

Der PraQ: Ein Instrument zur Einschätzung von Praktikumsqualität – Skriptqualität

Daniel Rehfeldt, Volker Brüggemann & Volkhard Nordmeier

Freie Universität Berlin, Fachbereich Physik, Arnimallee 14, 14195-Berlin

daniel.rehfeldt@fu-berlin.de | volker.brueggemann@fu-berlin.de | volkhard.nordmeier@fu-berlin.de

Kurzfassung

Um naturwissenschaftliche Experimentierpraktika verlässlich evaluieren zu können, wurde an der FU Berlin der PraQ-Fragebogen entwickelt. Dieses Instrument dient der fächerübergreifenden Qualitätssicherung von Laborpraktika und befindet sich derzeit in der Validierungsphase. Als Teil des Validierungsprozesses wurden die Teilskalen »Skriptqualität für die Versuchsvorbereitung« und »allgemeine Skriptqualität« in einer quantitativen Studie mit Studierenden und ExpertInnen untersucht. Für beide Skalen konnte so erfolgreich eine Inhaltsvalidierung vorgenommen werden. Bezüglich der Konstruktvalidität lieferten die Ergebnisse Hinweise auf mögliche Optimierungen im weiteren Entwicklungsprozess.

1. Einleitung

Seit Beginn des Bologna-Prozesses ist eine regelmäßig stattfindende Qualitätssicherung an Hochschulen üblich [1]. Leider mangelt es noch häufig an der theoretischen Fundierung der dafür verwendeten Evaluierungsinstrumente [2]. Für verschiedene Lehrformen wie Vorlesungen und Seminare wurden bereits neue Messinstrumente erstellt und abschließend validiert, die fächerübergreifend eine Qualitätssicherung ermöglichen [3] [4].

In diesem Prozess wurden naturwissenschaftliche Praktika zunächst nicht beachtet, obwohl diese elementarer Bestandteil der experimentellen Ausbildung in den Naturwissenschaften sind [5].

Aus diesem Grund wird an der FU Berlin an der Konstruktion eines fächerübergreifenden Evaluationskonzeptes für naturwissenschaftliche Praktika gearbeitet. Hierzu sollen die Qualitätsdimensionen *Lernzuwachs*, *Lehrkompetenz* und *Lernumgebung* für naturwissenschaftliche Experimentalpraktika erfasst werden, die sich jeweils in diverse Teildimensionen aufspalten lassen [6] [7].

Als Operationalisierung wurde der Fragebogen PraQ (früher PraKo [6]) entwickelt: Das Instrument besteht aus 140 Items in 40 Skalen, die in zwei Fragebögen unterteilt sind [8].

Zurzeit befindet sich PraQ in der Validierungsphase. Das bedeutet unter anderem, dass die verschiedenen Skalen als Bestandteile des Gesamtinstruments zu validieren sind. Teil davon war eine Studie, die zwei Skalen des Fragebogens PraQ-B zum Thema *Skriptqualität* auf Inhalts- und Konstruktvalidität prüfen sollte.

2. Didaktischer Hintergrund

2.1 Skriptqualität

Als Grundlage für die Bildung von Hypothesen bezüglich der Skalen und eine angemessene Planung

der Studie musste zunächst festgestellt werden, welche verschiedenen Aspekte in dem Konstrukt *Skriptqualität* enthalten sind.

Die Literaturrecherche blieb hierbei aber zunächst weitestgehend erfolglos. Aktuelle Forschungen zur Qualität von Experimentierpraktika befassen sich vor allem mit dem Praktikum als Ganzem oder stellen multimediale Alternativen zum klassischen Skript vor und kritisieren dabei vor allem das Alter des Skripts [9], aber keine genauen Kriterien, die die Wirksamkeit des Skripts als Lernbegleiter betreffen. Detailliert erarbeitete Qualitätsmerkmale, die sich allein auf ein klassisches Skript beziehen, stehen nicht im Fokus dieser Arbeiten.

Zur Erarbeitung solcher Qualitätskriterien wurden für die Fragebogen-Entwicklung Bezüge des Praktikums skripts zum Medium »Lehrbuch«, zu den anerkannten Zielen von Praktika und zur Motivationspsychologie von Medien gezogen. Der Bezug zum Lehrbuch wurde über einen strukturellen sowie einen funktionalen Vergleich der beiden Medien hergestellt. Ersteres betrifft den Aufbau und die strukturierenden Elemente von Skripten, letzteres die Aufgaben, die mit dem Skript verfolgt werden sollen. Die Struktur von Skripten wurde anhand von aktuellen Praktikumsmaterialien analysiert und mit der Beschreibung struktureller Elemente (Text, Aufgaben, Abbildungen...) nach Merzlyn [10] verglichen. Dabei zeigte sich, dass die verwendeten Elemente nicht nur dieselben sind, sondern auch zu sehr ähnlichen Zwecken eingesetzt werden. Auch im Vergleich mit älteren Beschreibungen [11] fanden sich die gleichen Parallelen.

Um die durch ein Skript erfüllbaren Aufgaben bestimmen zu können, wurde die gleiche Analogie genutzt. Hierzu mussten die durch strukturell ähnliche Medien, also Lehrbücher, erfüllbaren Funktionen [12] mit den Anforderungen an Experimentierpraktika als Ganzem [13] verglichen werden. Die Schnittmenge dieser Aufgaben sollte diejenigen Anforderungen darstellen, die an ein Praktikum gestellt

werden und die durch ein zugehöriges Skript auch tatsächlich erfüllt werden können.

2.2 Kriterien der Skriptqualität

Die soeben beschriebenen, möglichen Funktionen eines Skriptes wurden nun in Kriterien umformuliert und stellen damit den ersten Teil eines entstehenden Kriterienkataloges für die Beurteilung von Skripten dar. Da diese Liste sich jedoch ihrer Entwicklung entsprechend vor allem an den Zielen von Praktika orientiert, wurde noch kaum auf die Qualität als textliches Medium eingegangen. Um dies zu gewährleisten, wurde ein weiterer Abschnitt hinzugefügt. Dieser stellt Qualitätskriterien aus der Lehrbuchforschung dar und orientiert sich an geforderten Gestaltungsmerkmalen aus der zugehörigen Literatur [10] [14] [15].

Als letzter Beitrag für den Katalog diente die Arbeit zur Motivationspsychologie nach Prenzel [16]. Der Grund hierfür war, dass Forderungen nach Motivation der Lernenden sowohl in den Praktikumszielen als auch in der Lehrbuchforschung verankert sind, diese jedoch sehr vage blieben. Dementsprechend wurden die von Prenzel genannten »sechs Formen der Demotivation« in positive Merkmale umformuliert. Von diesen wurden drei als durch einen lehrbuchartigen Text erfüllbare Kriterien in den Katalog aufgenommen.

2.3 Vergleich des Katalogs mit den Evaluationskalen

Der fertige Katalog diente als theoretische Ausgangslage für einen Vergleich mit den beiden zu prüfenden Skalen des PraQ-B Fragebogens. Dabei wurde untersucht, inwieweit die Items des Fragebogens und der Kriterienkatalog sich überschneiden und an welchen Stellen eventuelle Abweichungen bestanden. Abweichungen zwischen den theoretischen Qualitätsaspekten eines Skriptes und den durch die Skalen gemessenen Konstrukten mussten als eine von zwei möglichen Gefährdungen der Inhaltsvalidität nach Messick [17] angesehen werden: Gefährdungen erster Art liegen dann vor, wenn unnötige oder irrelevante Aspekte der Skriptqualität mit in dessen Bewertung eingehen (»construct irrelevant variance«). Dies hätte dementsprechend eine weniger gültige Evaluation des Skripts zur Folge, als es der Realität entspräche, Verzerrungen sowohl negativer wie positiver Art wären die Folge. Gefährdungen zweiter Art dagegen liegen vor, wenn wesentliche, qualitätsbestimmende Merkmale der Skriptqualität nicht oder unzureichend in die Bewertung eingehen (»construct underrepresentation«). Dies würde dazu führen, dass Skripte nur auf Basis eines nichtrepräsentativen Ausschnitts ihrer Qualität evaluiert würden, Verzerrungen der Interpretation sind hier wahrscheinlich.

Der qualitative Vergleich lieferte nur eine eindeutige Abweichung: Für das letzte Item der Skala »allgemeine Skriptqualität« aus dem PraQ konnte im Kriterienkatalog keine Entsprechung gefunden werden

(»Das Skript macht den Stellenwert der Sachverhalte im Praktikum für das weitere Studium klar.«).

Im Gegenzug konnten für jedes dem Katalog und damit der Theorie entstammende Kriterium Items der beiden Skalen gefunden werden, die anscheinend den gleichen Qualitätsaspekt erfassen (vgl. [18] S.28).

Dieser qualitative Vergleich von theoretisch erarbeitetem Kriterienkatalog und den zwei Skriptskalen lieferte also einen Beitrag zur Einschätzung der inhaltlichen Gültigkeit des Bewertungsinstrumentes PraQ-B.

Unklar bleibt hier allerdings, (a) inwieweit evaluierende Studierende der Praktika tatsächlich valide Einschätzungen in den einzelnen Aspekten vornehmen können. Sowohl subjektive Präferenzen, als auch mögliche mangelnde Expertise mit guten Lehrwerken lassen hier auch gegenteiliges vermuten. Zudem ist unklar, ob (b) die formulierten Indikatoren auch alle von Lehrwerkexperten als relevant bzw. gleich-relevant für Skriptqualität angesehen werden. Nicht-relevante Einschätzungen würden hier die theoretischen Vorarbeiten in ihrer Interpretation umformen (Gefährdung der Inhaltsvalidität erster Art), heterogen relevante Indikatoren sprächen für die Notwendigkeit einer quantitativen Gewichtung der Indikatoren, bevor ein gültiger Gesamtscore für die Skriptqualität berechnet werden könnte. Zu diesem Zweck wurde eine quantitative Validierungsstudie mit Studierenden und ExpertInnen durchgeführt, die im Folgenden beschrieben wird, beginnend mit den aus den vorherigen Abschnitten generierten Hypothesen.

3. Hypothesen

Das vorrangige Ziel der Studie war die Überprüfung der beiden Skalen auf Inhaltsvalidität. Außerdem wurde erwartet, dass die Auswertung der Daten Schlussfolgerungen zur Konstruktvalidität ermöglichen würde.

Zu beiden Aspekten ergaben sich aus der theoretischen Vorarbeit Hypothesen, die in der Studie zu prüfen waren. Zur Inhaltsvalidität wurde folgendes angenommen:

- **H1.1** Alle Frage-Items bis auf das letzte Item der Skala *allgemeine Skriptqualität* (»Stellenwert«) sind relevant für die Erfassung von Skriptqualität.
- **H1.2** Das letzte Frage-Item (»Stellenwert«) ist irrelevant.
- **H1.3** Die Skalen erfassen alle relevanten Aspekte der Skriptqualität.

Bezogen auf die Konstruktvalidität wurden zwei Annahmen gemacht:

- **H2.1** Die Gesamtbeurteilungen des Skriptes der Studierenden im Praktikum stimmen mit ExpertInnen-Beurteilungen überein.
- **H2.2** Die Beurteilungen des Skriptes in Einzelaspekten folgen in beiden Gruppen dem gleichen Antwortmuster.

4. Planung der Studie

4.1 Methode und Probandenauswahl

Wie aus den theoretischen Ausführungen hervorgeht, konnte bei den ProbandInnen nicht von der Kenntnis eines überprüften Standards zur Skriptqualität ausgegangen werden. Daher wurden ExpertInnen befragt. Es sollte sichergestellt werden, dass die gewonnenen Daten trotz der Umstände durch ein möglichst umfangreiches Theorie- und Handlungswissen im relevanten Themenfeld entstanden. Als Stichprobe wurden $N = 30$ wissenschaftliche MitarbeiterInnen aus Arbeitsgruppen deutscher Hochschulen ausgesucht, welche sich mit Physikdidaktik beschäftigten, also mit Theorien des Lehrens und Lernens, sowie des Einsatzes von Medien zur Unterstützung des Lernens von Physik vertraut sind. Von PraktikumsleiterInnen wurde bewusst abgesehen, da diese erstens eine gewisse Befangenheit bezüglich ihrer eigenen beruflichen Tätigkeit aufweisen könnten und zweitens i.d.R. keine didaktische Grundausbildung aufweisen, die für die Beurteilung der Qualität eines Mediums als Lerngegenstand allerdings notwendig ist.

Bei der ausgewählten Methode handelte es sich um ein Expertenrating. Die einzelnen Frage-Items der Skalen wurden durch die ExpertInnen auf ihre Relevanz für die Skriptqualität hin eingeschätzt.

Um zudem auch Daten für die Prüfung der Konstruktvalidität erhalten zu können, fand in der Studie eine weitere Befragung statt: Parallel zur Einschätzung der Skalenitems wurden die Skalen auch an einem exemplarisch ausgewählten Auszug eines Praktikumskriptes angewandt. Die so gewonnenen Beurteilungen des Skriptauszuges sollten später verglichen werden mit denen aus einer zweiten Probandengruppe: Das Skript wurde durch die am zugehörigen Praktikum teilnehmenden Studierenden beurteilt, hier anhand der zu prüfenden Skalen in ihrer Originalform. Die Probandengruppe der Studierenden umfasste 42 Personen, die ein physikalisches Grundpraktikum im Physikstudium (Mono) besuchten.

4.2 Fragebogendesign

Die beiden Skalen wurden im Fragebogen in ihrer Form exakt übernommen, sowohl um eine Einschätzung der Skalenitems ohne Verfälschungen zu ermöglichen, als auch um bei der Beurteilung des Skriptes durch ExpertInnen und Studierende vergleichbare Daten zu erhalten.

Bei der Expertenbefragung wurde allerdings zur Messung der Inhaltsvalidität eine weitere Skala hinzugefügt. Dabei handelte es sich um eine sechsstufige Likert-Skala, die den Fragebogen in einer weiteren Spalte neben den Originalskalen ergänzte. Die ExpertInnen mussten hier einschätzen, ob die in den Skalenitems gemachten Aussagen eher unwichtig (Werte 1–3) oder wichtig (Werte 4–6) für die Skriptqualität sind. Im Hinblick auf die später gewählten Auswertungsverfahren wurde bewusst auf die Möglichkeit verzichtet, eine neutrale Antwort zu geben.

Weiterhin beinhaltete der Fragebogen im Anschluss an die Skalen eine Kommentarzeile. Hierbei wurde dazu aufgefordert, bei vorliegender Kritik an einzelnen Items der Skala diese auszuformulieren. Auch sollten Ergänzungen gemacht werden, falls Aspekte der Skriptqualität nicht in der Skala erfasst worden wären.

5. Auswertung

5.1 Inhaltsvalidität

Als erster Teil der Auswertung wurden die Expertenratings der Skalenitems bezüglich ihrer Wichtigkeit für Skriptqualität betrachtet. Als Maß für die Relevanz der Items sollte der Median der Ratings dienen: Bei einem Median von 4 oder mehr wurde das betreffende Item von der Mehrheit der Probanden für relevant befunden.

Wie Abbildung 1 zu entnehmen ist, lag der Median der Ratings bei allen Skalenitems bei einem Wert von 4 oder mehr, sie wurden also alle für relevant befunden. Darüber hinaus ergab die qualitative Auswertung des Kommentaritems (»Kritisieren und Ergänzen«), dass die ExpertInnen vor allem ihre bereits quantitativ ausgefüllten Einschätzungen verbalisierten. Darüber hinausgehende Kritik oder Ergänzungsvorschläge für etwaige neue Indikatoren wurden nicht vollzogen, allerdings wurde das Item auch nur selten zur Kommentierung genutzt.

Der nächste Schritt der Auswertung bestand in der Analyse der Kommentierungen durch die ExpertInnen. Hier ist festzustellen, dass sich keine der vorgenommenen Anmerkungen auf eine Anpassung einzelner Skalenitems oder nicht erfasste Aspekte der Qualität bezog.

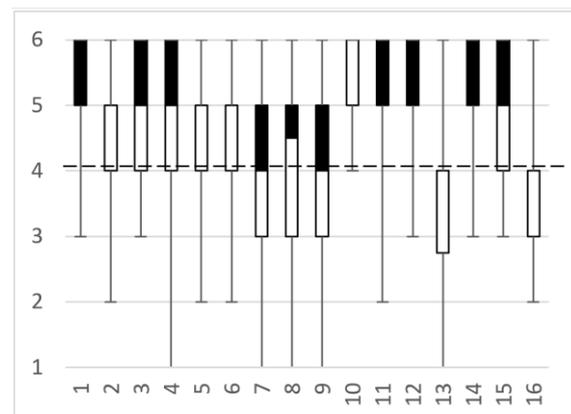


Abb. 1: Expertenratings im Boxplot. Der Median aller Items liegt über dem Cutoff-Wert 4, d. h. die ExpertInnen sehen alle im PraQ-B realisierten Items als wichtig für die Skriptqualität an.

5.2 Konstruktvalidität (t-Test)

Die Gesamtbeurteilungen des Skriptes wurden je TeilnehmerIn ermittelt. Dies geschah durch die Berechnung eines nach Faktorladungen [6] gewichteten Mittelwerts der einzelnen Antworten (jeweils für Skala 1 und 2 getrennt) und führte zu einem Gesamt-

score je Skala. Die so gewonnen Gesamtbeurteilungen der beiden Probandengruppen wurden anschließend durch Zweistichproben t-Tests für unabhängige Stichproben verglichen. Die Erfüllung der notwendigen Vorbedingungen wurde für die Normalverteilungsannahme durch den Shapiro-Wilk-Test sowie für die Varianzhomogenität durch den Levene-Test geprüft und bestätigt.

Für Skala 1 (»Skriptqualität für die Versuchsvorbereitung«, Item 1-3) ergab der t-Test keine signifikanten Unterschiede zwischen Studierenden ($M = 4.1$; $SE = 0.2$) und Didaktik-ExpertInnen ($M = 3.7$; $SE = 0.2$). ($t(70) = -1.36$; $p = .18 > .05$).

Auch für Skala 2 (»allgemeine Skriptqualität«, Item 4-16) ergab der t-Test, dass das Skript im Durchschnitt von den Studierenden ($M = 4.2$; $SE = 0.2$) besser beurteilt wurde als von den ExpertInnen ($M = 3.7$; $SE = .2$). Dieser Unterschied ist signifikant: $t(70) = -2.33$; $p = .03 < .05$. Es handelt sich mit $d = 0.57$ um einen mittleren Effekt.

5.3 Konstruktvalidität (explorative Auswertung)

Um auf Hypothese 2.2 einzugehen, wurde weiterhin eine explorative Auswertung vorgenommen. Dafür wurden die Mittelwerte der Antworten bezüglich der einzelnen Items zwischen den beiden Probandengruppen visuell verglichen (siehe Abbildung 2). Betrachtet man das Antwortmuster der beiden Gruppen, so fällt auf:

- Die mittlere Einschätzung durch die Studierenden lag bei allen Items höher oder auf dem gleichen Niveau wie die durch die ExpertInnen. Dies bestätigt die Befunde aus dem Gruppenvergleich der Skalenmittelwerte auch auf Einzelitemebene.
- Die Antwortmuster beider Probandengruppen ähnelten sich in weiten Teilen.

- Große Abweichungen im Muster ergeben sich nur für Items 7 und 8 (zweite Skala). Zudem ist die Expertenbeurteilung bei Items 11 und 16 schlechter ausgeprägt als die der Studenten.

Eine inhaltliche Analyse dieser Items zeigte allerdings, dass bei den Antworten durch die ExpertInnen für diese vier Items angezweifelt werden kann, ob hier in einer validen Weise die angedachten Aspekte gemessen wurden.

Die Formulierungen von Item 7 und 16 beziehen sich sehr stark auf den tatsächlichen Ablauf des Praktikums, der von den ExpertInnen durch die Art der Befragung nicht beurteilt werden konnte (»Ich habe die Theorie aus dem Skript auch bei der Versuchsdurchführung genutzt.«, »Das Skript macht den Stellenwert der Sachverhalte im Praktikum für das weitere Studium klar.«).

Item 8 (»Das Skript (ggf. Zusatzmaterial) unterstützte mich dabei, ordnungsgemäße Protokolle/ Versuchsberichte anzufertigen.«) bezieht sich auf einen Abschnitt des Skriptes, der aufgrund der Zwangsweisen Eingrenzung des Skriptauszuges in der Expertenbefragung nicht vorgelegt werden konnte.

Bei Item 11 (»Die Lernziele des jeweiligen Versuches werden im Skript verdeutlicht.«) schließlich ist zu vermuten, dass die Verwendung von Begriffen aus der Fachsprache der Didaktik zu verschiedenen Assoziationen in beiden Probandengruppen führte, wodurch in den beiden Gruppen letztendlich nicht das gleiche Merkmal erfasst wurde.

6. Fazit

Für beide Skalen konnte eine erfolgreiche Inhaltsvalidierung vorgenommen werden. Dabei zeigte sich

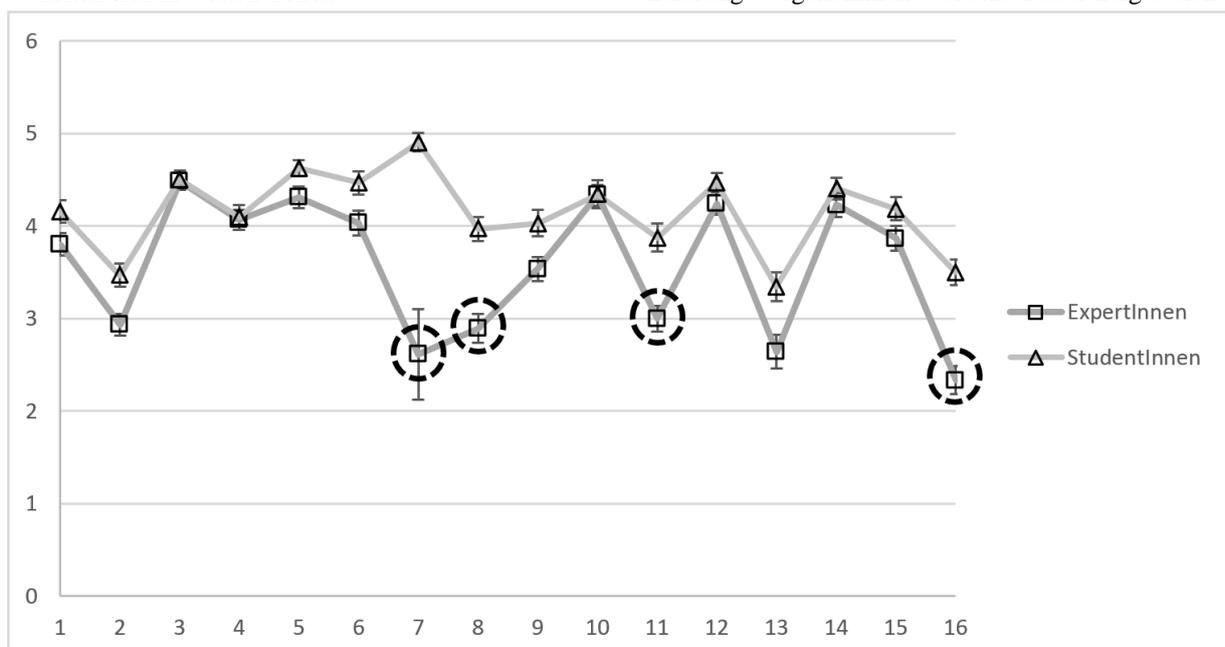


Abb. 2: Mittelwerte der Beurteilungen mit Standardfehler. (Anm.: Die Trendlinien dienen lediglich der Visualisierung der Antwortmuster beider Gruppen und stellen keinen mathematischen Sachverhalt dar. Die umkreisten Datenpunkte sind in ihrer Gültigkeit anzuzweifeln.)

sogar entgegen der Erwartung, dass auch Frage-Item 16 von den Experten als wichtiger Aspekt zur Erfassung von Skriptqualität angesehen wurde. Zudem gab es weder in den theoretischen Vorarbeiten noch in der Studie Hinweise auf unbeachtete Aspekte.

In Bezug auf die Konstruktvalidität ergab die Auswertung für die erste Skala keine Auffälligkeiten: Der t-Test lieferte für die Skriptbeurteilungen keinen signifikanten Unterschied.

Die andererseits für die zweite Skala festgestellte Abweichung muss auf ihre Ursachen hin untersucht werden. Die explorative Auswertung der Antwortmuster gibt einen ersten Eindruck von möglichen Ursachen. Besonders interessant scheinen die vier diskutierten Datenpunkte. Im Falle der Items 7, 8 und 16 ist ein systemischer Fehler in der hier vorgenommenen Befragung zu vermuten, der sich aber in zukünftigen Erhebungen leicht prüfen und vermeiden ließe und nicht die beiden untersuchten Skalen betreffen würde. Für diese wäre es wichtiger herauszufinden, ob (wie vermutet) die Formulierung des Items 11 zu verschiedenen Assoziationen seitens der unterschiedlichen Probandengruppen führt.

7. Literatur

- [1] H. R. Friedrich (2005), „Der Bologna-Prozess nach Bergen,“ *Die Hochschule*, Bd. 2, S. 114-135.
- [2] N. Csonka (2014), „Evaluation von Lehrveranstaltungen an der Humboldt-Universität zu Berlin,“ *Schriftenreihe zum Qualitätsmanagement an Hochschulen*, Bd. 8.
- [3] E. Braun, B. Gusy, B. Leidner und B. Hannover (2008), „BEvaKomp - Berliner Evaluationsinstrument für selbsteingeschätzte studentische Kompetenzen,“ *Diagnostica*, Bd. 54, Nr. 1, S. 30-42.
- [4] F. Thiel, I. Blüthmann und R. Watermann (2012), „Konstruktion eines Fragebogens zur Erfassung der Lehrkompetenz (LeKo),“ in *eues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten. Teil I Evaluation: Veranstaltungsevaluation.*, Berlin, Raabe.
- [5] D. Psillos und H. Niedderer (2002), „Issues and Questions Regarding the Effectiveness of Labwork,“ in *Teaching and Learning in the Science Laboratory*, Springer, S. 21-30.
- [6] D. Rehfeldt, T. Mühlenbruch und V. Nordmeier (2015), „Fragebogen zur Erfassung von Praktikumsqualität (PraQ): Faktorielle Validierung,“ in *Authentizität und Lernen - das Fach in der Fachdidaktik: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung 2015*, Kiel, IPN.
- [7] D. Rehfeldt, T. Mühlenbruch und V. Nordmeier (under review), „Praktikumsqualität messen - Validierungsstudie,“ in *Berlin-Brandenburger Beiträge zur Bildungsforschung 2016*, Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien, Peter Lang.
- [8] D. Rehfeldt und V. Nordmeier (under review), *Theoretically and empirically based Evaluation of laboratory courses – PraQ questionnaire*, proceedings of GIREP 2015 conference.
- [9] H. Borawski und H. Heinke (2005), „Entwicklung eines Physikpraktikums für Studierende der Biologie,“ in *Didaktik der Physik - Beiträge zur DPG Frühjahrstagung*, Berlin.
- [10] G. Merzyn (1994), *Physikschulbücher, Physiklehrer und Physikunterricht*, Kiel: Institut für die Pädagogik der Naturwissenschaften.
- [11] D. Sujew (1986), *Das Schullehrbuch*, Berlin: Volk und Wissen Volkseigener Verlag.
- [12] E. Kircher, R. Girwidz und P. Häußler (2015), *Physikdidaktik*, Berlin: Springer Spektrum.
- [13] K. Haller (1999), *Über den Zusammenhang von Handlungen und Zielen. Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*, Berlin: Logos Verlag.
- [14] H. Fischler (1979), „Das Schulbuch im Physikunterricht der Sekundarstufe I. Ungelöste Probleme mit dem Physikbuch,“ *Lehrmittel Aktuell: Informationen für die Unterrichtspraxis*, S. 28-41.
- [15] E. Vanecek (1995), „Zur Frage der Verständlichkeit und Lernbarkeit von Schulbüchern,“ in *Schulbuchforschung*, Frankfurt am Main, Peter Lang - Europäischer Verlag der Wissenschaften, S. 195-215.
- [16] M. Prenzel (1997), „Sechs Möglichkeiten, Lernende zu demotivieren,“ in *Wege zum Können. Determinanten des Kompetenzerwerbs*, Bern, Verlag Hans Huber, S. 32-44.
- [17] S. Messick (1995), „Validity of psychological assessment: Validation of inferences from persons' responses and performances as scientific inquiry into score meaning,“ *American Psychologist Volume 50 Issue 9*, S. 741-749.
- [18] V. Brüggemann (2015, unveröffentlicht), *Validierung einer Skala zur Erfassung der Skriptqualität in Physik-Experimentierpraktika*, FU Berlin.
- [21] U. Diemer, B. Baser und H.-J. Jodl (1999), *Computer im Praktikum: Moderne physikalische Versuche*, Berlin: Springer Verlag.