GIREP & EPEC 2015
July 6 – 10

Uniwersytet Wrocławski

INTERNATIONAL CONFERENCE GIREP EPEC 2015
July 6-10, Wrocław, Poland

THE JUBILEE OF THE 70TH ANNIVERSARY OF THE POLISH ACADEMIC COMMUNITY IN WROCŁAW

Europhysics Conference
The Conference of International Research Group on Physics Teaching (GIREP)
European Physical Society - Physics Education Division (EPS PED),
University of Wrocław (UWr)

Key Competences in Physics Teaching and Learning

Proceedings

Wrocław 2016

girep

EPS

Published by

Institute of Experimental Physics

University of Wrocław

Pl. M. Borna 9, 50-204 Wrocław, Poland

ISBN: 978-83-913497-1-7

INTERNATIONAL CONFERENCE GIREP EPEC 2015
July 6-10, Wrocław, Poland

THE JUBILEE OF THE 70TH ANNIVERSARY OF THE POLISH ACADEMIC COMMUNITY IN WROCŁAW

Europhysics Conference
The Conference of International Research Group on Physics Teaching (GIREP)
European Physical Society - Physics Education Division (EPS PED),
University of Wrocław (UWr)

Key Competences in Physics Teaching and Learning

Proceedings

Editors

Ewa Dębowska, Tomasz Greczyło

Wrocław 2016

Committees

Honorary patronage
Polish Physical Society, Warsaw, Poland
Prof. dr. hab. Marek Bojarski, Rector of University of Wrocław, Poland
Cezary Przybylski, Marshal of Lower Silesia Voivodeship, Poland

Scientific Advisory Committee
Mojca Čepič, EPS-PED Committee, University of Ljubljana, Slovenia
Costas Constantinou, EPS-PED Committee, University of Cyprus, Cyprus
Ewa Dębowska, Chair of the Organizing Committee, University of Wroclaw, Poland
Leoš Dvořák, GIREP and ICPE Committee member, Charles University in Prague, Czech Republic
Ton Ellermejer, MPTL Committee member, CMA , Amsterdam, The Netherlands
Francisco Esquembre, MPTL Committee member, University of Murcia, Spain
Hendrik Ferdinande, EPS-PED Committee member, retired at University of Gent, Belgium
Raimund Girwidz, MPTL President, Universiy of Ludwigsburg, Germany
Zofia Gołąb-Mayer, Jagiellonian University, Cracow, Poland
Claudia Haagen- Schuetzenhoefer, GIREP Vicepresident, University of Graz, Austria
Zdeňka Koupilová, EPS-PED Committee member, Charles University in Prague, Czech Republic
Robert Lambourne, ICPE Committee member, The Open University, United Kingdom
Ian Lawrence, GIREP past-Vicepresident, Institute of Physics, United Kingdom
Marisa Michelini, GIREP President, University of Udine, Italy
Cesar Eduardo Mora Ley, LAPEN President, National Polytechnic Institute, Mexico
Andreas Mueller, University of Geneva, Switzerland
Hodeo Nitta, ICPE President, Tokyo Gakugei University, Japan
Wim Peeters, GIREP Vice-president, DKO (vzw) and PONTon vzw, Belgium
Gorazd Planinšič, EPS-PED past Chair, University of Ljubljana, Slovenia
Julias Salinas, IACPE and CIAREF President, Argentina
David Sands, EPS-PED Chair, University of Hull, United Kingdom
Dagmara Sokołowska, GIREP General Secretary, Jagiellonian University, Cracow, Poland
Fatih Taşar, iSER President, Gazi University, Turkey
Urbaan Titulaer, EPS-PED Committee member, retired at University of Linz, Austria
Laurence Viennot, EPS-PED Committee member, retired at University Denis Diderot Paris 7, France
Nicola Vittorio, EPS-PED Committee member, University of Rome, Italy
Stamatis Vokos, APS T-TEP Chair, University of Seattle, USA
Els de Wolf, University of Amsterdam, the Netherlands
Dean Zollmann, AAPT representative, Kansas State University, USA

Local Organizing Committee
Małgorzata Bacia, Director of Lower Silesian Centre for Teacher Training, Wrocław, Poland
Ewa Dębowska, Chair of the Organizing Committee, University of Wrocław, Poland
Tomasz Greczyło, Chair of the Local Organizing Committee, University of Wrocław, Poland
Bernard Jancewicz, Chair of Wrocław Division of Polish Physical Society
Jerzy Jarosz, University of Silesia, Katowice, Poland
Elżbieta Kawecka, Computer Assisted Education and Information Technology Centre, Warsaw, Poland
Wojciech Małecki, Director of the Regional Examination Board in Wroclaw, Poland
Piotr Skurski, University of Lodz, Poland
Dagmara Sokołowska, Jagiellonian University, Cracow, Poland
Bartosz Strzelczyk, University of Wrocław, Poland

Contents

Theoretically and Empirically Based Evaluation of Laboratory Courses – PraQ Questionnaire

Daniel Rehfeldt , Volkhard Nordmeier
Physics education research, Freie Universität Berlin, Germany

Abstract
Since 1999, when the Bologna Process was introduced, a claim for permanent screening of learning outcomes of higher education has started. Nevertheless, evaluation nowadays is mostly not based on theoretical models, and rarely accepted by participants.Two theoretically based psychological test instruments were created at FU Berlin in order to change this. Unfortunately, they ignored science laboratories. Keeping the significance of lab courses in mind, e.g. for enhancement of experimental competences, the theory based evaluation of those would clearly fill a gap. Therefore we constructed a theoretical framework and an economically usable questionnaire called PraQ.The questionnaire is based on a theoretical model for lab quality, containing three main dimensions: (a) learning gains, (b) teaching practices of tutor and (c) lab material. PraQ consists of 140 items in 40 scales, divided into two questionnaires (15min each). PraQ-A measures the growth in content knowledge, scientific inquiry practices, communication etc. PraQ-B consists of teaching practices of lab tutors, with abilities like explaining properly or summarizing. Furthermore it contains the material-dimension, which covers the lab script, integration and basic experimental material. Analysis of the piloting data demanded an exploratory factor analysis for new and highly modified items and scales. Our data source for piloting consisted of several science labs across Germany, including different disciplines and different universities, resulting in $N_A = 241$, $N_B = 237$.
We could show an eight-factor solution very close to the intended structure, having good reliability among the scales. Therefore, PraQ questionnaire now can be used widely for research on labs while additional validation is still in progress.

Keywords
Science labs, evaluation, lab theory, PraQ, LeKo, BeVakomp, experimental practice, experimental competences.

Purpose of the questionnaire: Evaluation is crucial

Since 1999, when the Bologna Process was introduced, a claim for permanent screening of learning outcomes of higher education has started (Friedrich, 2005; Hopbach, 2007). Higher education should crucially become more competence oriented in terms of educational output, especially in science education at university level. Nevertheless, nowadays common evaluation is mostly not based on profound theoretical background, and rarely accepted by participants (Csonka, 2014). As Marsh & Roche, (1997, p. 1) state, “adopting a broad construct-validation approach, recognizing […] effective teaching […]” is needed.

Academic relevance: Gap of Knowledge

Lab courses are the core of experimental training in science education (Psillos & Niedderer, 2002). Keeping this in mind, the theory based evaluation of those courses would clearly fill a gap. They aim to teach experimental practices (e.g. Schreiber, Theyßen, & Schecker, 2012) – an essential part of further studies and the future work of the students.
In order to achieve a more theoretically based evaluation of university courses in general, two theoretically based psychological test instruments, called LeKo (Thiel, Blüthmann, & Watermann, 2012) and BeVaKomp (Braun, Gusy, Leidner, & Hannover, 2008), were created, aiming at lectures and seminars of all subjects.
Unfortunately, they ignored science laboratories for their theoretically based and comprehensive research. Severe differences between science labs and classic lectures or seminars might be the cause for that. Labs are well organized courses, in which participants often have a weekly repetition of their expected actions within the course. It starts with the preparation at home with a lab script. This preparation is crucial for passing the oral pre-exam at the beginning of lab attendance. Moreover, experimental groups are small with a good staff-student-ratio, which enhances the relevance of good teaching practises by the lab teachers, e.g. regarding diagnosis of learning gains or proper explanations of the apparatus used.
Nevertheless, the LeKo and BevaKomp questionnaire build a knowledge base for constructing a theoretically grounded instrument for labs. LeKo consists of several self-reporting scales regarding teaching competences of lecturers and seminar leaders in higher education. These teaching competences consist of pedagogical as well as didactical dimensions of teaching in higher education. Some of these might be important for lab teachers, too.
BEvaKomp evaluates lectures and seminars from another perspective, looking at the gain of (meta-) competences of the students due to the respecting course. This is also realized by self-reporting scales. The

(meta-) competences are oriented on higher education key competences (Braun & Gusy, 2006), such as content knowledge, methodical competence, cooperation competence, personal competence etc. Some of these competences might associate with laboratory goals as well. Both instruments were theory-driven and evaluated empirically.
In general, evaluation forms are often realized as self-reporting scales, opening the question for the validity of such measures. As discussed by Braun et al. (2008, p. 32), a valid measure is reached by testing for construct validity of the instrument, which demands for a *validation study*. Following Messick (1995, p. 745), construct validity as a global concept can be divided into different aspects: content, substantive, structural, generalizability, external and consequential aspects. The content aspect means the content of a questionnaire being representative to the topic being measured. Substantive aspect is about, whether the assessment task itself takes the participant to think about the constructs measured. Structural aspect investigates how the relationship between scoring structure and latent structure behaves. Whether or not such structures or interpretations can be generalized across target groups is the task of generalizability aspect.
New instruments have to be tested, whether they correlate with existing measures. This is part of the external validity aspect, divided into convergent (similar construct measured) and divergent (distinct construct measured) sides. Finally, if instruments are implemented in evaluation regularly, it is important to look for consequential aspects, meaning, if actions based on the results of the measurement are fair and unbiased.
When those validity criteria are met, we can assume quite good validity, which is not effected by the method of self-reporting measurement too much (Lucas & Baird, 2006). For a broader discussion, see Braun et al. (2008, p. 32f).
Their validity is one reason, why LeKo and BEvaKomp were successfully applied at universities, especially for lectures and seminars, and now replace the standard evaluation form university wide.
Our goal was to construct a theoretical framework and an economically usable and valid questionnaire. Our target group are undergraduate lab participants, since those courses are the first step during science education at higher education for gaining experimental skills. Furthermore, the concept of undergraduate courses is more comparable than advanced ones among universities.
Aiming on a theoretically and empirically based evaluation instrument, two things are needed, (1) a literature and expert-validated theoretical model and (2) an empirical validation study.

Theoretical framework: A model for lab quality

Our theoretical framework is a model for lab quality, which was based on literature review in the first step (Rehfeldt, Mühlenbruch, & Nordmeier, 2015). We found several equalities in the organization of undergraduate labs and literature also mentioned equalities on a content oriented basis (Gutzler, Rehfeldt, & Nordmeier, 2014). For instance, labs share similar goals (Haller, 1999; Zwickl, Finkelstein, & Lewandowski, 2013), the discipline contents overlap (Lagowski, 2002, p. 1) and there is even an international consent for a »prototypic culture of experimental investigation« (translated from Emden, 2011, p. 34). Therefore, our theoretical model should aim not only on physics, but also on undergraduate science labs in general, e.g. chemistry or physiology labs.
Our model contains three main dimensions: (a) **learning gains** (growth of competences), (b) **teaching practices** of tutor and (c) lab material, with a learning-theoretical influence of quality of **learning environment** and teaching practices on learning gains. Figure 1 shows also the respective sub dimensions: For **learning gains**, content knowledge is a standard goal for lab courses, as are inquiry practices in general, since no other undergraduate course works with own experience in the experimentation process that much. Communication competence, e.g. written communication, is targeted by labs, since lab reports are sort of the first academic paper a student has to write. The sub dimension *Assessment* is based on the assessment of adequacy, plausibility and validity of experimental results required for lab course (Kreiten, 2012), which accounts for critical interpretation of your own experimental results and its implications.
Some meta-skills potentially gained in the lab course are cooperation and personal competence. Cooperation is fostered by lab courses, since experimental groups are small and work together the whole time of attendance. Personal skills consist of motivation for labwork and time management for being able to plan the preparation at home, the lab attendance and the post-processing, which all are quite time consuming and therefore train time management skills.
The *teaching practices* are divided into three sub dimensions, which represent the basic teaching skills a lab tutor should have. On one hand side, these are the more didactical dimension supporting knowledge and learning, e.g. explaining, repeating, linking etc. (Thiel et al., 2012). On the other hand, a more pedagogical factor is to motivate and create a learning atmosphere, e.g. promote self-guided learning, illustrating relevance etc. (ibid.) and to promote interaction of groups, e.g. handling disruptions (ibid.), diagnosis of learning process (Ditton & Merz, 1995; Janke, 2006) etc.

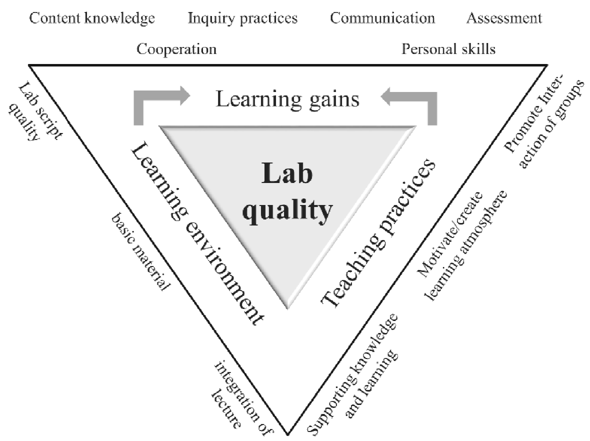

Figure 1. Theoretical model of lab quality. The three main dimensions are learning gains of the students (competences), teaching practices of the lab tutor and quality of the learning environment. Together with their respective sub dimensions, these build a model for evaluating quality of lab courses. Since learning environment and teaching practices are expected to influence learning gain, the letter one is seen as the main dimension of quality.

Regarding the *learning environment* dimension, three sub dimensions are implemented: As Zastrow (2001, p. 11) and Nagel (2009, p. 116) point out that quality of lab script has an impact on learning quality in lab courses, since it is the main source for the students' preparation at home. Integration of lecture was defined as crucial by Fraser & McRobbie in 1995, meaning the extent to which lab activities are integrated with lectures, which is quality criterion for the global goal of linking theory and practice (Haller, 1999).
The three main dimensions with their respective sub-dimensions build a theoretical framework for lab quality from three different quality perspectives: Material input, teaching input and learning output.
For designing the PraQ questionnaire, we operationalized these lab quality dimensions.

Methods: Questionnaire and validation study

The measuring tool itself consists of 140 statements (called items) in 40 scales, divided into two questionnaires PraQ-A and PraQ-B (15min each). The first part measures learning gains (a) regarding growth in content knowledge, scientific inquiry practices, communication etc. The second part contains the teaching practices of lab tutors (b) measuring abilities like explaining properly, summarizing or emphasizing relevance. Furthermore, it contains the material-dimension (c) which covers the lab script, integration of the lecture, and basic experimental material (Fraser & McRobbie, 1995; Kreiten, 2012).
For this study, results of PraQ-B are shown in details. Therefore, table 1 shows some example items of the questionnaire measuring teaching practices and material.
The validation study started in 2014 (design see fig. 2), where the literature-based items of PraQ-A and -B were reviewed by experts, namely lab instructors of physics-, chemistry-, veterinary science- and biotech-labs (substantive validity aspect). This was done by cognitive interviews (Prüfer & Rexroth, 2005) and aimed on content validity by examining the relevance and the acceptance of the items for labwork in the different fields.
The pilot study as a subpart of the validation for both questionnaires (see fig. 2) started late 2014 and ended in mid-2015. This part was meant to extract the item-structure of the instrument, handling the question, which items belong together to form so called *factors*, which could be seen as latent constructs behind the statements in the items. Some of these factors have been investigated previously, as a few scales came from established instruments. These were excluded from the pilot analysis and will only be implemented in the next step of the validation study.
For now, piloting results of PraQ-B are shown, looking at the factor structure. The analysis of the piloting data demands an exploratory factor analysis[1] for the new and highly modified items and scales.
Analyses of validation 1 and 2 are still ongoing. Validation 1 handles the question, whether found factors remain constant with a different sample (structural validity aspect) or within different labs (generalizability validity aspect). Validation 2 looks for construct validity by studying the relations between PraQ and established instruments, measuring the same or similar constructs (convergent validity) or distinct constructs (discriminant validity).

[1] To be precise a principal component analysis was done, which leads to similar results most of the time (Field, 2013, p. 638). For readability and recognition, the term »EFA« is used.

Table 1. Example Items for PraQ-B within validation-1-study.

Hypothetical construct	Rough construct description	Example item
Summarizing	Lab tutor summarizes essential aspects of theory and experiment	*The lab tutor highlights crucial aspects of the experimental setup (e.g. within discussions, while experimenting etc.)*
Check understanding	Lab tutor interacts with students to let them check their understanding	*Before continuing, the lab tutor assures, that central aspects of the experimental setup are clear.*
Illustrating relevance	Lab tutor illustrates relevance of labwork and lab topics	*The lab tutor emphasizes the relevance of theoretical issues for future studies.*
Improving self-efficacy	Lab tutor tries to improve self-efficacy of students	*The lab tutor shows trust in student's abilities.*
Diagnosis: Basic attitude	Lab tutor takes time to diagnose learning progress	*The lab tutor takes his time to explain things to students which are not or poorly understood.*
Diagnosis: just in time	Lab tutor recognizes instantly, whether comprehension is met.	*The lab tutor recognizes instantly, when a student can't follow.*
Lab script quality	Lab script is structured, useful for preparation, labwork and postprocessing	*The lab script supports me in gaining a good overview of the experimental setup.*
Integration of lecture	Extent to which laboratory activities are integrated with non-laboratory and theory lectures (See Fraser, 1995, p. 297)	*The laboratory work is related to the topics I study in the lecture.*

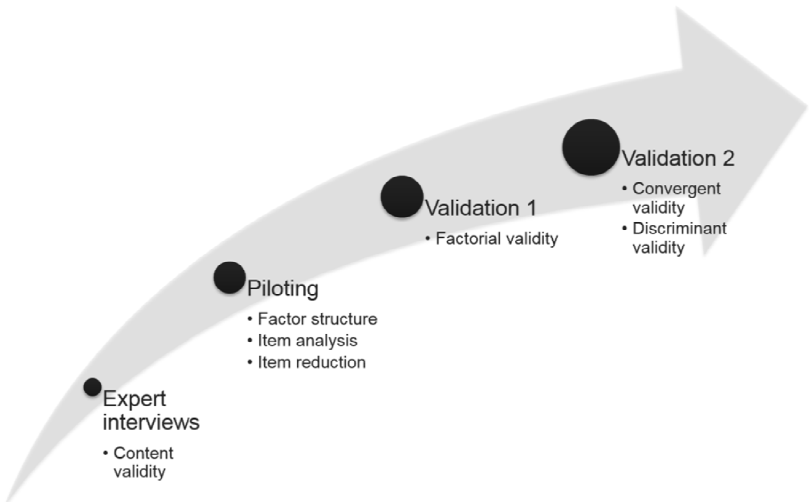

Figure 2. Design of validation study. Expert Interviews combined with a broad literature review formed the content of the theoretical model and the questionnaire. The piloting examines the structure of the questionnaire, while validation 1 analyses, whether the structure can be confirmed using a different sample. Validation 2 tries to connect PraQ measures with established instruments.

Data Source: Labs across Germany and Austria

The data source for piloting the PraQ consist of several science labs across Germany and Austria, including different disciplines (physics, chemistry, biotechnology, physiology) and different universities (Berlin, Potsdam, Wildau, Kiel, Bielefeld, Wuppertal, Tübingen, Trier, Cologne, Munich, Aachen and Vienna), resulting in N = 237 for the learning gain dimension of PraQ-A and N = 241 for teaching practices and material dimensions of PraQ-B.

Analysis & results: Good factor structure and very good reliability estimates for PraQ-B

PraQ-B was analysed by using an exploratory factor analysis (EFA), which acts on the basis of PEARSON correlations among the items and gives information on how much an item belongs to a factor (loading). Prior to and during the analysis, some items had to be dropped, caused by extremely low or high means (no gain of information due to the item), not enough correlations with other items or only low loadings (isolated item), low communalities (bad reliability of item) or practically significant double loadings (ambiguous relationship). This is summarized by table 2 and led to the elimination of eight items, which don't seriously affect content validity. Remaining 37 Items were analysed in EFA.

Table 2. Eliminated items within piloting PraQ-B and reasons for elimination, TRUE = 1; FALSE = 0.

	reasons for elimination			
Item	communality < .60	At least one crossloading >.40	No loading > .40	Not matching any factor
Feedback on Experiment	0	0	1	1
Exam feedback	0	0	1	1
Advice for experiments	0	0	0	1
Advice is focused	1	0	0	0
Explaining theory	0	1	0	1
Explaining experiment	0	0	1	0
Script shows lab report	1	0	0	0
Script combines theory and practice	1	0	0	0

A flexible approach was used, oriented on EV1 criterion and content considerations, to compare ten- to seven-factor solutions. Rotation method was chosen oblique (direct oblimin), because constructs for teaching competences are expected to correlate with each other.[1]

The factor structure of the inductive scales concerning teaching practices and material dimension shows an eight-factor solution very close to an expected structure with 74% variance explained by the factors. According to the respective items, scale names were chosen (see table 3). The reliability estimates among the scales are very good, with an α ranging from .85 to .95, indicating low measurement errors.

Conclusion: PraQ questionnaires piloting was successful: Good factor structure and great reliability estimates

The items of the PraQ questionnaire show a reasonable and reliable structure in terms of theoretical modelling. Results for PraQ-B were shown, those for PraQ-A where comparable (results for PraQ-A unpublished,). Therefore, PraQ questionnaire is now available for routine research on labs.[2]

The PraQ questionnaire is theoretically based and the empirical grounding is progressing well. The items were created by literature review and rated adequately be experts. Content of PraQ-B consists of quality of teaching practices of lab tutors and quality of lab material, such as the lab script. Teaching practices contain skills in summarizing, diagnosing, improving self-efficacy of students etc. The quality of lab script is determined by

[1] Assumptions for performing an EFA were met by the data: Normality was ok (regarding Skewness and Kurtosis). KMO measure was perfect (> .90), indicating a sufficient sample size. Bartlett Test was significant, indicating that correlations between items were large enough for EFA.

[2] Contact: danreh@zedat.fu-berlin.de

aspects such as a good structure, usability for preparation, labwork and postprocessing etc. (see table 1 for example items). PraQ-A examines learning gains of students.

Table 3. Factor structure of PraQ-B. Items are displayed on the left, factors with label on the right side. High factor loadings indicate items with higher contribution to the factor and therefore for the construct measured. All relevant loadings are at least .50 with only four very small crossloadings (see explanation below), showing quite differentiable constructs.

Item 1	.62								Summarizing	$\alpha = .90$
Item 2	.83									
…	…									
Item 6	.66									
Item 7		-.82							Check understanding	$\alpha = .85$
Item 8		-.69								
Item 9		-.57								
Item 10			.77						Illustrating relevance	$\alpha = .87$
Item 11			.80							
Item 12			.80							
Item 13				.78					Improving self-efficacy	$\alpha = .86$
Item 14				.79						
Item 15				.84						
Item 16					.55				Diagnosis: Basic attitude	$\alpha = .86$
Item 17					.53	.32				
Item 18						.85			Diagnosis: Just in time	$\alpha = .86$
Item 19						.84				
Item 20						.74				
Item 21							.86		lab script quality	$\alpha = .95$
…							…			
Item 25					-.34		.66			
Item 26					-.36		.62			
…							…			
Item 32							.79			
Item 33			.37				.41			
Item 34								.82	Integration of lecture	$\alpha = .85$
Item 35								.90		
Item 36								.81		
Item 37								.73		

Loadings < .30 are not displayed. The benchmark for loading interpretation is: Not important: < .30 | minimal: [.30, .40] | OK: [.40, .50] | practically significant: > .50 (Hair et al., 1998, p. 111). Grey colour indicates at least minimal loadings, dark grey colour highlights crossloadings, meaning two loadings for one item which are greater or equal to .30.

The ongoing validation study still needs to show, whether the item structure is invariant across samples (validation 1, see fig. 2). A confirmatory factor analysis will examine whether the found structure leads to a good model fit with a different sample of lab participants. This will be done for PraQ-A and -B.
Furthermore, inspection of construct validity is needed, proving, if PraQ really measures, what is intended. For this purpose, PraQ-A experimental competence scale is to be correlated with similar measures, namely a test on experimental competence (Straube, in print) and an experimental self-efficacy scale (Schroedter & Körner, 2012). Middle to high correlations would indicate good convergent construct validity here.
On the divergent side, PraQ-A and -B are to be compared with a measure of lab social climate (adapted from Albrecht, 2011), we expect low correlations here.
When validation is done, different labs can be compared in terms of lab quality. A first question could be for instance, whether chemistry labs differ substantially from physics labs and how one might be able to learn from each other.

Acknowledgments

The study for PraQ is located within project TSL and is supported by BMBF, project SUPPORT.

References

Albrecht, A. (2011). Längsschnittstudie zur Identifikation von Risikofaktoren für einen erfolgreichen Studieneinstieg in das Fach Physik. Berlin: Freie Universität Berlin.

Braun, E., & Gusy, B. (2006). Perspektivender Lehrevaluation. In G. Krampen & H. Zayer (Eds.), Didaktik und Evaluation in der Psychologie (pp. 152–166). Hogrefe Verlag.

Braun, E., Gusy, B., Leidner, B., & Hannover, B. (2008). Das Berliner Evaluationsinstrument für selbsteingeschätzte, studentische Kompetenzen (BEvaKomp). Diagnostica, 54(1), 30–42. http://doi.org/10.1026/0012-1924.54.1.30

Csonka, N. (2014). Evaluation von Lehrveranstaltungen an der Humboldt-Universität zu Berlin: Praxisleitfaden für Evaluationsbeauftragte an Fakultäten und Instituten. In Schriftenreihe zum Qualitätsmanagement an Hochschulen (Vol. 8). Stabsstelle Qualitätsmanagement der Humboldt-Universität zu Berlin.

Emden, M. (2011). Prozessorientierte Leistungsmessung des naturwissenschaftlich-experimentellen Arbeitens: eine vergleichende Studie zu Diagnoseinstrumenten zu Beginn der Sekundarstufe I. Berlin: Logos.

Field, A. (2013). Discovering statistics using IBM SPSS statistics. Sage.

Fraser, B. J., & McRobbie, C. J. (1995). Science Laboratory Classroom Environments at Schools and Universities: A Cross-National Study. Educational Research and Evaluation, 1(4), 289–317.

Friedrich, H. R. (2005). Der Bologna-Prozess nach Bergen. Die Hochschule, (2), 114–135.

Gutzler, T., Rehfeldt, D., & Nordmeier, V. (2014). TSL: Bedarfsanalyse in Praktika: Ein "neues" Werkzeug zur Strukturerfassung. In S. Bernholt (Ed.), Naturwissenschaftliche Bildung zwischen Science- und Fachunterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in München 2013. Münster: LIT.

Hair, J. F., Anderson, R. E., Tatham, R. L., & William, C. (1998). Multivariate data analysis (5th ed.). NJ: Prentice Hall.

Haller, K. (1999). Über den Zusammenhang von Handlungen und Zielen. Eine empirische Untersuchung zu Lernprozessen im physikalischen Praktikum. Berlin: Logos.

Hopbach, A. (2007). Qualifikationsrahmen für deutsche Hochschulabschlüsse. In Benz, Kohler, & Landfried (Eds.), Handbuch Qualität und Lehre. Berlin.

Kreiten, M. (2012). Chancen und Potenziale web-basierter Aufgaben im physikalischen Praktikum. Universität zu Köln, Köln. Retrieved from http://kups.ub.uni-koeln.de/4719/

Lagowski, J. J. (2002). THE ROLE OF THE LABORATORY IN CHEMICAL EDUCATION. Presented at the International Conference on Chemical Education, Beijing.

Lucas, R. E., & Baird, B. M. (2006). Global Self-Assessment. In M. Eid & E. Diener (Eds.), Handbook of multimethod measurement in psychology (pp. 29–42). Washington, DC, US: American Psychological Association.

Marsh, H. W., & Roche, L. A. (1997). Making students' evaluations of teaching effectiveness effective: The critical issues of validity, bias, and utility. American Psychologist, 52(11).

Messick, S. (1995). Validity of psychological assessment: validation of inferences from persons' responses and performances as scientific inquiry into score meaning. American Psychologist, 50(9), 741.

Nagel, C. C. (2009). eLearning im Physikalischen Anfängerpraktikum. Berlin: Logos.

Prüfer, P., & Rexroth, M. (2005). Kognitive Interviews (No. 15). Mannheim: Zentrum für Umfragen, Methoden und Analysen. Retrieved from http://www.gesis.org/fileadmin/upload/forschung/publikationen/gesis_reihen/howto/How_to15PP_MR.pdf

Psillos, D., & Niedderer, H. (2002). Issues and Questions Regarding the Effectiveness of Labwork. In D. Psillos & H. Niedderer (Eds.), Teaching and Learning in the Science Laboratory (pp. 21–30). Netherlands: Springer. Retrieved from http://link.springer.com/chapter/10.1007/0-306-48196-0_4

Rehfeldt, D., Mühlenbruch, T., & Nordmeier, V. (2015). Fragebogen zu Praktikumskompetenzen (PraKo): Erforschung naturwissenschaftlicher Praktika. In S. Bernholt (Ed.), Heterogenität und Diversität - Vielfalt der Voraussetzungen im naturwissenschaftlichen Unterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung 2014 (pp. 417–419). Kiel: IPN.

Schreiber, N., Theyßen, H., & Schecker, H. (2012). Diagnostik experimenteller Kompetenz: ein Verfahrensvergleich. In S. Bernholt (Ed.), Konzepte fachdidaktischer Strukturierung für den Unterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011 (pp. 263–265). Münster: LIT.

Schroedter, S., & Körner, H.-D. (2012). Entwicklung eines Fragebogens zur Selbstwirksamkeitserwartung beim Experimentieren (SWE_EX). In S. Bernholt (Ed.), Konzepte fachdidaktischer Strukturierung für den Unterricht: Gesellschaft für Didaktik der Chemie und Physik. Jahrestagung in Oldenburg 2011 (pp. 164–166). Münster: LIT.

Straube, P. (in print). Kompetenzen der Erkenntnisgewinnung Physik-Lehramtsstudierender. Freie Universität Berlin: Dissertation.

Thiel, F., Blüthmann, I., & Watermann, R. (2012). Konstruktion eines Fragebogens zur Erfassung der Lehrkompetenz (LeKo). In B. Berendt & H. P. Voss (Eds.), Neues Handbuch Hochschullehre. Lehren und Lernen effizient gestalten. [Teil] I. Evaluation. Veranstaltungsevaluation. Berlin: Raabe.

Zastrow, M. U. (2001). Interaktive Experimentieranleitungen. Berlin: Logos.

Zwickl, B. M., Finkelstein, N., & Lewandowski, H. J. (2013). The process of transforming an advanced lab course: Goals, curriculum, and assessments. American Journal of Physics, 81(1), 63–70. http://doi.org/10.1119/1.4768890

Affiliation and address information

Daniel Rehfeldt
Physics Education Research
Department of Physics
Freie Universität Berlin
Arnimallee 14
14195 Berlin, Germany
e-mail: danreh@zedat.fu-berlin.de