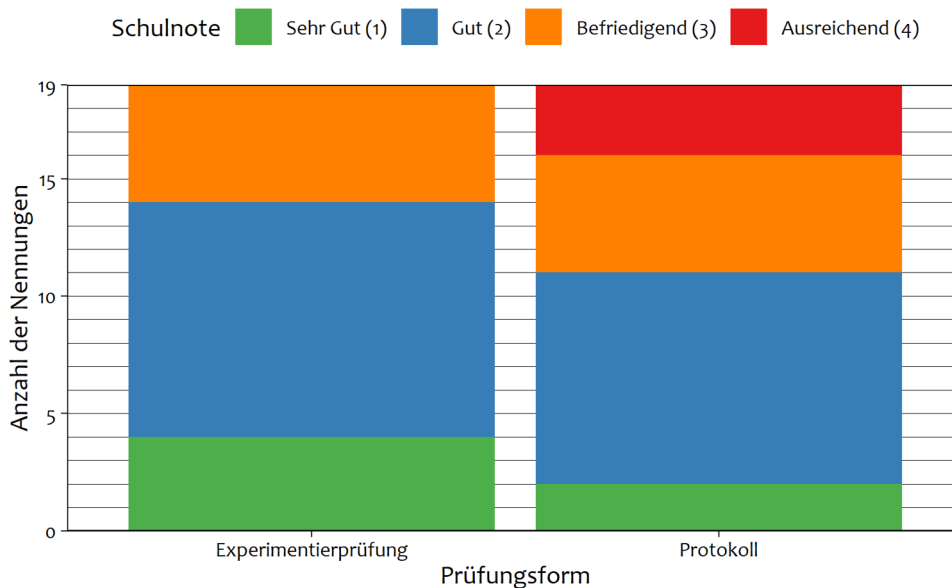


Abbildung 7: Vergebene Schulnoten für die beiden Prüfungsformen

Tabelle 3: Kategorial zusammengefasste Begründungen für die Noten in Abbildung 7

Note	Experimentierprüfung	Versuchsprotokoll
Sehr Gut (1)	Wichtige Kompetenzen für Lehrberuf (3x)	Standardformat, wiss. Schreiben (3x)
Gut (2)	Prüfungssituation, Übungsmöglichkeiten (4x)	Standardformat, wiss. Schreiben (3x)
	Wichtige Kompetenzen für Lehrberuf (3x)	Geringe Relevanz für Lehrberuf (1x)
	Unklare Bewertungskriterien (1x)	
	Sonstige Einzel-Begründungen (2x)	Sonstige Einzel-Begründungen (3x)
Befriedigend (3)	Prüfungssituation, Übungsmöglichkeiten (3x)	Standardformat, wiss. Schreiben (1x)
		Geringe Relevanz für Lehrberuf (1x)
		Unklare Bewertungskriterien (1x)
Ausreichend (4)	–	Geringe Relevanz für Lehrberuf (1x)

Anmerkung: Mehrfachzuordnungen sind möglich.

7 Zusammenfassung und Ausblick

Die beiden „großen“ Prüfungsformate in den schulbezogenen Experimentierpraktika, das Versuchsprotokoll und die videografierte Experimentierprüfung prüfen aufgrund ihrer unterschiedlichen Gestaltung theoretisch erwartbar unterschiedliche Kompetenzen ab. Dies ist auch vielen Studierenden klar. Ein:e Proband:in äußert dazu: „Die Prüfungsformen unterscheiden sich maßgeblich daran, welche Kompetenzen sie überprüfen (sollen). Bei der Experimentierprüfung geht es

vor allem um Experimentierkompetenzen, implizit wird aber besonders auch die Fähigkeit zum lauten Denken geprüft. Bei Versuchsprotokollen geht es dagegen vor allem um die theoretische und schriftliche Auseinandersetzung mit dem Versuch und dessen Präsentation, implizit werden hier aber vor allem sprachliche und textgestalterische Fähigkeiten geprüft.“

In der berichteten Befragung wurde dazu die studentische Sicht auf diese beiden Prüfungsformen in verschiedenen Kategorien erfragt. Die Online-Befragung erreichte 19 Studierende aus 3 Lehrveranstaltungen und bildet somit eine typische Kohortengröße ab. Aufgrund der stetigen Veränderung der Lehrveranstaltungen und erwartbaren Lehrenden-Effekten kann jedoch nicht von allzu großer Repräsentativität ausgegangen werden.

Zentraler Unterschied zwischen den Prüfungsformen ist die Relevanz in Bezug auf den späteren Lehrberuf, wo die videografierte Experimentierprüfung durch die Studierenden deutlich positiver bewertet wird. Das Versuchsprotokoll wird hingegen vor allem von den Personen für relevant gehalten und positiv bewertet, die es als wissenschaftlichen Standard erkennen und daher eine (anders bezogene) Relevanz oder zumindest Notwendigkeit herleiten.

Eine weitere, in die Gesamtschau auf die Prüfungsformen eingehende Kategorie ist die Einbettung in die vorherige Lehrveranstaltung, also der vorhandenen Übungsmöglichkeiten vorher und des Feedbacks nachher. In beiden Kategorien zeigt sich bei einem Teil der Studierenden Tendenzen, die videografierte Experimentierprüfung schwächer zu beurteilen. Es fehle vielfach an Übungsmöglichkeiten, vor allem für das erklärende laute Denken. Die Bewertung werde häufig nicht anhand der vorhandenen Kriterien anhand des aufgezeichneten Videos erläutert.

Zusammenfassend sind die Unterschiede zwischen den Prüfungsformen aus Sicht der Studierenden bei Weitem nicht so deutlich ausgeprägt wie hypothetisch vermutet. Insbesondere die wahrgenommene Schwierigkeit zeigt kaum Unterschiede, obwohl hier gezielt nach antizipierten schwierigkeiterzeugenden Merkmalen gefragt wurde. Lediglich das Erklären mathematischer Sachverhalte fällt den Studierenden signifikant schwerer, was aber ein Nebeneffekt mangelnder Übung sein kann.

Bei einer weiteren Überarbeitung der schulbezogenen Experimentierpraktika im Department Physik der Universität Paderborn werden die videografierten Experimentierprüfungen sicherlich nicht die schriftlichen Versuchsprotokolle ersetzen – dies ist aufgrund ihres unterschiedlichen Profils aus Lehrendensicht auch kaum wünschenswert, zumal die Befragung zeigt, dass auch unter den Studierenden beide Prüfungsformen Anhänger:innen finden. Nächste Schritte sind vielmehr einerseits die Anpassung des Feedbacks zu den Experimentierprüfungen – hier müssen Wege gefunden werden, wie den Studierenden kriteriengeleitet verständlicheres Feedback gegeben werden kann – sowie die Schaffung gezielterer Übungsformate, die die als schwierig empfundene Versprachlichung physikalischer Erklärungen üben. Weitere Erhebungen könnten z. B. Auskunft darüber geben, ob die hier retrospektiv aufgenommenen Wahrnehmungen mit denen in oder kurz vor/nach der Prüfungssituation übereinstimmen. Eine längsschnittliche Erhebung könnte Auskunft über Lern- oder Gewöhnungseffekte bei beiden Formen geben. Eine Überarbeitung des Testinstruments könnte darüber hinaus erfragen, inwiefern diese Prüfungssituationen auch als Lernsituation wahrgenommen werden, ob z. B. eines der beiden Formate zu einer intensiveren Auseinandersetzung mit den fachlichen Inhalten führt.

Literatur

- Anderson, L. W., & Krathwohl, D. R. (Hrsg.). (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Bauer, A. B., & Sacher, M. D. (2018). Kompetenzorientierte, universitäre Laborpraktika: Das Paderborner Physik Praktikum (3P). *PhyDid B – Didaktik der Physik – Beiträge zur DPG-Frühjahrstagung Würzburg*, 65–72.
- Biggs, J., & Tang, C. (2011). *Teaching for Quality Learning at University: What the Student Do* (4. Aufl.). Maidenhead: Open University Press.

- Bornschein, B. (2016). Schreibausbildung in der Physik: Erste Erfahrungen am Schreiblabor des House of Competence. In A. Hirsch-Weber & S. Scherer (Hrsg.), *Wissenschaftliches Schreiben in Natur- und Technikwissenschaften: Neue Herausforderungen der Schreibforschung* (S. 143–154). Wiesbaden: Springer Spektrum.
- Cordes, N. (2017). Versuchsprotokolle in der Lehre. *Deutsche Universitätszeitung*, 73(4), 81–83.
- Dubs, R. (2011). Besser schriftlich prüfen: Prüfungen valide und zuverlässig durchführen. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre: Lehren und Lernen effizient gestalten*. Berlin: Raabe.
- Haller, K. (1999). *Über den Zusammenhang von Handlungen und Zielen: Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*. Berlin: Logos.
- Hollander, M., & Wolfe, D. A. (1973). *Nonparametric statistical methods. Wiley series in probability and mathematical statistics: Applied probability and statistics*. New York: Wiley.
- Hucke, L. (1999). *Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums*. Berlin: Logos.
- Hüther, M. (2005). *Evaluation einer hypermedialen Lernumgebung zum Thema Gasgesetze: Eine Studie im Rahmen des Physikpraktikums für Studierende der Medizin*. Berlin: Logos.
- KMK, Ständige Konferenz der Kultusminister der Länder in der Bundesrepublik Deutschland (2008). *Ländergemeinsame inhaltliche Anforderungen für die Fachwissenschaften und Fachdidaktiken in der Lehrerbildung*. Bonn.
- Krabbe, H. (2015). Das Versuchsprotokoll als fachtypische Textsorte des Physikunterrichts. In S. Schmörlzer-Eibinger & E. Thürmann (Hrsg.), *Schreiben als Medium des Lernens: Kompetenzentwicklung durch Schreiben im Fachunterricht* (S. 57–174). Münster: Waxmann.
- Kulgemeyer, C., Riese, J., Borowski, A., Schreiber, N., & Vogelsang, C. (2018). Performanztests in der naturwissenschaftlichen Lehrerbildung. In C. Maurer (Hrsg.), *Qualitätsvoller Chemie- und Physikunterricht: Normative und empirische Dimensionen* (S. 46–49). Regensburg: Universität Regensburg.
- Miller, G. E. (1990). The Assessment of Clinical Skills/Competence/Performance. *Academic Medicine*, 65(9), 63–67.
- Nagel, C. (2009). *eLearning im physikalischen Anfängerpraktikum*. Berlin: Logos.
- Neumann, K. (2004). *Didaktische Rekonstruktion eines physikalischen Praktikums für Physiker*. Berlin: Logos.
- Präsidium der Universität Paderborn (Hrsg.). (2016a). *Besondere Bestimmungen der Prüfungsordnung für den Bachelorstudiengang Lehramt an Haupt-, Real-, Sekundar- und Gesamtschulen mit dem Unterrichtsfach Physik an der Universität Paderborn. Amtliche Mitteilungen* (Bd. 142.16). Paderborn: Universität Paderborn.
- Präsidium der Universität Paderborn (Hrsg.). (2016b). *Besondere Bestimmungen der Prüfungsordnung für den Bachelorstudiengang Lehramt an Gymnasien und Gesamtschulen mit dem Unterrichtsfach Physik an der Universität Paderborn. Amtliche Mitteilungen* (Bd. 144.16). Paderborn: Universität Paderborn.
- Präsidium der Universität Paderborn (Hrsg.). (2017). *Besondere Bestimmungen der Prüfungsordnung für den Masterstudiengang Lehramt an Gymnasien und Gesamtschulen mit dem Unterrichtsfach Physik an der Universität Paderborn. Amtliche Mitteilungen* (Bd. 92.17). Paderborn: Universität Paderborn.
- Riewerts, K. (2016). Schreiben und Naturwissenschaften in der Hochschule: Unvereinbare Gegensätze oder fruchtbare Zusammenarbeit? In A. Hirsch-Weber & S. Scherer (Hrsg.), *Wissenschaftliches Schreiben in Natur- und Technikwissenschaften: Neue Herausforderungen der Schreibforschung* (S. 109–120). Wiesbaden: Springer Spektrum.
- Sacher, M. D., Probst, H. M., & Reinhold, P. (2015). Neue Wege zur Konzeption eines kompetenzfördernden, studierendenzentrierten Laborpraktikums. In Zentrum für Hochschuldidaktik (Hrsg.), *Tagungsband zum 2. HDMINT Symposium 2015: DiNa-Sonderausgabe* (S. 106–113). Ingolstadt: Technische Hochschule Nürnberg Georg Simon Ohm.
- Sacher, M. D., Probst, H. M., Reinhold, P., & Schaper, N. (2015). Entwicklung eines kompetenzorientierten physikalischen Laborpraktikums. In S. Hartz & S. Marx (Hrsg.), *Leitkonzepte der Hochschuldidaktik: Theorie – Praxis – Empirie* (S. 128–136). Bielefeld: Bertelsmann Verlag.
- Sander, F. (2000). *Verbindung von Theorie und Experiment im physikalischen Praktikum: Eine empirische Untersuchung zum handlungsbezogenen Vorverständnis und dem Einsatz grafikorientierter Modellbildung im Praktikum*. Berlin: Logos.
- Schaper, N., Hilkenmeier, F., & Bender, E. (2013). *Umsetzungshilfen für kompetenzorientiertes Prüfen: HRK-Zusatzgutachten*. Bonn: Hochschulrektorenkonferenz.
- Schröder, M. (2015). *Kompetenzorientiert prüfen: Zum Lernergebnis passende Prüfungsaufgaben. Nexus Impulse für die Praxis* (Bd. 4). Bonn: Hochschulrektorenkonferenz.

- Szabone Varnai, A., & Reinhold, P. (2017). Experimentelle Handlungen von Lehramt-Studierenden im physikalischen Praktikum. In C. Maurer (Hrsg.), *Implementation fachdidaktischer Innovation im Spiegel von Forschung und Praxis* (S. 728–731). Regensburg: Universität Regensburg.
- Tesch, M., & Duit, R. (2004). Experimentieren im Physikunterricht – Ergebnisse einer Videostudie. *Zeitschrift für Didaktik der Naturwissenschaften*, 10, 51–69.
- Theyßen, H. (1999). *Ein Physikpraktikum für Studierende der Medizin: Darstellung der Entwicklung und Evaluation eines adressatenspezifischen Praktikums nach dem Modell der Didaktischen Rekonstruktion*. Berlin: Logos.
- Tiemann, R., & Körbs, C. (2014). Die Fragebogenmethode, ein Klassiker der empirischen didaktischen Forschung. In D. Krüger, I. Parchmann & H. Schecker (Hrsg.), *Methoden in der naturwissenschaftsdidaktischen Forschung* (S. 283–295). Berlin: Springer.
- Wass, V., Vleuten, C. van der, Shatzer, J., & Jones, R. (2001). Assessment of clinical competence. *The Lancet*, 357, 945–949.
- Welzel, M., Haller, K., Bandiera, M., Hammelev, D., Koumaras, P., Niedderer, H., Paulsen, A., Bécu-Robinault, K., & Aufschnaiter, S. v. (1998). Ziele, die Lehrende mit dem Experimentieren in der naturwissenschaftlichen Ausbildung verbinden: Ergebnisse einer europäischen Umfrage. *Zeitschrift für Didaktik der Naturwissenschaften*, 4(1), 29–44.
- Westphal, W. H. (1938). *Physikalisches Praktikum: Eine Sammlung von Übungsaufgaben für die physikalischen Übungen an Universitäten und Hochschulen aller Gattungen*. Wiesbaden: Vieweg+Teubner Verlag.
- Wildt, J., & Wildt, B. (2011). Lernprozessorientiertes Prüfen im „Constructive Alignment“: Ein Beitrag zur Förderung der Qualität von Hochschulbildung durch eine Weiterentwicklung des Prüfungssystems. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre: Lehren und Lernen effizient gestalten*. Berlin: Raabe.
- Wunderlich, A., & Szczyrba, B. (2011). Kompetenzorientiertes Prüfen – transparent, komplex und fair. In B. Berendt, A. Fleischmann, N. Schaper, B. Szczyrba, M. Wiemer & J. Wildt (Hrsg.), *Neues Handbuch Hochschullehre: Lehren und Lernen effizient gestalten*. Berlin: Raabe.
- Zastrow, M. U. (2001). *Interaktive Experimentieranleitungen: Entwicklung und Evaluation eines Konzeptes zur Vorbereitung auf das Experimentieren mit Messgeräten im physikalischen Praktikum*. Berlin: Logos.

Autor

Dr. David Woitkowski. Ehem. Universität Paderborn, Didaktik der Physik, Paderborn; E-Mail: david.woitkowski@gmail.com



Zitiervorschlag: Woitkowski, D. (2021). Videografiertes Experimentieren als Ergänzung zum Versuchsprotokoll – Studentische Wahrnehmung eines performativen Prüfungsverfahrens. *die hochschullehre*, Jahrgang 7/2021. DOI: 10.3278/HSL2132W. Online unter: wbv.de/die-hochschullehre