

Skalen zur Messung von Praktikumsqualität: Konfirmatorische Analyse der Struktur und Konstrukte

Daniel Rehfeldt, Volkhart Nordmeier

Freie Universität Berlin, Fachbereich Physik, Arnimallee 14, 14195-Berlin
daniel.rehfeldt@fu-berlin.de | volkhart.nordmeier@fu-berlin.de

Kurzfassung

Das Lehrevaluationsinstrument *PraQ* mit Ratingskalen zur Selbst- und Fremdeinschätzung von Aspekten der Praktikumsqualität liefert in einem theoretischen und kompetenzorientierten Rahmen Maße für die Qualität von Praktika. Zielgruppe sind alle naturwissenschaftlichen Anfängerpraktika bis zum vierten Fachsemester an Hochschulen, zunächst begrenzt auf den deutschsprachigen Raum.

Das *3L-Modell* der Praktikumsqualität definiert hierbei den relevanten Lernzuwachs, die Lehrkompetenz der Betreuenden und die Qualität der Lernumgebung von Praktika.

Die Testevaluationsstudie umfasst Studien zur inhaltlichen Adäquatheit der Konstrukte und Items, strukturelle Untersuchungen der Zuordnung von Items und der Konstruktrennbarkeit, externe Vergleiche mit etablierten Messinstrumenten sowie Multiperspektivitätsanalysen.

Gemäß Expert*innenurteilen darf der *PraQ* inhaltlich als Messinstrument aller relevanten Aspekte der Praktikumsqualität verstanden werden. Strukturell zeigen sich in explorativer und konfirmatorischer Faktorenanalyse gut den Items zuzuordnende, reliable Faktoren, die eine getrennte Interpretation der Einzelkonstrukte zulassen. Vergleiche mit Alternativmodellen zeigen die sehr gute Passung des Grundmodells zu den erhobenen Daten. Dies bedeutet, dass Praktikumsleitende bei der Interpretation ihrer Evaluationsergebnisse die Konstrukte getrennt interpretieren dürfen und so gesicherter mögliche Interventionen im Praktikum an eben jenen Konstrukten ansetzen.

Bezogen auf die fachdidaktische Forschung lässt sich schlussfolgern, dass die *PraQ*-Skalen inhaltlich umfassend und strukturell adäquat ihre betreffenden Konstrukte messen, was sie bereits als Messinstrument auch für andere Kontexte, in denen die enthaltenen Konstrukte Relevanz besitzen, nutzbar macht.

1. Einleitung: Praktikumsqualität messen

Seit der Bologna-Reform gehört die Überprüfung und Weiterentwicklung im Bereich der Lehre an den Hochschulen zum allgemeinen Konsens (z. B. [1], [2]). Zentrales Prüfinstrument, um die weitreichenden Struktur- und Studienprogrammanpassungen nach Bologna [3] zu begutachten ist neben der Studiengangakkreditierung die interne Evaluation. Allerdings mangelt es an einer theoretischen Fundierung und Passung der meisten Evaluationsinstrumente und in Folge dessen auch an der Akzeptanz derselben [4]. Aus hochschulstruktureller Perspektive sollte auf Evaluationsinstrumente zurückgegriffen werden, die im universitären Regelbetrieb auch ökonomisch einsetzbar sind. Aus wissenschaftlicher Perspektive sollten diese zudem gültige Interpretationen in Bezug auf die Lehrrealität zulassen (»Validität«). Für Vorlesungen und Seminare wurden bereits solche Instrumente zur fächerübergreifenden Lehrevaluation konstruiert und validiert (z. B. [5], [6]).

Für naturwissenschaftliche Praktika fehlt ein solches, interpretationsevaluiertes Instrument. Eine Übertragbarkeit obiger Instrumente ist auf Grund der speziellen Lehrform »Praktikum« zudem nicht gegeben [3, S. 91]. Praktika bilden aber den Kern der experimentellen Ausbildung in den Naturwissen-

schaften (z. B. [7]) und werden oftmals kritisch in ihrer Lernwirksamkeit betrachtet (z. B. [8, S. 10]). Gerade dort wäre also eine passgenaue und aussagekräftige Lehrevaluation von besonderem Nutzen, auch um Lehrverbesserungen zu initiieren.

Unser Ziel war es daher, einen theoretischen Rahmen und ein Instrument für die Erfassung von Qualität von Praktika zu entwickeln und eine Testevaluation gemäß gewünschter Interpretationen vorzunehmen (»Validierung«):

Ziel 1: Es soll eine fundierte empirische Erfassung aller für die Lehre relevanten Aspekte der Qualität von Praktika realisiert werden, die zeit- und materialökonomisch verläuft. Darüber soll eine Etablierung des Instruments als *reguläres Evaluationsinstrument von Praktika in Hochschulen* möglich werden.

Ziel 2: Es sollen verschiedene Interpretationen der Testergebnisse empirisch abgesichert werden. Demnach soll es möglich sein, mit der Interpretation der Evaluationsergebnisse *fachdidaktisch Forschung betreiben* zu können.

Der Einsatz des Instruments ist dabei für naturwissenschaftliche Anfängerpraktika bis zum 4. Semester an Hochschulen vorgesehen.

Mit der literaturgestützten Erarbeitung eines theoretischen Modells der Praktikumsqualität fand für die Erfassung von praktikumsrelevanten Facetten der erste Schritt statt (Abb. 1). Das *3L-Modell* definiert als Qualitätsdimensionen den *Lernzuwachs* als Kompetenzerwerb der Studierenden, die didaktische und pädagogische *Lehrkompetenz* der Betreuenden und die *Lernumgebung* von Praktika [9].



Abb. 1: Theoretisches 3L-Modell der Praktikumsqualität in drei Hauptdimensionen: Kompetenzzuwachs (Lernerfolg), Lehrkompetenz und Lernumgebung. Letztere zwei wirken hypothetisch auf ersteres ein.

Das 3L-Modell ist so allgemein »naturwissenschaftlich« gehalten, dass die Qualität von Anfängerpraktika der Fächer Physik, Chemie, Biologie und Veterinärmedizin modelliert werden soll. Die Gemeinsamkeiten der Praktika dieser Disziplinen und damit eine Generalisierung des Modells lassen sich sowohl (a) auf der organisatorischen Ebene als auch (b) auf der inhaltlichen Ebene begründen.

- (a) Organisatorisch weisen die meisten dieser Praktika eine ähnliche personelle und materielle Struktur auf. Die Vorbereitung auf den wöchentlichen Praktikumstermin geschieht oft in Heimarbeit mit einem Praktikumsskript. Die Lehre in den Praktika wird von mehreren Personen als Praktikumsbetreuer*innen durchgeführt, die eine vergleichsweise kleine Studierenden-Gruppe betreuen (Betreuungsschlüssel oft 10:1; [8, S. 135]). Experimentiert wird hierbei häufig in Tandems. Die Arbeit an den Experimenten folgt in der Regel einem wöchentlichen Turnus, den irgendwann alle Studierenden durchlaufen haben. Ausgewertet wird schließlich meist wieder in Heimarbeit, in der ein Versuchsbericht angefertigt wird. Die strukturellen Gemeinsamkeiten konnten auch in Studien [10] bestätigt werden, als mit einer neuen Methode die zeitliche Verlaufsstruktur zwischen Praktika der Physik, Chemie und Veterinärmedizin verglichen wurden.
- (b) Inhaltliche Überschneidungen existieren auf Basis der dargestellten Ziele von naturwissenschaftlichen Praktika. Die Fachgebiete überlappen zudem thematisch [11, S. 1] und metho-

disch gibt es einen internationalen Konsens über eine »prototypische Basiskultur experimentellen Arbeitens« [12, S. 34].

Je nach Praktikum und Curriculum können natürlich unterschiedliche Schwerpunkte der Ziele von Praktika gelegt werden.

Aufgrund der Vorteile in Bezug auf die Durchführungs- und Auswertungsökonomie wurde als Messinstrument ein Rating-Fragebogen mit Selbsteinschätzungsitems gewählt (*PraQ*: **Praktikumsqualität**). Der Fragebogen *PraQ-A* umfasst in einer 15-minütigen Erhebung den selbsteingeschätzten Kompetenzzuwachs der Studierenden (vgl. 2.1), der *PraQ-B* soll ebenso in 15min Auskunft über die Lehrqualität des Betreuungspersonals und die im Praktikum eingesetzten Medien geben (vgl. 2.2).

In einem breit angelegten Testevaluationsdesign wurde nun geprüft, ob die Messung mittels PraQ gültige Interpretationen in Bezug auf die enthaltenen Inhalte, die Struktur, die Trennbarkeit der Konstrukte und die Vorhersage anderer Testleistungen liefern kann (vgl. 3).

2. Konstrukte des PraQ-Fragebogens

2.1 PraQ-A: Selbsteingeschätzter Kompetenzzuwachs

Subdimension 1.1: Fachwissen

Der Bereich Fachwissen beschreibt die Fähigkeit, im Praktikum behandelte Inhalte, experimentelle Besonderheiten und Modelle sowohl deklarativ als auch vernetzend erläutern zu können. Gerade durch den erwarteten Theorie-Praxis-Transfer, ein Grobziel nach Haller [13], können hierbei andere Vernetzungen vorgenommen werden, als etwa in Vorlesungen, was daher als Qualitätskriterium verstanden werden kann. Anders als alle anderen Konstrukte der Ergebnisdaten wird beim Fachwissen keine Veränderung modelliert, sondern ein Zustand zu Ende des Praktikums. Dies ist dem Umstand geschuldet, dass Studierende mit umfangreichen Vorerfahrungen durchaus bereits zu Praktikumsbeginn alle inhaltlichen Aspekte beherrschen können. Würde nun eine »Verbesserung« des Fachwissens als Indikator für Praktikumsqualität verwendet, so wäre ein Fehlschluss möglich.

Für die inhaltliche Operationalisierung des Konstrukts wurde sich an Lagowskis »domain knowledge« [11], Staufenschiehls »FEFRA« [14] und für die Theorie-Praxis-Verknüpfung an Arbeiten von [15, S. 32] und dem Effektivitätsmodell von Psillos & Niedderer [7, S. 23] orientiert.

Subdimensionen 1.2: Experimentelle Kompetenz

Der Bereich Experimentelle Kompetenz beschreibt verschiedene Teilkompetenzen, die im Umgang mit dem Experiment und seiner Vor- und Nachbereitung verbessert werden können. Für viele ist dies sicherlich das Kerngeschäft der Praktika, z. B. in Hinblick auf die Studienordnungen oder Hallers [13] Grobziel (2), der Erwerb experimenteller Fähigkeiten. Der

Erwerb experimenteller Kompetenz ist daher ein Qualitätskriterium von Praktika. Strukturell wurde sich hierbei am Modell experimenteller Kompetenz nach Schreiber et al. [16] orientiert, welches zunächst eine sehr kleinschrittige Modellierung des Experimentierprozesses erlaubt. Für Praktika erscheinen insbesondere die Schritte *Fragestellung klären, Erwartungen formulieren, Versuchsplan entwerfen, Geräte zusammenstellen, Versuchsanordnung aufbauen, Messungen durchführen, Messungen dokumentieren, mit Problemen und Fehlern umgehen, Rohdaten graphisch & tabellarisch darstellen, mit Daten Berechnungen anstellen und Ergebnisse interpretieren* relevant. Es ist aber zu betonen, dass es keine strikte, sprunglose Reihenfolge der Schritte gibt. Diese Schritte bilden hypothetisch trennbare Konstrukte der Subdimension Experimentelle Kompetenz. Das eigenständige Bilden von Hypothesen spielt für Grundpraktika keine bzw. eine untergeordnete Rolle und wird meist erst in den Fortgeschrittenen-Praktika forciert, weshalb dieses Konstrukt nicht Bestandteil des 3L-Modells ist.

Für die konkrete Operationalisierung der Konstrukte wurde sich abermals an Arbeiten von Schreiber [17] orientiert. Maiseyenko [18, p. 4] lieferte dazu z. B. den Indikator *Zusammenhänge/Muster innerhalb gewonnener Daten ("Zahlen") erkennen* für die Skala *Rohdaten graphisch, tabellarisch darstellen*. Chang et al. [19, S. 1229] gaben Anregung für die Operationalisierung des Konstrukts *Versuchplanung & Geräte zusammenstellen*, z. B. mit dem Indikator *Experimente nach eigenem Versuchsplan durchführen*.

Auf Aspekte von *nature of science* wurde bewusst verzichtet, da diese eine Metaebene höher liegen und mutmaßlich implizit in den Konstrukten der übergreifenden Kompetenzen (Methoden, Kommunikation, Kooperation) enthalten sind.

Subdimension 1.3: Kommunikationskompetenz

Als Kommunikationskompetenz im Praktikum wird einerseits die schriftliche, andererseits die mündliche Komponente wissenschaftlicher Kommunikation verstanden. Es geht hierbei um die Verbesserung der Fähigkeiten, wissenschaftliche Texte wie Protokolle zu verfassen, sowie um die adressatengerechte Kommunikation von Messergebnissen und Interpretationen. Gerade das Schreiben von Protokollen in den ersten Semestern bildet eine Art erste Übung für das Verfassen wissenschaftlicher Publikationen und nimmt viel Zeit im Praktikum in Anspruch. Auch hier lässt sich ein Bezug zu den übergeordneten Zielen von Praktika nach Haller [13] finden, nämlich in der Dimension *Weiterentwicklung der sozialen Kompetenz, Persönlichkeit und Motivation*, hier heißt es z. B. »allgemeine Fähigkeiten der Kommunikation und der Interaktion weiterzuentwickeln« [ebd., S. 213]. Deshalb lässt sich die Kommunikationskompetenz als Qualitätskriterium von Praktika definieren.

Für die Operationalisierung der Konstrukte wurde sich am Kompetenzbereich *Kommunikation* des Kompetenzmodells für Praktika von Kreiten [20] orientiert, etwa beim Indikator *Fragen fachsprachlich zu diskutieren* (vgl. auch [21, S. 8–10]: *Argumentation*). Lagowski [11, S. 4, 9] und Braun et al. [6] bildeten die Orientierung für Indikatoren wie *Wortbeiträge verständlicher formulieren* oder *Protokolle verständlicher schreiben*. Die AAPT [22, S. 484] lieferte den Indikator *in Protokollen präziser formulieren*.

Subdimension 1.4: Bewertungskompetenz

Bei der Bewertungskompetenz wird, im Unterschied zum schulischen Verständnis, in Praktika lediglich auf das innerfachliche Bewerten Bezug genommen (vgl. [20, S. 67]). So umfasst dieser Kompetenzbereich die Verbesserung in kritischer Interpretation von Daten und die Einschätzung der Grenzen der eigenen Ergebnisse. Erneut gab hier Kreiten [20] mit ihrem Kompetenzmodell für Praktika die hauptsächliche Strukturierung. Vergleichend mit Haller [13] ergeben sich einige Überschneidungen mit ihrem Grobziel (3), *Methoden wissenschaftlichen Denkens kennenlernen*, dort heißt es z. B. »eine kritische Herangehensweise bei der Interpretation von Daten zu entwickeln« [ebd., S. 212]. Der Zuwachs an Bewertungskompetenz als Teil naturwissenschaftlicher Handlungskompetenz kann daher als Qualitätskriterium für Praktika modelliert werden.

Für die konkrete Operationalisierung der Konstrukte wurde sich an Arbeiten von Kreiten [20], Lagowskis *higher order cognition* [11, S. 4] und der KMK [23] im Vergleich mit Maiseyenko [18, S. 16] orientiert.

Subdimension 1.5: Kooperationskompetenz

Unter Kooperationskompetenz wird das Vermögen definiert, mit Mitstudierenden derselben Praktikumsgruppe kooperativ agieren zu können. Da Praktika als eine der wenigen universitären Lehrveranstaltungsformen hoch kooperativ angelegt sind (viel Tandearbeit, geringe Gruppengrößen, vgl. [3, S. 31]), wird dieser Bereich vornehmlich von Praktika tangiert und bildet daher ein Qualitätskriterium. Grundlage für die Operationalisierung bildete eine Adaption der Skala *Kooperation* von Braun et al. [6].

Subdimension 1.6: Personalkompetenz

Als Personalkompetenz beschreibt Erpenbeck [24] vor allem intrinsisch-motivationale Aspekte als produktive Einstellung des Individuums gegenüber Lernen und Selbstentwicklung. Dies lässt sich mit Bestandteilen von Hallers *Weiterentwicklung der sozialen Kompetenz, Persönlichkeit und Motivation* assoziieren (»Interesse zu entwickeln«; »zu motivieren«, [13, S. 213]).

Speziell im Bereich der Praktika wurde diese Sichtweise noch erweitert um den Bereich des Zeitmanagements [25]. Die Ausbildung intrinsischer Motivation durch das Praktikum liefert hierbei wichtige Hinweise auf ein lernförderliches Klima, die Ver-

besserung des eigenen Zeitmanagements ist ein essentieller Faktor im Praktikum, da der enorme zeitliche Aufwand der Vor- und Nachbereitung [26] je Woche ein gutes Zeitmanagement fordert. Daher sind beide als Bestandteile der Qualität von Praktika anzusehen.

Die konkrete Operationalisierung der beiden Konstrukte erfolgte für die intrinsische Motivation erneut über eine Adaption des Instruments nach Braun et al. [6], Skala *Personalkompetenz*. Für das *Zeitmanagement* wurde die Skala nach Wild & Schiefele [25] für Praktika adaptiert.

2.2 PraQ-B: Evaluation der Lehrkompetenz und des Begleitmaterials

Subdimension 2.1: Verständlich erklären (adaptiert nach Thiel et al. [5])

Der Bereich *verständlich erklären* umfasst die Fähigkeit des/r Betreuenden, Begriffe und Sachverhalte nachvollziehbar veranschaulichen und erklären zu können, auch bei Verständnisproblemen oder Nachfragen. Schlüsselzeitpunkte sind hierbei das Vorgespräch und etwaige Nachbesprechungen im Praktikum. Die Relevanz liegt darin, dass den Studierenden so das Verständnis der Theorie erleichtert wird, was einen Einfluss auf die Subdimension 1.1 nahelegt und somit als Qualitätsbeitrag angesehen werden kann.

Subdimension 2.2: Zusammenfassen & Wiederholen (adaptiert nach [ebd.])

Der Bereich *zusammenfassen und wiederholen* betrifft die Tätigkeiten des/r Betreuenden, die Aspekte der Theorie oder des Experiments zusammenzufassen und zu pointieren, die besonders zentral sind. Dies dient den Studierenden als Organisierungshilfe beim Lernen.

Subdimension 2.3: Verknüpfungen herstellen (adaptiert nach [ebd.])

Der Bereich *Verknüpfungen herstellen* umfasst das Herstellen von Bezügen aktueller Anforderungen und Themen zu vergangenen Themen. Dies kann einerseits das Vorwissen der Studierenden sein, andererseits bereits im Praktikum behandelte Sachverhalte. Dies erleichtert vor dem Hintergrund einer konstruktivistischen Perspektive den Lernprozess.

Subdimension 2.4: Verständnis überprüfen (adaptiert nach [ebd.])

Beim Konstrukt *Verständnis überprüfen* geht es darum, dass der/die Betreuende den Studierenden durch gezielte Fragen ermöglicht, das eigene Verständnis des Sachverhalts zu reflektieren und einen etwaigen Erklärungsbedarf auszumachen. Dies unterstützt die Studierenden in der Selbstdiagnose des Lernprozesses und kann daher vertieftes Lernen begünstigen. Gerade im Vorgespräch kann dies zu einem besseren Verständnis der Theorie und des experimentellen Aufbaus im Praktikum beitragen.

Subdimension 2.5: Gute Lernatmosphäre herstellen (adaptiert nach [ebd.])

Dieses Konstrukt umfasst die Fähigkeit des/r Betreuenden, ein lernförderliches Klima im Praktikum herzustellen, was Lernprozesse begünstigt. Hierbei wird die zwischenmenschliche Ebene zwischen Betreuer*in und Studierenden erfasst, was von Belang für die Lernqualität ist.

Subdimension 2.6: Zur Selbsttätigkeit anregen (adaptiert nach [ebd.])

Dieser Bereich zielt auf eine Aktivierung der Studierenden durch den/die Betreuenden ab, welches zu einer persönlichen Auseinandersetzung mit dem Lerngegenstand führt. So kann etwa verhindert werden, dass ein Tandem von Studierenden nicht kooperativ, sondern einzeln abwechselnd und pausierend arbeitet, was die Lernzeit maximiert.

Subdimension 2.7: Relevanz verdeutlichen (adaptiert nach [ebd.])

Bei oftmals stark innerfachlich orientierten Praktika kann es vorkommen, dass Studierende den Sinn der Versuche in Frage stellen. Gute Betreuer*innen betonen daher die Relevanz der Inhalte und Methoden für Anwendungszusammenhänge bzw. für das weitere Studium der Teilnehmer*innen.

Subdimension 2.8: Ermutigen und Selbstwirksamkeit stärken (adaptiert nach [ebd.])

Gerade bei den ersten Erfahrungen mit einer neuen Experimentierkultur können bei den Studierenden Probleme durch das Nicht-Bewältigen von experimentellen Aufgaben bzw. die Vermutung desselben entstehen. Ein/e gute/r Betreuer*in stärkt daher die Selbstwirksamkeitserwartung durch den Ausdruck von Vertrauen und Ermutigungen.

Subdimension 2.9: Effizient mit Störungen umgehen (adaptiert nach [ebd.])

Auch in der universitären Lehre und gerade bei ungünstigen räumlichen Bedingungen können Störungen und Unruhe in der Praktikumsgruppe entstehen, welche es von dem/r Betreuenden zu bewältigen gilt.

Subdimension 2.10: Diagnostik vornehmen

Die Praktikumsbetreuung verlangt eine adäquate Diagnostik des Lernfortschritts [27], bedingt durch die enge Zusammenarbeit mit den Studierenden innerhalb vergleichsweise kleiner Experimentiergruppen und der Relevanz des Umgangs mit (teuren) Praktikumsgeräten. So sollte ein/e Praktikumsbetreuer*in Probleme im Verständnis und im Experiment frühzeitig erkennen, um dann eine passende Maßnahme ergreifen zu können. Selbst wenn diese gewählte Maßnahme aus Nicht-Helfen besteht, so sollte sie dennoch bewusst geschehen. Im Rahmen des PraQ-B wird hier gemäß Pilotierung (siehe 3.1) zwischen zwei faktorenanalytisch trennbaren Konstrukten unterschieden [9]. Auf der einen Seite betrifft die *situative Diagnostik* das sofortige Erkennen von Lernschwierigkeiten. Auf der anderen Seite wird die *Grundeinstellung zur Diagnostik* derart

beschrieben, dass eine grundsätzlich der Diagnostik offene Haltung eingenommen und sich Zeit dafür reserviert wird.

3. Gesamtdesign der Testevaluation

»Validierung«

Um diese vielfältigen qualitätsrelevanten Konstrukte im Praktikum so messen zu können, dass die Ergebnisinterpretationen zu gültigen Schlüssen führen, wurde sich für ein umfassendes Testevaluationsdesign entschieden.

Gemäß dem argument-based approach der Validität (z. B. [28]–[30]) wird an dieser Stelle festgelegt, welche Schlussfolgerungen aus den Ergebnissen von Erhebungen mit dem PraQ-Instrument gezogen können werden sollen.

Erste Anforderung an die Interpretierbarkeit der Testergebnisse ist, dass die komplette lernrelevante Praktikumsqualität in möglichst allen Facetten und ohne qualitätsirrelevante Aspekte bestimmt werden kann. »Construct irrelevant variance« (Messick, [29]) meint hier genau den Fall, dass konstruktfernde Einflüsse oder gar konstruktfernde Variablen mit gemessen werden, was es auszuschließen gilt. Ebenso zu vermeiden ist die »construct underrepresentation« [ebd.]. Gemeint ist hiermit das »Vergessen« von wesentlichen Aspekten der Qualität von Praktika, etwa das Nicht-Erfassen von experimentellen Fähigkeiten, obwohl es zentrales Ziel von Praktika ist (vgl. [13], [21]–[22]).

Die sich hieraus ergebenden Forschungsfragen für die Validierung sind:

FF1 (Content): Welche Indikatoren des PraQ sind nicht konstruktrelevant für Praktikumsqualität?

FF2 (Content): Welche Aspekte von Praktikumsqualität fehlen im PraQ?

Des Weiteren sollen Ergebnisse der Messungen verschiedener Konstrukte der Praktikumsqualität voneinander trennbar interpretiert werden dürfen. Damit könnte nämlich eine etwaige Verbesserung des Praktikums an just denjenigen Konstrukten ansetzen, die sich als problematisch ergeben haben. Dazu gehört auch eine klare Item-Konstrukt-Zuordnung (AERA et al., [28]):

FF3 (Structural): Welche Konstrukte lassen sich beim PraQ theoretisch wie empirisch voneinander getrennt interpretieren?

Für jedes neue Messinstrument liegt ein Vergleich mit bereits etablierten und testevaluierten Instrumenten nahe (»external aspect«, Messick, [29]). Da der PraQ-Fragebogen die Qualitätsfacetten von Praktika über Selbsteinschätzungen von teilnehmenden Studierenden misst, sollte diese Erhebungsform zudem kritisch hinterfragt werden. Die Erhebung mittels Selbsteinschätzung ist zwar deutlich ökonomischer

und stärker mit den institutionellen Anforderungen vereinbar. Es ist aber wichtig abzusichern, dass die Methode keine Verzerrungen bewirkt, um Fehlinterpretationen zu vermeiden. So ist etwa nicht bekannt, wie stark Kompetenzselbsteinschätzungen mit tatsächlicher Kompetenz zusammenhängen, oder ob sie eher Bestandteile von Selbstwirksamkeitserwartung messen.

Wenn die Messung mittels Selbsteinschätzung als tatsächlicher Kompetenzzuwachs interpretierbar sein soll, so müsste sich demnach empirische Evidenz finden lassen, die Zusammenhänge zwischen dem PraQ und etablierten Kompetenzmessinstrumenten aufzeigt. Analoges gilt, falls eher eine Interpretation gemäß Selbstwirksamkeitserwartung gelten können soll.

Wäre für Kompetenz mit Messung per Leistungstest ein großer positiver Zusammenhang wünschenswert, so sollte dieser keinesfalls für ein »allgemeines Praktikums-klima« zutreffen. Dieses Konstrukt beschreibt die soziale Atmosphäre im Praktikum und stellt einen möglichen Störeinfluss für eine objektivierte Interpretation der Messung von Aspekten der Praktikumsqualität dar. Studierende könnten durch etwa ein schlechtes soziales Klima auch gewillt sein, im PraQ-Fragebogen davon unbeeinträchtigte Bereiche (wie etwa die Skriptqualität) herabgesetzt zu bewerten.

Ähnlich also wie der konvergente Zusammenhang zu einem Leistungstest sollte sich ein divergenter Zusammenhang zum Praktikums-klima ergeben, sofern eine störungsarme Interpretation erwünscht ist.

FF4 (External): Wie hängen die Skalen des PraQ statistisch mit der Messung durch Leistungstests zur Kompetenzmessung zusammen?

FF5 (External): Wie hängen die Skalen des PraQ statistisch mit der Messung von allgemeinem Praktikums(sozial)klima zusammen?

Schließlich soll die Evaluation von Praktika allein aus Sicht der Studierenden erhoben werden, wie es institutionell auch verankert ist. Um eine zu einseitige Interpretation der Messergebnisse zu vermeiden, kann es daher lohnen, auch Urteile von dritter Stelle mit denen der Studierenden zu vergleichen (Multiperspektivität, [31, S. 141]).

FF6 (Multiperspective): In welchen Facetten der Praktikumsqualität unterscheiden sich Einschätzungen von Studierenden gegenüber Einschätzungen von Expert*innen?

3.1 Bisherige Testevaluationsergebnisse

FF1 & 2

Die Forschungsfragen vali1 und 2 wurden im Rahmen von kognitiven Interviews von Praktikums-expert*innen untersucht. Die Stichprobe bestand aus

fünf Praktikumsleiter*innen im Berliner Raum, die Praktika für die Fächer Physik, organische Chemie, physikalische Chemie, Biotechnologie und Veterinärmedizin leiteten. Charakteristische Leitfrage beim Begutachten aller Items der PraQ-Fragebögen war »Welche dieser Aussagen sind für ihr Praktikum relevant, welche nicht? Warum?«. Dabei sollte die Passung der bestehenden Items zur Praktikumsrealität kritisch geprüft werden. Bei übermäßiger Ablehnung eines Items (ab 3 von 5) wurde sich für einen Itemausschluss entschieden, sofern die angeführten Begründungen auf Passungsprobleme zurückzuführen waren. Am Ende dieses Teils konnte FF1 mittels der ausgeschlossenen Items beantwortet werden: Insgesamt wurden 9 Items von ursprünglich 141 ausgeschlossen, davon vier im gesamten Bereich *Experimenteller Kompetenz*¹, zwei zur *Bewertungskompetenz*, eines beim Betreuungskonstrukt *Verknüpfungen herstellen* und zwei bei *lerndienliches Feedback geben*².

Die Folgefrage war »Haben wir etwas vergessen, was Sie gerne noch ansprechen möchten?« und zielt damit direkt auf FF2 ab. Zusätzlich wurden Studierendenanmerkungen während der Pilotierung in die Erweiterung des Itempools einbezogen.

Informelle Interviews mit den Autor*innen des LeKo [5] und des BevaKomp [6] zu Itemformulierungen und kognitive Interviews mit drei Studierenden zur Itemverständlichkeit rundeten diesen Schritt ab.

Inhaltlich darf der PraQ demnach als Messinstrument für Praktikumsqualität verstanden werden.

FF6

Zur sechsten Forschungsfrage wurde bereits ein kleiner Beitrag in Bezug auf die multiperspektivische Interpretation der Skriptqualität geleistet, zweier Konstrukte des PraQ-B. Ergebnis war hier, dass Physikdidaktiker*innen die stichprobenartig ausgewählten Skripte fast identisch bewerteten, wie Studierende im betreffenden Praktikum, also in der Interpretation von genügend Urteilsvermögen der Studierenden ausgegangen werden kann ([32], [33]).

FF3

Im Rahmen einer Pilotierungsstudie wurde der gemäß den Ergebnissen von FF1 & 2 angepasste Fragebogen im deutschsprachigen Raum eingesetzt. Die Stichprobe bestand aus 237 Studierenden für den PraQ-A in vier Praktika der Physik, drei der Chemie und einem der Biologie. Sie bestand ferner aus 241 Studierenden für den PraQ-B, wobei in drei Physik- und einem Physik-Nebenfach-Praktikum erhoben wurde.

Ergebnis der strukturellen Auswertung mittels explorativer Faktorenanalyse war eine erwartungsgerechte Faktorstruktur, einige Items mussten aufgrund

schlechter struktureller Passung³ entfernt werden. Dies waren 12 von 80 Items beim PraQ-A und acht von 46 Items beim PraQ-B; je mit sehr geringem Verlust an inhaltlicher Konstruktbreite verbunden, außer für das Konstrukt *lerndienliches Feedback geben* (Ladungen aller Items zu gering).

Im Rahmen dieses Beitrags wird nun auf die Forschungsfrage FF3 derart genauer eingegangen, dass eine konfirmatorische Analyse der in der Pilotierung erhaltenen Faktorstruktur ansteht.

3.2 Forschungsfrage zur Trennbarkeit der Konstruktinterpretationen

Für FF3 wurde bereits eine explorative Faktorenanalyse (EFA) für diejenigen hypothetischen Skalen durchgeführt, die neu konstruiert oder stark von der Ursprungsquelle abgewandelt wurden [34]. Damit wurden neue Strukturen entdeckt und vermutete untermauert.

Gemäß Rost [35] folgt idealerweise daraufhin eine konfirmatorische Faktorenanalyse (CFA), die das in der EFA etablierte Modell a priori in die Analyse implementiert und die Passung des Modells zur neuen Stichprobe prüft.

Diese kann Aufschluss geben über...

- die Güte der Messmodelle der einzelnen Konstrukte (latente Faktoren).
- die Passung des Modells auf die vorliegenden Daten insgesamt.
- die empirische Trennbarkeit der Faktoren.

Die ursprüngliche Forschungsfrage enthält also statistisch drei Teilkomponenten, die es im Folgenden zu analysieren gilt. Dabei ist a) Voraussetzung für b) [36, S. 129] und c) kann nur dann analysiert werden, wenn das Modell überhaupt in sich stimmig ist, also die theoretische Zuordnung von Items zu Faktoren sich empirisch wiederfinden lässt.

4. Methode

4.1 Erhebung und Probandenauswahl

Die Erhebung mittels PraQ-Fragebögen wurde analog zur Pilotierung⁴ in diversen Hochschulen im deutschsprachigen Raum vorgenommen (Abb. 1). Zielgruppe waren Anfängerpraktika (max. viertes Semester) in den Naturwissenschaften. Es wurde je Praktikum entweder der PraQ-A oder der PraQ-B eingesetzt, sobald mindestens 40% der Lehrveranstaltungszeit im Semester verstrichen war. In der Regel fanden die Erhebungen innerhalb der letzten drei Praktikumstermine statt. Die Erhebung selbst wurde als klassische Lehrevaluation mit Papierfragebögen in den Praktikumsräumen durchgeführt, oft vor den experimentellen Phasen im Praktikum. An nicht Berlin-Brandenburgischen Standorten wurde

¹ Hier wurden v. a. zu stark an der Physik orientierte Items entfernt (Fehlerrechnung, Statistik).

² Dieses Konstrukt musste im Rahmen der Pilotierungsergebnisse komplett entfernt werden.

³ Kommunalität < .60; mind. eine Nebenladung > .40; Keine Ladung > .40; Kein Zusammenhang mit Faktor (Einzelitem)

⁴ Dies waren die Erhebungen, die mittels explorativer Faktorenanalyse ausgewertet wurden.

durch die Praktikumsleitenden oder die Betreuenden erhoben. Zum Zweck der Durchführungsobjektivität wurde ein Manual für die standardisierte Durchführung verwendet. Bezüglich des Datenschutzes wurde das Deckblatt ohne Items gestaltet, um nicht von außen einsehbar zu sein.

Für die Erhebung wurde eine Stichprobengröße von $N > 250$ je Fragebogentyp angestrebt.



Abb. 2: Standorte der Praktika, in denen die Erhebungen für FF3 stattfanden (Bildquelle: Quickslide, Strategy Compass GmbH).

4.2 Auswertungsmethode: Konfirmatorische Faktorenanalyse

Je Fragebogen wurde eine eigene konfirmatorische Faktorenanalyse (CFA) angestrebt. In dieser wird die Item-Faktor-Zuordnung, die so genannte Faktorstruktur definiert. Die Faktorstruktur stammte hierfür wie erläutert aus den explorativen Faktorenanalysen der Pilotierungsdaten. Bezüglich der nur wenig veränderten Skalen nach Thiel et al. [5] wurde auf eine explorative Analyse verzichtet und stattdessen die Ergebnisse der Analysen von [ebd.] für die Festlegung der theoretischen Faktorstruktur verwendet.

Bei der konfirmatorischen Analyse wird im Wesentlichen geprüft, ob die theoretische Kovarianzmatrix der empirischen gleicht. Erstere beschreibt die korrelativen Zusammenhänge zwischen den Items gemäß erfolgter EFA, letztere umfasst die empirisch in der Validierungserhebung gemessenen Zusammenhänge.

Voraussetzungen für eine CFA mit dem ML-Schätzer [37, S. 106] und einem *missing-data-handling* mittels FIML [36, S. 370] sind neben der Erfüllung der Normalverteilungsannahme je Item eine *missing-Rate* von unter 5% je Proband und je

Item. Alle diese Voraussetzungen waren nach der Datenbereinigung erfüllt.

Die Kriterien für eine gelungene Übereinstimmung des theoretischen Modells mit den empirischen Daten fasst Tabelle 1 zusammen.

Kriterium
Korrelationen der Faktoren in $[-.90; +.90]$ (Chang, 2011, S. 1227)
Standardisierte Faktorladungen in $ [.50; .95]$
Faktorvarianz sign. > 0
Standardisierte Residuen größtenteils vom Betrag < 2.58 (Brown, 2006)
Absolute Fit-Kriterien
SRMR < 0.10 (Byrne, 2001)
RMSEA < 0.10 (Brown, 2006, S. 87)
Inkrementelle Fit-Kriterien
CFI $> .90$ (Kline, 2005)
TLI $> .90$ (Brown, 2006, S. 87)
MI: sEPCs für korrelierte Residuen $< .20$ (Saris et al., 2009, S. 573)

Tab. 1: Kriterien an die Analyse innerhalb der CFA.

Besonders hervorzuheben sind hier die *fit-indices*, welche Auskunft über die Passung des Modells zu den Daten geben.⁵ Dies trägt gemeinsam mit dem Benchmarking für die Itemladungen ($> .50$) und den standardisierten *expected parameter changes* (sEPCs $< .20$; [38, p. 573] zur Beantwortung der statistischen Forschungsfragen a) (Einzelmessmodelle) und b) (Gesamtmodell) bei. Die Trennbarkeit der Faktoren wiederum betreffen vor allem die Faktorkorrelationen, die für klar trennbare Konstrukte unter $.90$ liegen sollten [19, S. 1227].

Technisch wurde die Analyse über das *package lavaan* in R realisiert (Version 0.5-18).

4.3 Stichprobe

Bezüglich des PraQ-A konnten für diese konfirmatorische Studie die Fachrichtungen Veterinärmedizin (Berliner Raum), Physik (Tübingen, Köln, Berlin, Potsdam, Kiel, Aachen) und Chemie (AC: Berlin) gewonnen werden. Für den PraQ-B konnten Erhebungen in Praktika der Veterinärmedizin (Berliner Raum), Chemie (AC: Wuppertal, PC: Berlin) und Physik (Berlin, Potsdam, Tübingen, Trier, Köln) realisiert werden. Die Stichprobengröße für den PraQ-A betrug $N_A = 576$ und $N_B = 260$ und lag damit jeweils über den angestrebten 250.

Die Stichprobe des PraQ-A setzte sich zusammen aus 226 Physik-, 153 Chemie-, 40 Biologie-, 26 Medizinstudierenden bei 131 Sonstigen oder nicht angegebenen Studiengängen. Die Fachrichtung der Praktika war hauptsächlich Physik (470 Studieren-

⁵ Für weitere Informationen zu den *fit-indices* sei auf die in der Tabelle angegebene Literatur oder Brown [36] verwiesen.

de), teils Chemie (28 Studierende) und teils Veterinärmedizin (20 Studierende). Es gab hierbei 112 Studierende in Blockpraktika.⁶

Die Stichprobe des PraQ-B setzte sich zusammen aus 71 Physik-, 40 Chemie-, 21 Biologie-, 69 Medizinstudierenden bei 59 Sonstigen oder nicht angegebenen Studiengängen. Die Fachrichtung der Praktika war hauptsächlich Physik (156 Studierende), teils Chemie (21 Studierende) und teils Veterinärmedizin (65 Studierende). Es gab hierbei 58 Studierende in Blockpraktika.

5. Auswertung

5.1 Einzel-Messmodelle der Faktoren

Da PraQ-A und -B gemeinsam 30 Skalen enthalten, wird hier auf die Ergebnisse des PraQ-B überblicksartig und auf die des PraQ-A genauer eingegangen.

Bevor das Gesamtmodell getestet werden kann (b), müssen zunächst die einzelnen Messmodelle je Faktor überprüft werden (a). Eine Zusammenfassung der relevanten fit-indices und Spezifizierungen der Messmodelle zeigt Tabelle 2. Dabei ist zu betonen, dass für alle dargestellten Modelle keine sEPCs < .20 für Item-Residualkorrelationen vorlagen.

Die Auswertung zeigt bezüglich der angelegten Kriterien an die Modellgüte (Tab. 1) viele fit-Parameter im Normbereich und nur wenige leicht über dem Normbereich. Überschreitungen der Obergrenze eines Kriteriums lagen nicht vor.

Es wurden im Rahmen der Analyse, auch auf Basis der *modification-indices*, korrelierte Residuen von Items dann zugelassen, wenn es inhaltlich plausible Gründe für eine über die Faktorvarianz hinausgehende Interkorrelation der Residuen gab, z. B. *analoges wording*.⁷ Items wurden dann entfernt, wenn ihre Ladung .50 unterschritt, korrelierte Residuen mit *allen* übrigen Items in den *modification-indices* vorgeschlagen wurden (überflüssiges Item, [36, S. 203]) oder gravierende Probleme mit der Modellpassung auf ein einzelnes Item zurückzuführen waren (hohe standardisierte Residuen). Zusätzlich gab es einen Fall, in dem zwei Items stark analogen wordings, aber unterschiedlicher Faktoren, sich so stark von der Varianz überlappten, dass die Trennbarkeit der betreffenden Konstrukte nicht mehr gewährleistet war. Die beiden Items luden also stark auf beiden Konstrukten und wurden, auch auf Grund des besseren Modellfits der neuen Lösung (vgl. Abb. 2), ent-

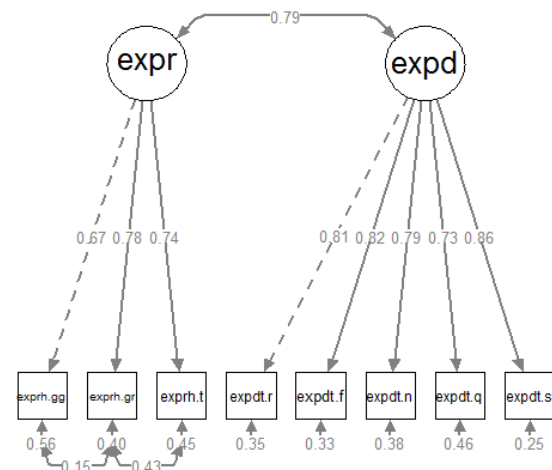


Abb. 3: Zwei-Faktoren-Modell für die Konstrukte EXP: Rohdaten darstellen und EXP: Daten interpretieren unter entfernen zweier Items. Die Korrelation der beiden Faktoren liegt nun unterhalb des kritischen Werts von .85 und lässt damit eine getrennte Interpretation der zugehörigen Konstruktausprägungen zu. Alle Ladungen sind zudem genügend groß (> .50).

fernt.⁸ Bei alledem wurde stets die Gefährdung von inhaltlicher Aussagekraft durch das Entfernen von Items mit bedacht. Eine starke Veränderung der inhaltlichen Breite des Konstrukts ist lediglich zu erwarten für die Skala *EXP: Rohdaten darstellen*. Hier wurden Items zum Messdaten auswerten, Erkennen von Zusammenhängen in Daten und zum Repertoire der Darstellungsformen entfernt. Übrig bleiben die Facetten »graphisch darstellen«, »tabellarisch darstellen« und »geeignete Darstellungen erstellen«.

Moderate Einbußen ergeben sich für das *theoretische* und *praktische Fachwissen*, das *Formulieren von Erwartungen* beim Experimentieren, das *Daten Interpretieren* und die *Kooperationskompetenz* mit je einem Itemausschluss. Für diese Fälle konnten jedoch stets noch vorhandene Items gefunden werden, die die inhaltliche Facette im Wesentlichen abdeckten.

Für den PraQ-B gab es analoge Ergebnisse, die in einer folgenden Publikation genauer beschrieben werden.

⁶ Für Studierende in Block- und Semesterpraktika konnte bereits eine Messinvarianz auf skalarem Niveau nachgewiesen werden. Die ersten Analysen bzgl. der Messinvarianz zwischen Physik- und Chemiestudierenden sehen zudem vielversprechend aus.

⁷ BsS.: Kommunikationskompetenz (mündlich): »...fällt es mir leichter, meine eigenen Eindrücke/Meinungen zu äußern.« und »...fällt es mir leichter, nachzufragen, wenn ich etwas nicht verstanden habe.«

⁸ Es handelte sich bei *EXP: Rohdaten darstellen* und *EXP: Daten interpretieren* um die Items »...erkenne ich Zusammenhänge/Muster innerhalb meiner gewonnenen Daten ("Zahlen") besser.« und »...erkenne ich Zusammenhänge/Muster in den grafischen Darstellungen meiner Daten besser.«

Faktor (# Items nach der Analyse)	alle Faktorladungen > .50 ?	SRMR	RMSEA	CFI	Korrelierte Residuen zugelassen?	# Items ausgeschlossen
Fachwissen: Theorie (3)		--- Identifiziertes Modell ---				1
Fachwissen: Praxis (3)		--- Identifiziertes Modell ---				1
EXP: Erwartungen formulieren (4)	erfüllt	0.01	0.04	1	Ja	1
EXP: Versuchsplanung & Geräte (5)	erfüllt	0.03	0.08	0.98	Nein	0
EXP: Gerätekenntnisse (3)		--- Identifiziertes Modell ---				0
EXP: Versuchsanordnung aufbauen (4)	erfüllt	0.01	0.08	0.99	Nein	0
EXP: Messungen durchführen (4)		--- Identifiziertes Modell ---				0
EXP: Messungen dokumentieren (3)		----- Identifiziertes Modell -----				0
EXP: Probleme lösen (4)	erfüllt	0.01	0.08	0.99	Nein	0
EXP: Rohdaten darstellen (3)		--- Identifiziertes Modell ---				3
EXP: Daten interpretieren (5)	erfüllt	0.02	0.06	0.99	Nein	1
Kommunikationskompetenz: schriftlich (5)	erfüllt	0.01	0.06	1	Ja	0
Kommunikationskompetenz: mündlich (6)	erfüllt	0.02	0.08	0.98	Ja	0
Bewertungskompetenz (5)	erfüllt	0.01	0.05	0.99	Nein	0
Kooperationskompetenz (4)	erfüllt	0.01	0.07	0.99	Nein	1
Personalkompetenz: Intrinsische Motivation (5)	erfüllt	0.02	0.1	0.98	Nein	0
Personalkompetenz: Zeitmanagement (4)	erfüllt	0.01	0.08	0.99	Nein	0

Legende:
im Normbereich |
Leicht über dem Normbereich
Stark über dem Normbereich

Tab. 2: Konfirmatorische Analyse der Einzel-Messmodelle (je Faktor) des PraQ-A. Dargestellt sind populäre fit-indices und einige Informationen zu den Items. Auf Grund von identischen wordings von Items wurden teils korrelierte Residuen zugelassen. Die Modification-Indices führten weiterhin zum Ausschluss von insgesamt 8 Items. Faktoren mit nur drei Items können mathematisch erst im Gesamtmodell geprüft werden (identifiziert).

5.2 Gesamtmodell und Alternativmodelle

Nach der erfolgreichen Etablierung von Einzelmessmodellen für die Konstrukte des PraQ-A konnte nun das Gesamt-Messmodell inklusive aller Items und Konstrukte des PraQ-A getestet werden. Das Gesamtmodell, das sich aus den unveränderten Einzelmessmodellen zusammensetzte, erreichte einen akzeptablen Modellfit mit allen Faktorladungen > .50, einem SRMR von .05, einem RMSEA von .04 und einem CFI von .93. Nach weiteren Modellspezifikationen mittels inhaltlich haltbarer korrelierter Residuen konnte das finale Messmodell noch signifikant optimiert werden (SRMR = .04, RMSEA = .04, CFI = .94, Δ CFI = .01) und enthielt auch keine Faktorkorrelationen von über .85 mehr.

Die Testung möglicher Alternativmodelle umfasste ein Modell, das die *Experimentelle Kompetenz* als einen übergreifenden Faktor annimmt, sowie ein Modell, das die gesamte Praktikumsevaluation als einen Faktor »Praktikumsqualität« annimmt, also sämtliche Items auf nur einem gemeinsamen Faktor laden. Zusätzlich wurden auf Einzelmesebene noch *EXP: Rohdaten darstellen* und *EXP: Daten interpretieren* als ein gemeinsamer Faktor modelliert, da diese die höchste Faktorkorrelation im Gesamtmodell aufwiesen.

Alle Modellvergleiche fielen gemäß χ^2 -Differenzentest und CFI-Differenzkriterium von Δ CFI \geq .01 [39, S. 251] zu Gunsten der Grundmodelle aus, wie Tabelle 3 genauer zeigt.

	Grundmodell	Alternativmodell
	17-Faktoren-Modell: Gesamtmodell***	1-Faktor-Modell: Praktikumsqualität
χ^2	4010.2	13879.7
df	2203	2339
$p <$	0.0001	
CFI	0.94	0.6
ΔCFI	0.34	
	9-Faktor-Modell: Experimentelle Kompetenz***	1-Faktor-Modell: Experimentelle Kompetenz
χ^2	1025.2	4121.7
df	520	556
$p <$	0.0001	
CFI	0.96	0.73
ΔCFI	0.23	
	2-Faktor-Modell: Daten interpretieren vs. Rohdaten darstellen***	1-Faktor-Modell: Umgang mit Daten
χ^2	47.1	85.7
df	17	18
$p <$	0.0001	
CFI	0.98	0.96
ΔCFI	0.02	

Tab. 3:

Modellvergleiche mittels χ^2 -Differenzentest und ΔCFI -Kriterium. Die fettgedruckten Modelle sind in beiden Tests signifikant besser als ihr Pendant.

Auch Modellvergleiche beim PraQ-B fielen stets zu Gunsten des ursprünglichen Messmodells aus.

5.3 Empirische Trennbarkeit der Faktoren

Eine empirische Trennbarkeit der Faktoren ist gemäß Rost [40, S. 213] dann anzunehmen, wenn die Interfaktorkorrelationen kleiner als .85 ausfallen, Chang [19, S. 1227] dagegen sieht dies bereits bei kleiner .90 als erfüllt an. Im Falle des PraQ-A liegen die Interfaktorkorrelationen im Intervall [.25, .79], sind also allesamt positiv und kleiner als .85. Die mittlere Interfaktorkorrelation beträgt .56. Die größte Korrelation besteht erwähnungsgemäß zwischen EXP: Rohdaten darstellen und EXP: Daten interpretieren, die niedrigste zwischen EXP: Probleme lösen und Zeitmanagement.

Auch beim PraQ-B ergaben sich Faktorkorrelationen, die eine Trennbarkeit der Konstrukte annehmen lassen, auch wenn hier teils höhere Korrelationen um .85 vorlagen.

5.4 Reliabilitätsanalyse

Gemäß Chang [19, S. 1227] sind Reliabilitätsmaße $> .60$ annehmbar im Rahmen einer CFA. Tabelle 4 stellt hierzu eine kleine Auswahl der Reliabilitätsmaße in absteigender Größe dar.

Zu erkennen ist, dass die Mehrzahl der Skalen eine hohe bis sehr hohe Reliabilität aufweisen, lediglich die PraQ-B Betreuer*innen-Skala zur *Selbsttätigkeit anregen* und die PraQ-B EXP-Skala *Rohdaten darstellen* liegen in einem niedrigen Intervall [.60, .70], was bei letzterer nicht zuletzt dem Entfernen von dreier Items und dem Zulassen korrelierter Residuen geschuldet ist.

Reliabilität Raykov's ρ	
Skriptqualität	0.93
Kommunikationskompetenz schriftlich	0.93
Personalkompetenz: Zeitmanagement	0.91
Zusammenfassen & Wiederholen	0.91
=====	
EXP: Erwartungen formulieren	0.82
EXP: Messungen durchführen	0.82
Gute Lernatmosphäre herstellen	0.81 [#]
=====	
Integration der Vorlesungsinhalte	0.77
Fachwissen: Praxis	0.75
zur Selbsttätigkeit anregen	0.69
EXP: Rohdaten darstellen	0.66

Legende:

Cronbach's α

Tab. 4: Reliabilität der Skalen (Auswahl).

6. Fazit

Nach der Auswertung kann die Forschungsfrage FF3 beantwortet werden:

FF3 (Structural): Welche Konstrukte lassen sich beim PraQ theoretisch wie empirisch voneinander getrennt interpretieren?

Dazu gehört statistisch:

- a) die Güte der Messmodelle der einzelnen Konstrukte (latente Faktoren)
- b) die Passung des Modells auf die vorliegenden Daten insgesamt
- c) die Empirische Trennbarkeit der Faktoren

Die Messmodelle der Einzelfaktoren (a) ergaben für alle beteiligten Konstrukte in den Bereichen *Fachwissen*, *Experimentelle Kompetenz*, *Kommunikationskompetenz*, *Bewertungskompetenz*, *Kooperationskompetenz* und *Personalkompetenz* für die Selbsteinschätzung der eigenen Kompetenz durch die Studierenden der Praktika einen annehmbaren bis sehr guten Modellfit (Tab. 2), sofern dieser prüfbar war. Für alle restlichen Faktoren und den fit der Messbeziehungen als Ganzes diente die Analyse des Gesamtmodells (b), die einen guten Fit (SRMR = .04, RMSEA = .04, CFI = .94) lieferte, also das

angenommene Messmodell und somit auch das theoretische 3L-Modell der Praktikumsqualität empirisch stützt, zudem sämtliche plausiblen Alternativmodelle signifikant schlechter als das Grundmodell ausfielen.

Dabei ergab die Analyse der Korrelationen der Faktoren im Gesamtmodell, dass diese gering genug waren, dass die Annahme trennbarer Konstruktinterpretationen (c) beibehalten werden darf. Lediglich der Faktor *Rohdaten darstellen* der Experimentellen Kompetenz wurde mit drei Itemausschlüssen inhaltlich arg beschnitten und weist eine geringe Reliabilität auf. Dieses Konstrukt darf also vor dem Hintergrund der verbliebenen Items wirklich nur noch als ein »darstellen« der Rohdaten verstanden werden und nicht mehr als der gesamte Umgang mit Rohdaten (vgl. 5.1).

Antwort FF3: *Es lassen sich alle modellierten Konstrukte (vgl. Abschnitte 2.1 und 2.2) theoretisch wie empirisch getrennt voneinander interpretieren.*

Darüber hinaus ergaben sich die Konstruktschätzungen als sehr messgenau, wie die hohen Reliabilitätskoeffizienten belegen.

Bezogen auf Ziel 1 bedeutet dies, dass Praktikumsleitende bei der Interpretation ihrer Evaluationsergebnisse des PraQ die Konstrukte getrennt interpretieren dürfen. Auch können so gesicherter die Interventionen an eben jenen Konstrukten ansetzen.

Bezogen auf Ziel 2 lässt sich schlussfolgern, dass die PraQ-Skalen inhaltlich umfassend und strukturell adäquat ihre betreffenden Konstrukte messen, was sie bereits als Messinstrument für die fachdidaktische Forschung nutzbar macht.

Einschränkend muss jedoch auf die recht physiklastige Stichprobe verwiesen werden, die eine klare Absicherung von Interpretationen in Chemie- und sonstigen Praktika nicht gewährleisten konnte.

7. Ausblick

Verbleibende Testevaluationsschritte gemäß Validierungsdesign sind nun, Vergleiche mit konkurrierenden Messinstrumenten zu realisieren, die ähnliche Konstrukte erfassen (externaler Aspekt). Ersteres wird über korrelative Vergleiche mit dem Kowadis-Test [41] konvergent und einem Maß der Praktikumszufriedenheit divergent realisiert.

Am Ende einer erfolgreichen Testevaluation stehen die inhaltlichen Forschungsfragen zur Analyse. Diese können Erkenntnisse über die Praktikumsqualität geben:

Was sind Stärken, Schwächen und die qualitative Bandbreite von Praktika in Deutschland in Bezug auf die Praktikumsqualität?

Was sind Unterschiede zwischen Chemie- und Physikpraktika in Bezug auf die Praktikumsqualität?

Was sind Unterschiede zwischen Block- und Semesterpraktika in Bezug auf die Praktikumsqualität?

Durch Erhebungen in 18 Praktika im deutschsprachigen Raum ist eine Generalisierbarkeit der Ergebnisse deutschlandweit naheliegend und wird neue Erkenntnisse über die Praktikumsrealität an deutschsprachigen Hochschulen liefern.

8. Literatur

- [1] H. R. Friedrich, "Der Bologna-Prozess nach Bergen," *Hochsch.*, no. 2, pp. 114–135, 2005.
- [2] A. Hopbach, "Qualifikationsrahmen für deutsche Hochschulabschlüsse," in *Handbuch Qualität und Lehre*, Benz, Kohler, and Landfried, Eds. Berlin, 2007.
- [3] E. Bosbach, Hochschulrektorenkonferenz, and Service-Stelle Bologna, Eds., "Bologna-Reader II Neue Texte und Hilfestellungen zur Umsetzung der Ziele des Bologna-Prozesses an deutschen Hochschulen," in *Neue Texte und Hilfestellungen zur Umsetzung der Ziele des Bologna-Prozesses an deutschen Hochschulen: Beiträge zur Hochschulpolitik*, 1. Aufl., vol. 5, Bonn: Hochschulrektorenkonferenz, 2007.
- [4] N. Csonka, "Evaluation von Lehrveranstaltungen an der Humboldt-Universität zu Berlin," *Schriftenreihe Zum Qual. Hochschulen*, vol. 8, 2014.
- [5] F. Thiel, I. Blüthmann, and R. Watermann, "Konstruktion eines Fragebogens zur Erfassung der Lehrkompetenz (LeKo)," in *Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten. [Teil] I. Evaluation. Veranstaltungsevaluation*, B. Berendt and H. P. Voss, Eds. Berlin: Raabe, 2012.
- [6] E. Braun, B. Gusy, B. Leidner, and B. Hannover, "Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp)," *Diagnostica*, vol. 54, no. 1, pp. 30–42, Jan. 2008.
- [7] D. Psillos and H. Niedderer, "Issues and Questions Regarding the Effectiveness of Labwork," in *Teaching and Learning in the Science Laboratory*, D. Psillos and H. Niedderer, Eds. Netherlands: Springer, 2002, pp. 21–30.
- [8] L. Hucke, *Handlungsregulation und Wissenserwerb in traditionellen und computergestützten Experimenten des physikalischen Praktikums*. Berlin: Logos, 1999.
- [9] D. Rehfeldt and V. Nordmeier, "Theoretically and empirically based Evaluation of laboratory courses – PraQ questionnaire," *Proc. GIREP-MPTL Int. Conf. TeachingLearning Phys. Wroclaw PL*, in print.
- [10] T. Gutzler, D. Rehfeldt, and V. Nordmeier, "TSL: Bedarfsanalyse in Praktika: Ein 'neues' Werkzeug zur Strukturierung," in *Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht: Gesellschaft für Didaktik*

- der Chemie und Physik. Jahrestagung in München 2013, Münster, 2014.
- [11] J. J. Lagowski, "THE ROLE OF THE LABORATORY IN CHEMICAL EDUCATION," presented at the International Conference on Chemical Education, Beijing, 2002.
- [12] M. Emden, *Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens: eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I*. Berlin: Logos, 2011.
- [13] K. Haller, *Über den Zusammenhang von Handlungen und Zielen. Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum*. Berlin: Logos, 1999.
- [14] T. Staufenbiel, "Fragebogen zur Evaluation von universitären Lehrveranstaltungen durch Studierende und Lehrende," *Diagnostica*, vol. 46, no. 4, pp. 169–181, Oct. 2000.
- [15] F. Sander, *Verbindung von Theorie und Experiment im physikalischen Praktikum – Eine empirische Untersuchung zum handlungsbezogenen Vorverständnis und dem Einsatz grafikorientierter Modellbildung im Praktikum*. Logos Berlin, 2000.
- [16] N. Schreiber, H. Theyßen, and H. Schecker, "Diagnostik experimenteller Kompetenz: ein Verfahrensvergleich," in *Konzepte fachdidaktischer Strukturierung für den Unterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011*, Münster, 2012, pp. 263–265.
- [17] N. Schreiber, *Diagnostik experimenteller Kompetenz: Validierung technologiegestützter Testverfahren im Rahmen eines Kompetenzstrukturmodells*, vol. 139. Berlin: Logos, 2012.
- [18] V. Maiseyenko, *Modellbasiertes Experimentieren im Unterricht: Praxistauglichkeit und Lernwirkungen*. Berlin: Logos, 2014.
- [19] H.-P. Chang, C.-C. Chen, G.-J. Guo, Y.-J. Cheng, C.-Y. Lin, and T.-H. Jen, "The development of a competence scale for learning science: Inquiry and communication," *Int. J. Sci. Math. Educ.*, vol. 9, no. 5, pp. 1213–1233, 2011.
- [20] M. Kreiten, "Chancen und Potenziale web-basierter Aufgaben im physikalischen Praktikum," Universität zu Köln, Köln, 2012.
- [21] B. M. Zwickl, N. Finkelstein, and H. J. Lewandowski, "The process of transforming an advanced lab course: Goals, curriculum, and assessments," *Am. J. Phys.*, vol. 81, no. 1, pp. 63–70, Jan. 2013.
- [22] American Association of Physics Teachers, "Goals of the Introductory Physics Laboratory," *Am. J. Phys.*, vol. 66, no. 6, p. 483, 1998.
- [23] KMK, *Bildungsstandards im Fach Physik (Chemie / Biologie) für den Mittleren Schulabschluss*. München: Luchterhand, 2005.
- [24] J. Erpenbeck, "Einführung," in *Handbuch Kompetenzmessung*, J. Erpenbeck and L. von Rosenstiel, Eds. Stuttgart: Schaefer-Poeschel, 2003, pp. IX–XL.
- [25] K.-P. Wild and U. Schiefele, "Lernstrategien im Studium: Ergebnisse zur Faktorenstruktur und Reliabilität eines neuen Fragebogens," *Z. Für Differ. Diagn. Psychol.*, 1994.
- [26] C. Metzger, A. Ladstätter-Weissenmayer, P. Heyken, R. Schulmeister, and M. Radmacher, "ZEITLast-Studie des Studiengangs B.Sc. Physik an der Universität Bremen, Wintersemester 2011/12." Universität Bremen, 2012.
- [27] H. Ditton and D. Merz, "Qualität von Schule und Unterricht," *Z. Für Pädagog.*, vol. 41, no. 4, pp. 507–529, 1995.
- [28] AERA, APA, and NCME, *Standards for educational and psychological testing*. Washington, DC: American Educational Research Association, 2014.
- [29] S. Messick, "Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning," *Am. Psychol.*, vol. 50, no. 9, p. 741, 1995.
- [30] Michael T. Kane, "Current Concerns in Validity Theory," *J. Educ. Meas.*, vol. 38, no. 4, pp. 319–342, 2001.
- [31] M. Gollwitzer and R. S. Jäger, *Evaluation kompakt*, 1st ed. Weinheim: Beltz, 2009.
- [32] V. Brüggemann, "Validierung einer Skala zur Erfassung der Skriptqualität in Physik-Experimentierpraktika, Abschlussarbeit zum Erwerb des akademischen Grades Master of Education," Freie Universität Berlin, 2015.
- [33] D. Rehfeldt, V. Brüggemann, and V. Nordmeier, "Der PraQ: Ein Instrument zur Einschätzung von Praktikumsqualität – Skriptqualität," *PhyDid B - Didakt. Phys. - Beitr. Zur DPG-Frühjahrstagung*, 2015.
- [34] D. Rehfeldt, T. Mühlenbruch, and V. Nordmeier, "Praktikumsqualität messen – Validierungsstudie," in *Berlin-Brandenburger Beiträge zur Bildungsforschung 2016*, J. Stiller and C. Laschke, Eds. Frankfurt am Main, Berlin, Bern, Bruxelles, New York, Oxford, Wien: Peter Lang, eingereicht.
- [35] D. H. Rost, *Interpretation und Bewertung pädagogisch-psychologischer Studien: eine Einführung*. Hogrefe & Huber, 2005.
- [36] T. A. Brown, *Confirmatory Factor Analysis for Applied Research*. Guilford Press, 2006.

- [37] D. Urban and J. Mayerl, *Strukturgleichungsmodellierung*. Wiesbaden: Springer Fachmedien, 2014.
- [38] W. E. Saris, A. Satorra, and W. M. van der Veld, “Testing Structural Equation Models or Detection of Misspecifications?,” *Struct. Equ. Model. Multidiscip. J.*, vol. 16, no. 4, pp. 561–582, 2009.
- [39]
- G. W. Cheung and R. B. Rensvold, “Evaluating Goodness-of-Fit Indexes for Testing Measurement Invariance,” *Struct. Equ. Model. Multidiscip. J.*, vol. 9, no. 2, pp. 233–255, 2002.
- [40] D. H. Rost, *Interpretation und Bewertung pädagogisch-psychologischer Studien: eine Einführung*, 3rd ed. Bad Heilbrunn: Hogrefe & Huber, 2013.
- [41] P. Straube, *Kompetenzen der Erkenntnisgewinnung Physik-Lehramtsstudierender*. Freie Universität Berlin: Dissertation, in print.