

Analysis of Six Reviews on the Quality of Instruments for the Evaluation of Interprofessional Education in German-Speaking Countries

Abstract

Background: More and more institutions worldwide and in German-speaking countries are developing and establishing interprofessional seminars in undergraduate education of health professions. In order to evaluate the different didactic approaches and different outcomes regarding the anticipated interprofessional competencies, it is necessary to apply appropriate instruments. Cross-cultural instruments are particularly helpful for international comparability. The Interprofessional Education working group of the German Medical Association (GMA) aims at identifying existing instruments for the evaluation of interprofessional education in order to make recommendations for German-speaking countries.

Methods: Systematic literature research was performed on the websites of international interprofessional organisations (CAIPE, EIPEN, AIPEN), as well as in the PubMed and Cinahl databases. Reviews focusing on quantitative instruments to evaluate competencies according to the modified Kirkpatrick competency levels were searched for. Psychometrics, language/country and setting, in which the instrument was applied, were recorded.

Results: Six reviews out of 73 literature research hits were included. A large number of instruments were identified; however, their psychometrics and the applied setting were very heterogeneous. The instruments can mainly be assigned to Kirkpatrick levels 1, 2a & 2b. Most instruments have been developed in English but their psychometrics were not always reported rigorously. Only very few instruments are available in German.

Conclusion: It is difficult to find appropriate instruments in German. Internationally, there are different approaches and objectives in the measurement and evaluation of interprofessional competencies. The question arises whether it makes sense to translate existing instruments or to go through the lengthy process of developing new ones.

The evaluation of interprofessional seminars with quantitative instruments remains mainly on Kirkpatrick levels 1 and 2. Levels 3 and 4 can probably only be assessed with qualitative or mixed methods. German language instruments are necessary.

Keywords: interprofessional education, evaluation, methods, validation, quality criteria

Jan P. Ehlers¹

Sylvia Kaap-Fröhlich²

Cornelia Mahler³

Theresa Scherer⁴

Marion Huber⁵

1 Witten/Herdecke University, Faculty of Health, Chair of Didactics and Educational Research in Health Science, Witten, Germany

2 University of Zurich, Faculty of Medicine, Dean's Office for Student Affairs, Zurich, Switzerland

3 University of Heidelberg, Medical Faculty, Department of General Medicine and Health Services, Heidelberg, Germany

4 Bern University of Applied Sciences, Health Division, Bachelor of Science Programme in Nursing, Bern, Switzerland

5 Zurich University of Applied Sciences, Interprofessional Teaching, Zurich, Switzerland

1. Introduction

Interprofessional collaboration in health care is growing in importance in German-speaking countries. At the same time, more and more institutions are establishing interprofessional teaching and learning scenarios in the education of health professions¹.

1.1. Background to the Evaluation of Interprofessional Education

With a history of merely 25 years, interprofessional education is a relatively new field in teaching and affiliated research using different teaching and learning approaches with respect to methodology (e.g. joint seminars or project work), seminar and course composition (e.g. different health professions with and without physicians) and content (e.g. understanding one's role, team building

or professional skills). The special edition of the GMS Journal for Medical Education [12] provides an excellent overview of projects in German-speaking countries. Systematic reviews on the evaluation of interprofessional education have shown positive effects on education as well as on the collaboration between different health professions [8], [17]. However, the studies' heterogeneity [8], a missing theoretical, competence-oriented interprofessional education framework [17] as well as missing comparable and standardized evaluation instruments posed a problem for general interpretation of the studies included in the reviews. Alongside different interprofessional competency frameworks have therefore been developed in recent years [4], [7], [11], [15]. Moreover, recommendations for planning and performing evaluation studies have been established [16] and various evaluation instruments developed and tested in different settings [3].

Prior to applying instruments, it needs to be decided which instrument is suitable for which learning outcome. In this respect, the Kirkpatrick four-level framework for evaluating the results of educational programmes serves as orientation. For the evaluation of interprofessional education [1], [5], the original four competence levels were developed further into six levels (see Table 1). In doing so, the original levels 2 "Learning" and 4 "Results" were each divided into two levels in order to distinguish learning results on the personal and the health-care level. Despite the fact that the model has been criticized and despite its limitations [20] it provides a basic and clear framework. In the meantime, the six levels have been frequently applied when evaluating educational measures in health professional training [20]. These levels therefore provide a good orientation for a practice-oriented selection of comparable evaluation instruments.

1.2. Research Objective

In order to be able to assess interprofessional teaching quality, it is necessary to use reliable and valid instruments. In a first step, existing instruments were to be identified in the literature and assessed regarding their basic, but also specific suitability for use in the German-speaking countries. Due to the manifold and heterogeneous publications in this field, it was decided not to consult individual studies but rather reviews.

Identified suitable instruments will then be made available for German-speaking countries. Based on the results, a practice-oriented handout with instruments for the evaluation of interprofessional seminars and projects shall be provided in the years to come.

The following research questions were derived from these considerations: Which instruments are recommended in the summarized review literature for the evaluation of interprofessional teaching on which Kirkpatrick evaluation level?

Which instruments are available in German and which can be recommended for translation into German respectively?

2. Method

A two phase methodological approach was necessary.

First Phase

As the field of interprofessional collaboration and teaching is developing rapidly, the websites of the relevant national and international interprofessional networks (CAIPE - Centre for the Advancement of Interprofessional Education, EIPEN - European Interprofessional Practice and Education Network, AIPEN - Australasian Interprofessional Practice and Education Network, NIPNET - Nordic Interprofessional Network) were scanned systematically for instruments for the evaluation of teaching activities.

Second Phase

In order to further address the research questions, the aim was to compile a systematic literature review. As mentioned in the Research Objective section, this systematic review was not to be based on individual studies, but rather on a meta-synthesis of reviews. The main focus was on review articles of instruments for evaluation as assessment tool, not for evaluation as complete control cycle. The evaluation tools were to be suitable for use on the teaching level, not on the collaboration level.

PubMed/Medline and CINAHL were identified as suitable literature databases. Database-specific search strategies were determined, including the search keys "review", "evaluation" and "interprofessional education" (see Attachment 1 for detailed search strategies). In a next step, publications found twice or being unsuitable as regards content were consensually eliminated by three experts of the working group. Only reviews were included which:

- Consult studies on evaluation instruments for interprofessional education
- Assign them to evaluation levels according to Kirkpatrick [5] and
- Describe the psychometric quality criteria.

Afterwards, two experts (MH, JE) independently checked the reviews being determined this way and included in the study by means of criteria stipulated in the Critical Appraisal Skills Programme Checklist (CASP; [19]). The reviews were checked for quality based on the quality of the research objective, literature research, publications, quality control and results. First, the experts filled in the checklist independently, then compared the results and after checking contentious reviews again agreed consensually on the reviews to be included.

Two independent working group experts (MH, JE) subsequently assessed the information on the test quality of the individual studies in the respective reviews. A brief characterization describes the suitable instruments determined this way. An adequate test quality description contains comprehensive details on objectivity, reliability, validity and on diagnostic psychometric properties (such as sensitivity and specificity).

Table 1: Kirkpatrick's competence levels modified for the evaluation of interprofessional education [1]

1. Reaction	Learners views on the learning experience and its interprofessional nature.
2a. Attitudes Beliefs	Changes in reciprocal attitudes or perceptions between participant groups. Changes in perception or attitude towards the value and/or use of team approaches to caring for a specific client group.
2b. Knowledge Skills	Including knowledge and skills linked to interprofessional collaboration.
3. Behaviors	Identifies individuals' transfer of interprofessional learning to their practice setting and changed professional practice.
4a. Client Benefit	Improvements in health or well-being of patients/clients.
4b. Organizational Change	Wider changes in the organisation and delivery of care.

Particularly successful tools as well as German-language tools were individually and critically discussed again and presented.

3. Results

First Phase

The scanning of the interprofessional networks yielded two repositories for evaluation tools:

- Canadian Interprofessional Health Collaborative [15]
- National Center for Interprofessional Practice and Education [3].

Both collections comprehensively present tools, but also show limitations. They do not only focus on interprofessional education, but also consider instruments for evaluating interprofessional collaboration. The access to these databases is rather open, enabling a free download of tools. However, the integrated tools are not subject to critical evaluation and quality management, rendering their validity and reliability uncertain. The review at hand therefore does not include any tools of these databases.

Second Phase

The PubMed search produced 24 hits while the Cinahl search produced 49 hits (see Attachment 1 and Figure 1). After scanning the abstracts, 63 publications could be excluded as they did not deal with the sought-for topic or did not comply with the requirements agreed upon (e.g. interprofessional collaboration or qualitative methods). Out of the ten suitable reviews, the studies of four reviews were not assigned to the Kirkpatrick levels [5]. The working group assessed the six remaining reviews using the CASP checklist and went through the full text in search of valid and suitable tools (see Table 2 and Figure 2).

The evaluation of the reviews shows different results regarding the quality of the included reviews and also regarding the recommended measuring instruments. Three reviews are high-quality reviews [9], [10], [14]. All three reviews are based on standards (e.g. Cochrane Guidelines for Systematical Reviews or Standards for evaluating the assessment tools included in the studies). Furthermore, these three reviews clearly assign identified evaluation tools to the Kirkpatrick levels. The review at hand also includes a medium-quality review [6] and two low-quality reviews [2], [20]. These three reviews primarily either performed an exploratory search or the literature review was part of a multi-method survey. The fact that the

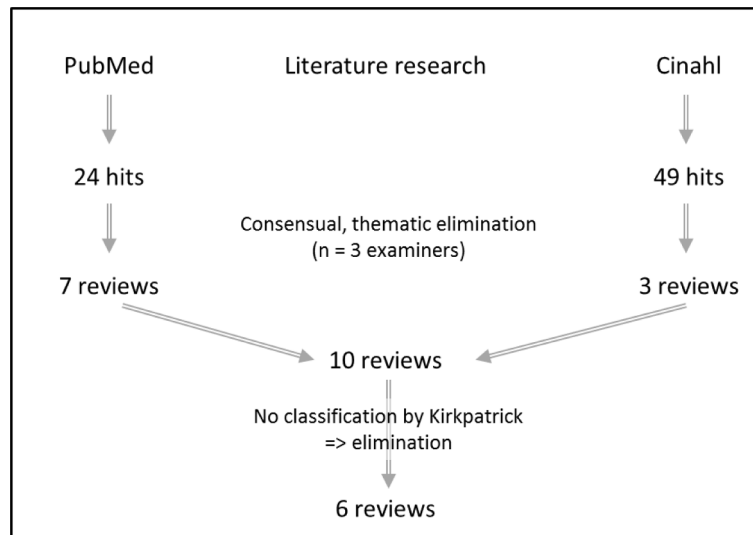


Figure 1: Flowchart demonstrating results of the literature search

studies included in these reviews were not critically examined significantly reduces the validity of the reviews. Moreover, an evaluation of the assessments regarding their psychometric properties took place only partly or not at all. Cronbach's Alpha is the value most mentioned in this context. There was no information on further psychometric properties available.

Havyer et al. [9] are the only ones to make a clear recommendation based on the test quality criteria. From their point of view, the Collaborative Healthcare Interdisciplinary Relationship Planning (CHRIP), the Readiness for Interprofessional Education Scale (RIPLS), the Communication and Teamwork Skills assessment (CATS) and the Teamwork Mini-Clinical Evaluation Exercise (T-MAX) tools meet the quality criteria. Their 2014 publication does expressly state but not explicitly recommend the Safety Attitudes Questionnaire (SAQ) and the Team Climate Inventory (TCI). Recommendations of Havyer et al. [10] refer to a specific competency framework of the Association of American Medical Colleges (AAMC). Cronbach's Alpha as a measure for reliability was the only indicator for quality criteria mostly stated by Havyer et al. [9], [10] for the included studies. They did not report on factor-analytical findings for indicating content validity.

After having determined own minimum test theory quality criteria, Oates and Davidson [14] conclude that only the Interprofessional Socialization and Valuing Scale (ISVS) meets all minimum quality standards. Five further instruments (Interprofessional Collaborator Assessment Rubric - ICAR, Attitudes Towards Teamwork in Training Undergoing Designed Educational Simulation - KidSIM ATTITUDES Questionnaire, Interdisciplinary Education Perception Scale - IEPS and University of the West of England Interprofessional Questionnaire - UWEIPQ) only partly comply with the minimum standards. Three instruments (RIPLS, Attitudes to Shared Learning - ASL and StudDat Questionnaire) do either not meet the standards at all or just insufficiently.

Gillan et al. [6] offer a comprehensive overview of evaluation instruments and categorize all items from the found

instruments. As the literature research was described in detail, this review can be regarded as medium-quality review. The clear assignment of the items contained in the evaluation tools to the Kirkpatrick levels [5] is particularly useful in practice. The authors therefore offer a very transparent overview and show that an assignment to levels 1 to 4a is possible. No item could be assigned to level 4b though.

Blue et al. [2] exclusively included reviews in their article. However, their description of the kinds of reviews they included in their review article is imprecise. A critical assessment of the included reviews is non-existent. The exploratory literature research, however, was performed as addition to a "mixed method" survey, the content of which does not refer to the present research objective. Therefore, only the review part of Blue et al. [2] was reviewed for quality.

Thistlethwaite et al. [20] were primarily looking for intervention studies containing instruments for evaluating interprofessional education and collaboration in their review. It should be noted that the authors searched directly in journals only and did not scan any databases. Moreover, the search keys of this literature research are not evident. Thistlethwaite et al. [20] recommend a multi-method survey of interprofessional competencies.

The overall conclusion is that a wide range of instruments has been developed, which are applied on different levels of the modified Kirkpatrick framework [2]. The scanning of full texts showed that presumably there is no quantitative evaluation instrument for level 4b [1].

Furthermore, validation and quality criteria, evidence level and evaluation levels according to the modified Kirkpatrick framework [1] were very heterogeneous.

Nevertheless, the following four instruments meeting the search requirements, i.e. assignment to one or more Kirkpatrick levels and/or interprofessional competencies, could be determined: Readiness for Interprofessional Learning Scale (RIPLS), University of the West of England Interprofessional Questionnaire (UWEIPQ), Interprofessional Socialization and Valuing Scale (ISVS) and Team

Table 2: Six review articles about the quality of evaluation tools for interprofessional education rated by CASP (Critical Appraisal Skills Programme Checklist)

Publications	Research questions	Literature research	Included publications	Quality assessment	Results	Quality of the review
GILLAN et al. (2011): The evaluation of learner outcomes in interprofessional continuing education: A literature review and an analysis of survey instruments	What IPE evaluation tools are available and how do they cover the levels of Kirkpatrick/Barr?	MEDLINE, CINAHL, Manual search. Search terms in the paper.	All studies, insofar as they contain instruments or data on known instruments.	Evaluation at item level.	Studies: 202 Tools: 33 (538 items, Lvl 1: 111, Lvl 2a: 336, Lvl 2b: 7, Lvl 3: 51, Lvl 4a: 2 (TCI and Garrad 2006), Lvl 4b: 0). Recommended tools: 0	Quality: Medium. Included studies were not critically appraised. No statement regarding the overall quality of the review possible.
HAVYER et al. (2013): Teamwork assessment in internal medicine: A systematic Review of validity evidence and outcomes	What tools are available in internal medicine for the teamwork in the field of education and training (medical students, doctors of all levels, interprofessional teams) and how are their quality criteria?	MEDLINE, MEDLINE In-process, CINAHL, Manual search. Search terms in the paper.	Original papers esp. in internal medicine but also IPE, only quantitative tools.	Validity framework, Kirkpatrick, MERSQI, Cochrane Criteria.	Studies: 179 Tools: 73 (53 with high validity, 13 with „patient outcomes“ correlated, mentioned above SAQ and TCI). Recommended tools: 0	Quality: High. Cochrane criteria, but inclusion of non-randomized studies. Inclusion criteria were critically appraised by their quality.
BLUE et al. (2015): Assessment and Evaluation in Interprofessional Education - Exploring the Field	How can IPE be best tested and evaluated at the moment? Where are still gaps to fill?	Interviews, research meetings, review, MEDLINE, ERIC, PsychInfo, CINAHL, ISI Web of Knowledge, Manual search, 1960-2012. Search terms in the paper.	Systematic reviews.	Not apparent.	Reviews: 7 Tools: 12 (Unknown tools should be presented better). Recommended tools: 0	Quality: Low. Review part of a multi-method study. Explorative study without quality assessment.
THISTLETHWAITE et al. (2015): An exploratory review of pre-qualification interprofessional education evaluations	What evaluation tools are available for IPE and how do they focus on outcomes and can they be allocated to Kirkpatrick?	Journal of Interprofessional Care, Medical Education, Focus on Health Professional Education and Nurse Education, Journal of Research in Interprofessional Education. No clear search strategy.	All studies as long as they contain instruments.	Not apparent.	Studies: 90 Tools: 4 (5 studies focussed on outcomes). Recommended tools: 0	Quality: Low. Only direct journal search, no databases. No mention of keywords. Results tables separately online. No critical appraisal of studies included.

(Continued)

Table 2: Six review articles about the quality of evaluation tools for interprofessional education rated by CASP (Critical Appraisal Skills Programme Checklist)

Publications	Research questions	Literature research	Included publications	Quality assessment	Results	Quality of the review
OATES and DAVIDSON (2015): A critical appraisal of instruments to measure outcomes of interprofessional education	Which evaluation tools with a high psychometric standard are available for IPE and collaborative practice among health care students?	CIHC inventory, Thannhauser et al. 2010, new search with same criteria.	Original papers with psychometric data.	QuAILS and own checklist.	Tools: 140 (9 included tools, only SVS covers standards and Lvl 2a+b, 3). Recommended tools: 0	Quality: High. Systematic approach. Test theoretic minimal quality criteria for included tools.
HAYVER (2016): Addressing the Interprofessional Collaboration Competencies of the Association of American Medical Colleges: A Systematic Review of Assessment Instruments in Undergraduate Medical Education	What is the validity and characteristics of tools for evaluating teamwork in medical students and interprofessional behavior?	MEDLINE, MEDLINE In-process, CINAHL, PsycINFO (1979-2014). Searchterms in digital appendix.	Original papers of all disciplines with medical students involved in IPE.	MERSQI, Framework, AAMC Competencies, Kirkpatrick.	Studies: 70 Tools: 64 Recommended tools: 4 (CHIRP, R/PLS, CATS, T-MEX)	Quality: High. Careful review of the psychometric properties of the instruments. Quality assessment of the included studies.

Remark: ¹ Lvl = Kirkpatrick Level

Publication	CA carried out	Results		Quality judgement of the review		
		n tools	Recommendation	high	medium	low
HAVYER et al. (2013)	yes	73	no	✗		
GILLAN et al. (2011)	yes	33	no		✗	
BLUE et al. (2015)	no	12	no			✗
THISTLETHWAITE et al. (2015)	no	4	no			✗
OATES und DAVIDSON (2015)	yes	140	no	✗		
HAYVER (2016)	yes	70	CHIRP, RIPLS, CATS, T-MEX	✗		

Figure 2: Overview of the consensual evaluation of six review articles by two examiners (CA=Critical Appraisal)

Climate Inventory (TCI). Attachment 2 contains a brief description of the individual instruments.

4. Discussion

The already existing online repositories of tools for evaluating interprofessionalism have shown that it is inevitable to focus on specific questions and quality assurance regarding usability and transparency. The present review reveals that the unsystematic approach in the scanned reviews and thus the missing critical assessment of the studies included in the reviews proved to be problematic. In conclusion, the heterogeneous situation regarding the quality of the reviews included in this study only allows for a limited recommendation. From the test theory perspective, the recommendation of Havyer et al. [9] should be considered. However, as it is not clearly evident how the recommendation came about, it should be treated with caution as well because in particular the use of RIPLS has been critically discussed in the meantime due to its quality criteria [13], [18]. None of the reviews mentions psychometric measures such as sensitivity or specificity regarding test methods. Moreover, construct validity verification has not been described for any of the mentioned instruments. The review only in part reports that a factor analysis was performed. The specific construct to be reviewed by means of the tools can only be measured content-wise but is not verifiable by means of test theory methods. A review of different evaluation instruments regarding concurrent validity would be beneficial at this point. Information on objectivity and inter rater reliability, respectively, is missing as well, but would be important with respect to the test quality. The fact that most of the time the authors only report on the instrument reliability indicates a lack of theoretical orientation towards a framework when developing the instruments. This orientation, however, would enable a validity check. Various instruments are available measuring interprofessional education by evaluating outcomes on different Kirkpatrick levels. Outcome and Kirkpatrick level have to be considered when selecting or applying an evaluation instrument. Only a few reliable and valid instruments are

available yet. Assigning an outcome to a competency framework may be helpful when choosing an instrument. Moreover, not all Kirkpatrick levels can be mapped by quantitative assessments. It seems to be necessary to include qualitative evaluation as well.

TCI and ISVS are mentioned in various reviews. Oates and Davidson [14] describe the ISVS as the only instrument meeting the test quality criteria. Moreover, the ISVS can be assigned to three Kirkpatrick levels and is also mentioned as instrument in all other reviews. The RIPLS is certainly considered as one of the instruments being applied most and it would also be available in German. However, it is not recommendable from a test theory perspective, as critically commented by Oates and Davidson [14] as well. The UWEIPQ is also an often-used instrument and available in German, but only partly meets the quality criteria.

Finally, the ISVS is the only instrument to be recommended from a test theoretical perspective at the moment. As this high-quality tool is already available and does not need to be redeveloped, the working group is currently translating and localizing the ISVS. It will also validate the German version.

5. Conclusion

The results of this review show that the transition from education to practice also needs to come more into focus. This includes evaluations on team level, an analysis of the transition and the development of suitable instruments and benchmarks for ensuring business success and organizational modifications (e.g. patient safety, job satisfaction, critical appraisal).

Another working group of the GMA Committee for Interprofessional Education is currently working on a German competency framework for interprofessional education. After the compilation and approval of this competency framework, the suitability of the recommended instruments shall be reviewed again. Afterwards, it can be decided on whether renewed, adapted recommendations or new instruments appropriate to the complexity shall be developed.

Notes

¹ In this context, health professions comprise professions in the fields of medicine, dentistry, nursing, therapy and diagnostics.

Competing interests

The authors declare that they have no competing interests.

Attachments

Available from

<http://www.egms.de/en/journals/zma/2017-34/zma001113.shtml>

1. Attachment 1.pdf (109 KB)
Strategies, search terms and results of the literature research
2. Attachment 2.pdf (94 KB)
Short presentation of individual instruments

References

1. Barr H, Freeth D, Hammick M, Koppel I, Reeves S. Evaluations of interprofessional education. London: United Kingdom Review of Health and Social Care; 2000.
2. Blue AV, Chesluk BJ, Conforti LN, Holmboe ES. Assessment and Evaluation in Interprofessional Education: Exploring the Field. *J Allied Health*. 2015;44(2):73-82.
3. Brandt BF. Update on the US national center for interprofessional practice and education. *J Interprof Care*. 2014;28(1): 5-7. DOI: 10.3109/13561820.2013.852365
4. Brewer M. Curtin University's faculty of health sciences interprofessional capability framework. Perth: Curtin University; 2011. Zugänglich unter/available from http://healthsciences.curtin.edu.au/local/docs/IP_Capability_Framework_booklet.pdf
5. Freeth D, Hammick M, Koppel I, Reeves S, Barr H. A critical review of evaluations of interprofessional education. London, UK: LTSN Health Science and Practice; 2002. S.63
6. Gillan C, Lovrics E, Halpern E, Wiljer D, Harnett N. The evaluation of learner outcomes in interprofessional continuing education: a literature review and an analysis of survey instruments. *Med Teach*. 2011;33(9):461-470. DOI: 10.3109/0142159X.2011.587915
7. Gordon F. Combined Universities Interprofessional Learning Unit: Final Report. Sheffield, UK: Sheffield Hallam University; 2006.
8. Hammick M, Freeth D, Koppel I, Reeves S, Barr H. A best evidence systematic review of interprofessional education: BEME Guide no. 9. *Med Teach*. 2007;29(8): 35-751. DOI: 10.1080/01421590701682576
9. Havyer RD, Nelson DR, Wingo MT, Comfere NI, Halvorsen AJ, McDonald FS, Reed DA. Addressing the Interprofessional Collaboration Competencies of the Association of American Medical Colleges: A Systematic Review of Assessment Instruments in Undergraduate Medical Education. *Acad Med*. 2016;91(6):865-888. DOI: 0.1097/ACM.0000000000001053
10. Havyer RD, Wingo MT, Comfere NI, Nelson DR, Halvorsen AJ, McDonald FS, Reed DA. Teamwork assessment in internal medicine: a systematic review of validity evidence and outcomes. *J Gen Intern Med*. 2014;29(6):894-910. DOI: 10.1007/s11606-013-2686-8
11. Interprofessional Education Collaborative Expert Panel. Core competencies for interprofessional collaborative practice. Washington, D.C.: Interprofessional Education Collaborative; 2011. Zugänglich unter/available from: <http://www.aacn.nche.edu/education-resources/ipcreport.pdf>
12. Klapper B, Schirlo C. Themenheft Interprofessionelle Ausbildung – herausgegeben von der Robert Bosch Stiftung und der Gesellschaft für Medizinische Ausbildung. *GMS J Med Educ*. 2016;33(2):Doc38. DOI: 10.3205/zma001037
13. Mahler C, Berger S, Pollard K, Krisam J, Karstens S, Szecsenyi J, Krug K. Translation and psychometric properties of the German version of the University of the West of England Interprofessional Questionnaire (UWE-IP). *J Interprof Care*. 2016 (in press)
14. Oates M, Davidson M. A critical appraisal of instruments to measure outcomes of interprofessional education. *Med Educ*. 2015;49(4):386-398. DOI: 10.1111/medu.12681
15. Orchard C, Bainbridge L, Bassendowski S, Stevenson K, Wagner SJ, Weinberg L, Curran V, Di Loreto L, Sawatzky-Girling B. A national interprofessional competency framework. Vancouver: Canadian Interprofessional Health Collaborative (CIHC); 2010.
16. Reeves S, Boet S, Zierler B, Kitto S. Interprofessional Education and Practice Guide No. 3: Evaluating interprofessional education. *J Interprof Care*. 2015;29(4):305-312. DOI: 10.3109/13561820.2014.1003637
17. Reeves S, Goldman J, Burton A, Sawatzky-Girling B. Synthesis of systematic review evidence of interprofessional education. *J All Health*. 2010;39(Supplement 1):198-203.
18. Schmitz CC, Brandt BF. The Readiness for Interprofessional Learning Scale: To RIPLS or not to RIPLS? That is only part of the question. *J Interprof Care*. 2015;29(6):525-526 DOI: 10.3109/13561820.2015.1108719
19. Singh J. Critical appraisal skills programme. *J Pharmacol Pharmacother*. 2013;4(1):76. DOI: 10.4103/0976-500X.107697
20. Thistlethwaite J, Kumar K, Moran M, Saunders R, Carr S. An exploratory review of pre-qualification interprofessional education evaluations. *J Interprof Care*. 2015;29(4):292-297. DOI: 10.3109/13561820.2014.985292

Corresponding author:

Prof. Dr. Jan P. Ehlers, M.A.
Witten/Herdecke University, Faculty of Health, Chair of Didactics and Educational Research in Health Science, Alfred-Herrhausen-Str. 50, D-58448 Witten, Germany, Tel.: +49 (0)2302/92-7786
jan.ehlers@uni-wh.de

Please cite as

Ehlers JP, Kaap-Fröhlich S, Mahler C, Scherer T, Huber M. Analysis of Six Reviews on the Quality of Instruments for the Evaluation of Interprofessional Education in German-Speaking Countries. *GMS J Med Educ*. 2017;34(3):Doc36. DOI: 10.3205/zma001113, URN: <urn:nbn:de:0183-zma0011137>

This article is freely available from

<http://www.egms.de/en/journals/zma/2017-34/zma001113.shtml>

Received: 2016-06-17
Revised: 2017-01-16
Accepted: 2017-05-04
Published: 2017-08-15

Copyright

©2017 Ehlers et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Auswertung von sechs Übersichtsartikeln über die Qualität von Evaluationsinstrumenten zur Beurteilung interprofessioneller Lehre im deutschsprachigen Raum

Zusammenfassung

Hintergrund: Sowohl weltweit als auch in den deutschsprachigen Ländern erfolgt zunehmend die Entwicklung und Etablierung von interprofessionellen Lehr- und Lernszenarien in den Ausbildungen der Gesundheitsberufe. Um die unterschiedlichen didaktischen Herangehensweisen wie auch die unterschiedlichen, situationsabhängigen Ergebnisse der Vermittlung interprofessioneller Kompetenzen zu evaluieren, ist es erforderlich, geeignete Instrumente einzusetzen. Dabei sind gerade interkulturelle Instrumente hilfreich, um eine internationale Vergleichbarkeit zu erzielen. Die GMA-Arbeitsgruppe „Interprofessionelle Ausbildung in den Gesundheitsberufen“ hat sich zum Ziel gesetzt, bestehende Instrumente zur Evaluation interprofessioneller Lehre zu identifizieren, um Empfehlungen für den deutschsprachigen Raum auszusprechen.

Methoden: Eine systematische Literaturrecherche auf den Internetseiten zur interprofessionellen Ausbildung internationaler Fachorganisationen (CAIPE, EIPEN, AIPEN) sowie in den Datenbanken PubMed und Cinahl wurde durchgeführt. Reviews mit Hauptaugenmerk auf quantitativen Instrumenten zur Evaluation des interprofessionellen Kompetenzerwerbs auf Grundlage der modifizierten Kirkpatrick Kompetenzstufen wurden gesucht. Diese wurden hinsichtlich ihrer psychometrischen Eigenschaften, Sprache/Land und Voraussetzungen des Instrumenteinsatzes untersucht und erfasst.

Ergebnisse: Es wurden sechs Reviews aus 73 Treffern der Literatursuche inkludiert. Eine Großzahl an Instrumenten wurde identifiziert, deren Gütekriterien und Einsatz allerdings sehr heterogen war, weshalb sie nicht in die Untersuchung aufgenommen wurden. Die Instrumente sind vorwiegend den Kirkpatrick Stufen 1, 2a & 2b zuzuordnen. Die meisten wurden in englischer Sprache entwickelt. Gütekriterien wurden selten ausführlich berichtet. Nur sehr wenige Instrumente sind in deutscher Sprache verfügbar.

Schlussfolgerung: Es ist schwierig geeignete deutschsprachige Instrumente zu finden. International gibt es unterschiedliche Ansätze und Zielsetzungen in der Evaluation und Messung der interprofessionellen Kompetenzen, die sich z.B. nicht allein auf die Ausbildung sondern oft auf das spätere Handeln im Team beziehen. Es stellt sich die Frage, ob es sinnvoll ist, bereits bestehende Instrumente zu übersetzen oder den langen Prozess der eigenen Entwicklung zu verfolgen.

Die Evaluation von interprofessioneller Lehre mit quantitativen Instrumenten erfolgt bisher vorwiegend auf den Stufen 1 und 2. Die Stufen 3 und 4 können größtenteils wahrscheinlich nur mit qualitativen Methoden oder „mixed methods“ erfasst werden. Neben dem auf deutsch vorliegendem UWEIPQ ist die Entwicklung weiterer Instrumente in deutscher Sprache oder die Lokalisierung (Übersetzung und Anpassung an nationale Gegebenheiten) englischer Instrumente (z.B. ISVS) notwendig.

Schlüsselwörter: interprofessionelle Ausbildung, Evaluation, Methoden, Validierung, Gütekriterien

Jan P. Ehlers¹

Sylvia Kaap-Fröhlich²

Cornelia Mahler³

Theresa Scherer⁴

Marion Huber⁵

1 Universität Witten/Herdecke, Fakultät für Gesundheit, Didaktik und Bildungsforschung im Gesundheitswesen, Witten, Deutschland

2 Universität Zürich, Medizinische Fakultät, Studiendekanat, Zürich, Schweiz

3 Universität Heidelberg, Medizinische Fakultät, Abteilung Allgemeinmedizin und Versorgungsforschung, Heidelberg, Deutschland

4 Berner Fachhochschule, Fachbereich Gesundheit, Bachelorstudiengang Pflege, Bern, Schweiz

5 Zürcher Hochschule für Angewandte Wissenschaften, Interprofessionelle Lehre (IPL), Zürich, Schweiz

1. Einleitung

In den deutschsprachigen Ländern gewinnt die interprofessionelle Zusammenarbeit im Gesundheitswesen zunehmend an Bedeutung. Parallel werden auch immer mehr interprofessionelle Lehr- und Lernszenarien in den Ausbildungen der Gesundheitsberufe¹ etabliert.

1.1. Hintergrund zur Evaluation von Interprofessioneller Lehre

Interprofessionelle Lehre (IPL) ist mit einer ca. 25-jährigen Geschichte ein relativ junger Bereich in der Lehre und der angegliederten Forschung. Hierbei finden sich ganz unterschiedliche Lehr- und Lernansätze, sowohl was die Methodik (z.B. gemeinsame Seminare oder Projektarbeiten), die Zusammensetzung der Veranstaltungen (z.B. verschiedene Gesundheitsberufe mit und ohne Mediziner) und auch die Inhalte (z.B. Rollenverständnis, Teambildung oder fachliche Kompetenzen) angeht. Einen sehr guten Überblick über Projekte im deutschsprachigen Raum bietet das Themenheft der *GMS Journal for Medical Education* [12]. Systematische Reviews zur Evaluation von IPL konnten zunehmend positive Effekte sowohl in der Ausbildung als auch in der späteren gemeinsamen Tätigkeit von unterschiedlichen Gesundheitsberufen aufzeigen [8], [17]. Probleme in der übergreifenden Interpretation stellten die Heterogenität der Studien [8], das Fehlen eines theoretischen, kompetenzorientierten interprofessionellen Rahmens für die Ausbildung [17] sowie vergleichbare und standardisierte Evaluationsinstrumente dar. Deshalb wurden in den letzten Jahren unterschiedliche interprofessionelle Kompetenzrahmen entwickelt [4], [7], [11], [15]. Außerdem wurden Empfehlungen für die Planung und Durchführung von Evaluationsstudien interprofessioneller Lehre ausgesprochen [16] und eine Vielzahl von Instrumenten zur Evaluation entwickelt und in unterschiedlichen Settings überprüft [3], [15].

Für die Anwendung von Instrumenten stellt sich die Frage, welche für welches Lernergebnis (Outcome) geeignet sind. Als Orientierung bietet sich das Kirkpatrick-Rahmenmodell an, das für die Evaluation von Bildungsmaßnahmen entwickelt wurde und auf vier Stufen die Lernergebnisse misst. Die ursprünglichen vier Evaluationsstufen wurden für die Evaluation von IPL [1], [5] zu sechs Stufen weiterentwickelt (siehe Tabelle 1). Dabei wurden die ursprüngliche Stufe 2 „Lerner“ und Stufe 4 „Ergebnisse“ in jeweils zwei Stufen aufgeteilt, um eine Unterscheidung der Lernergebnisse auf Personen- und Versorgungsebene zu erreichen. Trotz der Kritik an diesem Modell und der Limitationen [20] stellt es einen ersten klaren Rahmen dar. Die sechs Stufen finden inzwischen bei der Evaluation von Maßnahmen in den Gesundheitsberufen häufig Anwendung [20]. Somit stellen diese Stufen grundsätzlich eine gute Orientierung für die praxisorientierte Auswahl von vergleichbaren Evaluationsinstrumenten dar.

1.2. Fragestellung

Um die Qualität von IPL beurteilen zu können, ist es erforderlich reliable und valide Instrumente einzusetzen. In einem ersten Schritt sollten dazu bereits bestehende Instrumente in der Fachliteratur identifiziert und auf ihre grundsätzliche, aber auch spezielle Eignung für den deutschsprachigen Raum beurteilt werden. Aufgrund der vielfältigen und heterogenen Publikationen auf diesem Gebiet wurde beschlossen, nicht einzelne Studien heranzuziehen, sondern von Übersichtsarbeiten (Reviews) auszugehen

Geeignete Instrumente sollten anschließend für den deutschsprachigen Raum verfügbar gemacht werden. Aus den Ergebnissen soll in den folgenden Jahren eine praxisnahe Handreichung mit Instrumenten für die Evaluation von interprofessionellen Lehrveranstaltungen und Projekten geboten werden.

Vor diesem Hintergrund stellten sich folgende Forschungsfragen: Welche Instrumente werden in der zusammengefassten Übersichtsliteratur für die Evaluation interprofessioneller Lehre auf welcher Evaluationsstufe nach Kirkpatrick empfohlen?

Welche Instrumente stehen in deutscher Sprache zur Verfügung bzw. können für die Übersetzung ins Deutsche empfohlen werden?

2. Methode

Methodisch waren zwei Phasen des Herangehens notwendig.

1. Phase

Da sich das Feld der interprofessionellen Zusammenarbeit und Lehre rasant entwickelt, wurden systematisch die Internetseiten der einschlägigen nationalen und internationalen, interprofessionellen Netzwerke (CAIPE - Centre for the Advancement of Interprofessional Education, EIPEN - European Interprofessional Practice and Education Network, AIPEN - Australasian Interprofessional Practice and Education Network, NIPNET - Nordic Interprofessional Network) nach Instrumenten zur Evaluation von Lehraktivitäten durchsucht.

2. Phase

Zur weiteren Beantwortung der Fragestellung sollte ein systematisierter Literaturreview erstellt werden. Wie im Hintergrund beschrieben, sollten hier nicht einzelne Studien herangezogen werden, sondern im Sinne einer Metasynthese von Übersichtsarbeiten (Reviews) ausgegangen werden. Dabei wurde das Hauptaugenmerk auf Reviews zu Instrumenten für die Evaluation als Bewertungstool, nicht für die Evaluation als kompletten Steuerungszyklus gelegt. Der Einsatz der Evaluationstools sollte auf Lehrebene und nicht auf Zusammenarbeitsebene erfolgen.

Tabelle 1: Evaluationsmodell in Kompetenzstufen nach Kirkpatrick modifiziert für die Evaluation interprofessioneller Ausbildung [1]

Original version		Deutsche Übersetzung	
1. Reaction	Learners views on the learning experience and its interprofessional nature.	1. Reaktion	Vorstellung von Lernenden bezüglich der interprofessionellen Erfahrung und ihrem interprofessionellen Charakter.
2a. Attitudes Beliefs	Changes in reciprocal attitudes or perceptions between participant groups. Changes in perception or attitude towards the value and/or use of team approaches to caring for a specific client group.	2a. Einstellungen / Vorstellungen zur Zusammenarbeit	Veränderungen in der wechselseitigen Einstellung oder Wahrnehmung zwischen den teilnehmenden Gruppen. Veränderungen in der Wahrnehmung oder Haltung gegenüber dem Wert und/oder Nutzen eines Teamansatzes in der Versorgung einer bestimmten Personengruppe.
2b. Knowledge Skills	Including knowledge and skills linked to interprofessional collaboration.	2b. Wissen und Fertigkeiten	Einbeziehen von Wissen und Fertigkeiten, die mit interprofessioneller Zusammenarbeit verknüpft sind.
3. Behaviors	Identifies individuals' transfer of interprofessional learning to their practice setting and changed professional practice.	3. Verhalten	Identifiziert den Transfer interprofessionellen Lernens Einzelner in die berufliche Praxis sowie eine Veränderung professioneller Praxis.
4a. Client Benefit	Improvements in health or well-being of patients/clients.	4a. Patienten-Nutzen	Verbesserung der Gesundheit oder das Wohlbefindens von Patienten/Klienten.
4b. Organizational Change	Wider changes in the organisation and delivery of care.	4b. Organisations-bezogene Veränderungen	Weitreichende Veränderungen in der Organisation und der Gesundheitsversorgung.

Als passende Literaturdatenbanken für die Recherche wurden PubMed/Medline und CINAHL identifiziert. Es wurden datenbankspezifische Suchstrategien mit den Suchbegriffen „review“, „evaluation“ und „interprofessional education“ formuliert (genaue Suchstrategien siehe Anhang 1). In einem nächsten Schritt wurden konsensual die doppelt gefundenen oder thematisch nicht passenden Publikationen von drei Experten der Arbeitsgruppe eliminiert. Dabei wurden nur Reviews inkludiert, die:

- Studien über Evaluationsinstrumente für die interprofessionelle Ausbildung heranziehen
- diese den Evaluationsstufen nach Kirkpatrick [5] zuordnen und
- die psychometrischen Gütekriterien beschreiben.

Die so ermittelten, eingeschlossenen Reviews wurden danach kriteriengeleitet mittels der Critical Appraisal Skills Programme Checkliste (CASP, [19]) unabhängig von zwei Experten (MH, JE) auf ihre Qualität anhand der Güte der Forschungsfrage, der Literaturrecherche, der Publikationen, der Qualitätsüberprüfung und der Ergebnisse beurteilt, wobei zunächst ein unabhängiges Ausfüllen der Checkliste erfolgte, dann ein Vergleich der Ergebnisse und unter erneutem Hinzuziehen strittiger Reviews eine konsensuale Ergebnisfindung.

Die Angaben zur Testgüte der einzelnen Studien in den jeweiligen Reviews wurden anschließend von zwei unabhängigen Experten (MH, JE) der Arbeitsgruppe beurteilt. Die so ermittelten geeigneten Instrumente wurden in einer Kurzcharakteristik beschrieben. Als geeignete Beschreibung der Testgüte wurden umfassende Angaben zur Objektivität, Reliabilität, Validität sowie zu diagnostisch-psychometrischen Kennzahlen (wie Sensitivität und Spezifität) gewertet.

Besonders erfolgreiche Werkzeuge sowie in Deutsch vorliegende wurden nochmal einzeln kritisch diskutiert und vorgestellt.

3. Ergebnisse

1. Phase

Bei der Sichtung der interprofessionellen Netzwerke wurden zwei Repositorien für Evaluationstools gefunden:

- Canadian Interprofessional Health Collaborative [15]
- National Center for Interprofessional Practice and Education [3].

Beide Sammlungen stellen sehr umfangreiche Werkzeuge vor, zeigen allerdings auch Limitationen. So beschränken sie sich nicht allein auf die interprofessionelle Ausbildung, sondern berücksichtigen auch Instrumente zur Evaluation interprofessioneller Zusammenarbeit. Diese Datenbanken sind sehr offen gestaltet, sodass ein freies Hochladen von Werkzeugen möglich ist. Allerdings unterliegen die integrierten Werkzeuge keiner kritischen Evaluation und keinem Qualitätsmanagement, sodass keine sichere Aussage über die Validität und Reliabilität getroffen werden kann. Somit wurden keine der auf den Datenbanken befindlichen Instrumente in dieses Review inkludiert.

2. Phase

Die PubMed-Suche ergab 24 Treffer, die Cinahl-Suche 49 Treffer (siehe Anhang 1 und Abbildung 1). Nach Sichtung der Abstracts konnten 63 Publikationen ausgeschlossen werden, da sie nicht das gesuchte Thema behandelten oder nicht den vorher vereinbarten Vorgaben entsprachen (z.B. interprofessionelle Zusammenarbeit oder qualitative Untersuchungen). Von den zehn geeigneten Reviews erfolgte in vier Reviews keine Zuordnung der Studien zu den Kirkpatrick-Stufen [5].

Die sechs verbliebenen Reviews wurden mittels der CASP-Checkliste beurteilt und auf der Suche nach validen und passenden Tools im Volltext durchgearbeitet (siehe Tabelle 2 und Abbildung 2).

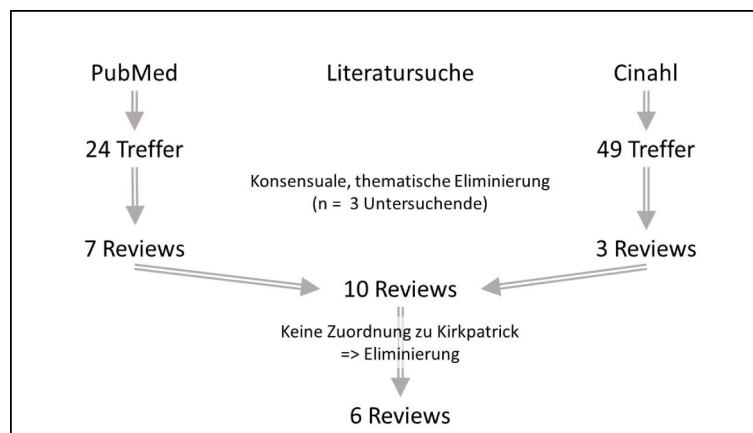


Abbildung 1: Schematische Darstellung der Ergebnisse der Literaturrecherche

Die Auswertung der Reviews zeigt unterschiedliche Ergebnisse bezüglich der Güte der inkludierten Reviews aber auch in Bezug auf empfohlene Messinstrumente. Drei Reviews weisen eine hohe Güte auf [9], [10], [14]. Allen drei Reviews liegen Standards zu Grunde (z.B. Cochrane Guidelines für Systematic Reviews oder Standards zur Beurteilung der in den Studien inkludierten Assessmenttools). Zudem bieten die drei Reviews eine klare Zuordnung der von ihnen identifizierten Evaluationsinstrumente zu den Kirkpatrick-Stufen. Es wurde ein Review mit mittlerer Güte [6] und zwei Reviews mit niedriger Güte [2], [20] in die vorliegende Übersichtsarbeit inkludiert. Bei allen drei Reviews fand in erster Linie eine explorative Suche statt oder der Literaturreview war Bestandteil einer Multimethoden-Untersuchung. Die in diesen Reviews inkludierten Studien wurden nicht kritisch gewürdigt, was die Aussagekraft der Reviews deutlich schmälert. Des Weiteren wurden die Assessments nur teilweise oder gar nicht bezüglich ihrer psychometrischen Eigenschaften hin beurteilt. Der meist genannte Wert in diesem Zusammenhang ist Cronbach's Alpha. Bezüglich weiterer psychometrischer Kennzahlen wurde nichts berichtet.

Einzig Hayer et al. [9] geben eine klare Empfehlung auf Basis der Testgütekriterien ab. Aus ihrer Sicht genügen die Instrumente Collaborative Healthcare Interdisciplinary Relationship Planning (CHRIP), Readiness for Interprofessional Learning Scale (RIPLS), Communication and Teamwork Skills assessment (CATS) und Teamwork Mini-Clinical Evaluation Exercise (T-MAX) den Gütekriterien. In ihrer Publikation von 2014 werden der Safety Attitudes Questionnaire (SAQ) und das Team Climate Inventory (TCI) ausdrücklich benannt, jedoch nicht explizit empfohlen. Hayer et al. [10] beziehen sich mit ihren Empfehlungen auf einen konkreten Kompetenzrahmen der Association of American Medical Colleges (AAMC). Als einziges Messgröße für Gütekriterien wurde von Hayer et al. [9], [10] für die inkludierten Studien mehrheitlich Cronbach's Alpha als Maß der Reliabilität genannt. Faktorenanalytische Untersuchungsergebnisse, zur Darstellung der Inhaltsvalidität, werden nicht benannt.

Oates und Davidson [14] kommen nach einer Festlegung eigener minimaler, testtheoretischer Gütekriterien zum Schluss, dass einzig die Interprofessional Socialisation

and Valuing Scale (ISVS) alle minimalen Gütestandards erfüllt. Fünf weitere Instrumente (Interprofessional Collaborator Assessment Rubric - ICAR, Attitudes Towards Teamwork in Training Undergoing Designed Educational Simulation - KidSIM ATTITUDES Questionnaire, Interdisciplinary Education Perception Scale - IEPS und University of the West of England Interprofessional Questionnaire - UWEIPQ) erfüllen die Minimalstandards nur teilweise, drei (RIPLS, Attitudes to Shared Learning - ASL und StudDat Questionnaire) erfüllen die Standards nicht oder nur ungenügend.

Gillan [6] bieten einen umfassenden Überblick über Evaluationsinstrumente und kategorisieren alle Items aus den gefundenen Instrumenten. Die Literatursuche wurde ausführlich beschrieben, sodass diesem Review eine mittlere Güte zugesprochen werden kann. Für die Praxis von großem Nutzen ist die klare Zuordnung der in den Evaluationstools enthaltenen Items zu den Kirkpatrick-Stufen [5]. Die Autoren bieten somit einen sehr transparenten Überblick. Es zeigt sich, dass eine Zuordnung zu den Stufen 1- 4a möglich ist. Zur Stufe 4b konnte kein Item zugeordnet werden.

Blue et al. [2] schlossen ausschließlich Reviews in ihre Arbeit ein. Es wird allerdings nur ungenau beschrieben, welche Arten von Reviews sie in ihre Übersichtsarbeit integriert haben. Eine kritische Beurteilung der integrierten Reviews fehlt ganz. Die explorative Literatursuche wurde allerdings als Zusatz zu einer „mixed-method“ Untersuchung durchgeführt, deren Inhalt sich jedoch nicht auf die vorliegende Fragestellung bezieht. Von daher wurde lediglich der Reviewanteil von Blue et al. [2] auf seine Güte hin untersucht.

Thistlethwaite et al. [20] suchten in ihrem Review in erster Linie nach Interventionsstudien, welche Instrumente zur Evaluation interprofessioneller Lehre und Zusammenarbeit enthielten. Es gilt anzumerken, dass lediglich eine Direktsuche in Zeitschriften durchgeführt wurde. Datenbanken wurden keine gesichtet. Zudem sind die Keywords zu dieser Suche nicht ersichtlich. Thistlethwaite et al. [20] empfehlen eine multimethodische Erhebung interprofessioneller Kompetenzen.

Insgesamt konnte festgestellt werden, dass eine breite Anzahl an Instrumenten entwickelt wurde, die auf unter-

Tabelle 2: Sechs Übersichtsartikel über die Qualität von Evaluationsinstrumenten zur Beurteilung interprofessioneller Lehre bewertet nach CASP (Critical Appraisal Skills Programme Checkliste)

Publikationen	Forschungsfrage	Literaturrecherche	Inkludierte Publikationen	Qualitätsüberprüfung	Ergebnisse	Qualität des Reviews
GILLAN et al. (2011): The evaluation of learner outcomes in interprofessional continuing education: A literature review and an analysis of survey instruments	Welche IPL Evaluationswerkzeuge liegen vor und wie decken sie die Stufen nach Kirkpatrick/Barr ab?	MEDLINE, CINAHL, manuelle Suche. Searchterms im Paper.	Alle Studien, soweit sie Instrumente enthielten oder Daten zu bekannten Instrumenten lieferten.	Auswertung auf Itemebene.	Studien: 202 Tools: 33 (538 items, Lvl 1: 111, Lvl 2a: 336, Lvl 2b: 7, Lvl 3: 51, Lvl 4a: 2 (TCI und Garrad 2006), Lvl 4b: 0). Empfohlene Tools: 0	Güte: Mittel. Inkludierte Studien wurden nicht kritisch gewürdigt. Keine Aussage bezüglich der Gesamtqualität des Reviews möglich.
HAVYER et al. (2013): Teamwork assessment in internal medicine: A systematic Review of validity evidence and outcomes	Welche Instrumente liegen in der Inneren Medizin vor zur Evaluation von Teamwork in Aus-, Fort- und Weiterbildung (Medizinstudierende, Ärzte aller Stufen, interprofessionelle Teams) und wie sind deren Gütekriterien?	MEDLINE, MEDLINE In-process, CINAHL, manuelle Suche. Searchterms im Paper.	Originalarbeiten v. a. Innere Medizin aber auch IPL, nur quantitative Tools.	Validity framework, Kirkpatrick, MERSQI, Cochrane Kriterien.	Studien: 179 Tools: 73 (53 mit guter Validität, 13 mit „patient outcomes“ korreliert, extra erwähnt SAQ und TCI). Empfohlene Tools: 0	Güte: Hoch. Cochrane Kriterien, allerdings Einchluss von nicht-randomisierten Studien. Inkludierte Studien wurden auf ihre Güte hin kritisch gewürdigt.
BLUE et al. (2015): Assessment and Evaluation in Interprofessional Education - Exploring the Field	Wie kann IPL zurzeit am besten überprüft und evaluiert werden? Wo sind noch Lücken zu füllen?	Interviews, For-schungsmeeting, Review, MEDLINE, ERIC, Psych-Info, CINAHL, ISI Web of Knowledge, manuelle Suche, 1960-2012. Suchwörter im Paper.	Systematische Reviews.	Nicht ersichtlich.	Reviews: 7 Tools: 12 (unbekannte Tools sollten besser vorgestellt werden). Empfohlene Tools: 0	Güte: Niedrig. Review Teil einer Multimethoden-Studie. Explorative Suche ohne Qualitätsbeurteilung.
THISTLETHWAITE et al. (2015): An exploratory review of pre-qualification interprofessional education evaluations	Welche Evaluationswerkzeuge für IPL gibt es und wie sehr fokussieren sie auf Outcomes und lassen sich nach Kirkpatrick einteilen?	Journal of Interprofessional Care, Medical Education, Focus on Health Professional Education and Nurse Education, Journal of Research in Interprofessional Education. Keine klare Suchstrategie.	Alle Studien soweit sie Instrumente enthielten.	Nicht ersichtlich.	Studien: 90 Tools: 4 (5 Studien auf outcomes ausgerichtet). Empfohlene Tools: 0	Güte: Niedrig. Nur direkte Journal-suche, keine Datenbanken. Keine Nennung von Keywords. Ergebnistabellen gesondert online. Keine kritische Würdigung inkludierter Studien.

(Fortsetzung)

Tabelle 2: Sechs Übersichtsartikel über die Qualität von Evaluationsinstrumenten zur Beurteilung interprofessioneller Lehre bewertet nach CASP (Critical Appraisal Skills Programme Checkliste)

Publikationen	Forschungsfrage	Literaturrecherche	Inkludierte Publikationen	Qualitätsüberprüfung	Ergebnisse	Qualität des Reviews
OATES und DAVIDSON (2015): A critical appraisal of instruments to measure outcomes of interprofessional education	Welche Evaluationswerkzeuge mit hohem psychometrischem Standard gibt es für IPL und kollaboratives Handeln bei Studierenden der Gesundheitsberufe?	CJHC inventory, Thammhauser et al. 2010, erneute Suche mit gleichen Kriterien.	Originalarbeiten mit psychometrischen Daten.	QuAILS und eigene Checkliste.	Tools: 140 (9 inkludierte tools, nur ISVS trifft Standards und deckt Lvl 2a+b, 3 ab). Empfohlene Tools: 0	Güte: Hoch. Systematische Vorgehensweise. Testtheoretische Mindestgütekriterien für inkludierte Instrumente.
HAYVER (2016): Addressing the Interprofessional Collaboration Competencies of the Association of American Medical Colleges: A Systematic Review of Assessment Instruments in Undergraduate Medical Education	Welche Validität und Charakteristika weisen Werkzeuge zur Evaluation von Teamwork bei Medizinstudierenden und interprofessionellem Handeln auf?	MEDLINE, MEDLINE In-process, CINAHL, PsycINFO (1979-2014). Searchteams in digitalem Anhang.	Originalarbeiten aller Disziplinen in denen Medizinstudierende in IPL involviert waren.	MERSQI, Framework, AAMC Competencies, Kirkpatrick.	Studien: 70 Tools: 64 Empfohlene Tools: 4 (CHIRP, R/PLS, CATS, T-MEX)	Güte: Hoch. Sorgfältige Überprüfung der psychometrischen Eigenschaften der Instrumente. Qualitätsbeurteilung der inkludierten Studien.

Anmerkung: ¹ Lvl = Kirkpatrick Level

Publikation	CA durchgeführt	Resultate		Qualitätsbeurteilung des Reviews		
		n tools	Empfehlungen	hoch	mittel	niedrig
HAVYER et al. (2013)	ja	73	nein	✗		
GILLAN et al. (2011)	ja	33	nein		✗	
BLUE et al. (2015)	nein	12	nein			✗
THISTLETHWAITE et al. (2015)	nein	4	nein			✗
OATES und DAVIDSON (2015)	ja	140	nein	✗		
HAYVER (2016)	ja	70	CHIRP, RIPLS, CATS, T-MEX	✗		

Abbildung 2: Übersicht über die konsensuale Beurteilung der sechs Review Artikel durch zwei Untersuchende (CA=Critical Appraisal)

schiedlichen Stufen des modifizierten Kirkpatrick-Frameworks eingesetzt werden [2]. Bei der Untersuchung der Volltexte wurde deutlich, dass für die Stufe 4b [1] vermutlich kein quantitatives Evaluationsinstrument existiert. Weiterhin bestand eine große Heterogenität hinsichtlich der Validierung und der Gütekriterien, des Evidenzgrads und der Evaluationsstufen nach dem abgeänderten Kirkpatrick Framework [1].

Trotzdem konnten die folgenden vier Instrumente eruiert werden, die den Anforderungen der Suche, d.h. Ausrichtung an einer oder mehreren Kirkpatrick Stufen und/oder interprofessionellen Kompetenzen gerecht wurden: Readiness for Interprofessional Learning Scale (RIPLS), University of the West of England Interprofessional Questionnaire (UWEIPQ), Interprofessional Socialization and Valuing Scale (ISVS) und das Team Climate Inventory (TCI). Eine Kurzdarstellung der einzelnen Instrumente findet sich in Anhang 2.

4. Diskussion

Bei der Betrachtung der bereits vorhandenen Online-Repositoryen über Evaluationswerkzeuge zur Interprofessionalität hat sich gezeigt, dass hinsichtlich der Verwendbarkeit und Übersichtlichkeit eine Fokussierung auf bestimmte Fragestellungen sowie eine Form der Qualitätssicherung unumgänglich sind. In der vorliegenden Untersuchung erwies sich die unsystematische Herangehensweise in den gesichteten Reviews und somit die fehlende kritische Beurteilung der in die Reviews inkludierten Studien als problematisch.

Zusammenfassend ist zu sagen, dass die heterogene Lage bezüglich der Güte der in dieser Arbeit inkludierten Reviews nur eingeschränkt eine Empfehlung zulässt. Aus testtheoretischer Sicht sollten die Empfehlungen von Haver et al. [9] berücksichtigt werden. Da jedoch nicht eindeutig ersichtlich ist, wie die Empfehlung zu Stande kam, gilt auch diese mit Vorsicht zu genießen, da insbesondere der Einsatz des RIPLS aufgrund seiner Gütekriterien inzwischen kritisch diskutiert wird [13], [18]. Keiner der Reviews benennt in Bezug auf die Testverfahren psychometrische Maße wie Sensitivität oder Spezifität.

Zudem ist für keines der genannten Instrumente eine Überprüfung der Konstruktvalidität beschrieben. Es wird lediglich teilweise beschrieben, dass eine Faktorenanalyse durchgeführt wurde. Zudem ist das konkrete Konstrukt, welches durch die Tools überprüft werden soll, nur inhaltlich erfassbar, jedoch nicht testtheoretisch nachgewiesen. Eine Überprüfung verschiedener Evaluationsinstrumente in Bezug auf die Übereinstimmung der Konstrukte (concurrent validity) wäre hier unterstützend. Auch fehlen Angaben zur Objektivität bzw. Inter-Rater-Übereinstimmung. Diese wären jedoch bezüglich der Testgüte von Bedeutung. Die Tatsache, dass meist nur die Reliabilität der Instrumente berichtet wird, deutet auf die noch mangelnde theoretische Orientierung an einem Rahmen bei der Entwicklung der Instrumente hin, die eine Validitätsüberprüfung erst erlauben würden.

Es gibt eine Vielzahl an Instrumenten zur Messung von IPL, die Outcomes auf unterschiedlichen Kirkpatrick-Stufen evaluieren. Man muss sich sowohl über Outcome wie auch Kirkpatrick-Ebene Gedanken machen, wenn man ein Evaluationsinstrument auswählt bzw. einsetzt. Es gibt (noch) nicht viele Instrumente, die als reliabel und valide erachtet werden können. Die Zuordnung eines Outcomes zu einem Kompetenzrahmen kann bei der Wahl eines Instruments hilfreich sein.

Zusätzlich können nicht alle Kirkpatrick-Stufen durch quantitative Einschätzungen abgebildet werden. Es erscheint notwendig, dass auch qualitative Evaluationen hinzugezogen werden.

TCI und ISVS werden in unterschiedlichen Reviews benannt. Von Seiten der Testgüte wird die ISVS als einzig den Kriterien genügendes Instrument von Oates und Davidson [14] beschrieben. Zudem kann die ISVS drei Kirkpatrick-Stufen zugeordnet werden. Des Weiteren wird die ISVS auch in allen anderen Reviews als Instrument benannt. Der RIPLS gilt zwar als eines der meist genutzten Instrumente, welches auch in deutscher Sprache vorläge, kann jedoch aus testtheoretischer Sicht nicht empfohlen werden, wie auch von Oates und Davidson [14] ebenfalls kritisch angemerkt wird. Der UWEIPQ, ein ebenfalls häufig genutztes Instrument, erfüllt allerdings nur teilweise die

Gütekriterien, liegt jedoch ebenfalls in deutscher Sprache vor.

Abschließend kann aus testtheoretischer Sicht einzig die ISVS derzeit empfohlen werden. Da hier ein hochwertiges Werkzeug bereits vorliegt, das nicht neu entwickelt werden muss, wird der ISVS von der Arbeitsgruppe gerade übersetzt, lokalisiert und in der deutschen Fassung validiert werden.

5. Schlussfolgerung

Der Übergang von Edukation zur Praxis muss, wie die Ergebnisse dieser Untersuchung zeigen, ebenfalls in den Fokus genommen werden. Dazu gehören Evaluationen auf der Teamebene, Überprüfung des Transfers und Entwicklung geeigneter Instrumente und Benchmarks für die Sicherstellung des Geschäftserfolgs bzw. der organisationsbezogenen Veränderungen (z.B. Patientensicherheit, Berufszufriedenheit, Critical Appraisal).

Ein deutscher Kompetenzrahmen zur IPL ist derzeit von einer anderen Arbeitsgruppe des GMA-Ausschusses für Interprofessionelle Ausbildung in Bearbeitung. Nach der Erstellung und Verabschiedung dieses Kompetenzrahmens soll die Passung der empfohlenen Instrumente erneut überprüft werden. Anschließend kann entschieden werden, inwiefern erneute, adaptierte Empfehlungen bzw. neue, der Komplexität entsprechende Instrumente entwickelt werden müssen.

Anmerkung

¹ Unter Gesundheitsberufe wird an dieser Stelle sowohl die Human- und Zahnmedizin wie auch die Pflege-, Therapie- und Diagnostische Berufe verstanden.

Interessenkonflikt

Die Autoren erklären, dass sie keine Interessenkonflikte im Zusammenhang mit diesem Artikel haben.

Anhänge

Verfügbar unter

<http://www.egms.de/en/journals/zma/2017-34/zma001113.shtml>

1. Anhang 1.pdf (112 KB)
Strategien, Suchwörter und Ergebnisse der Literaturrecherche
2. Anhang 2.pdf (94 KB)
Kurzdarstellung einzelner Instrumente

Literatur

1. Barr H, Freeth D, Hammick M, Koppel I, Reeves S. Evaluations of interprofessional education. London: United Kingdom Review of Health and Social Care; 2000.
2. Blue AV, Chesluk BJ, Conforti LN, Holmboe ES. Assessment and Evaluation in Interprofessional Education: Exploring the Field. *J Allied Health*. 2015;44(2):73-82.
3. Brandt BF. Update on the US national center for interprofessional practice and education. *J Interprof Care*. 2014;28(1): 5-7. DOI: 10.3109/13561820.2013.852365
4. Brewer M. Curtin University's faculty of health sciences interprofessional capability framework. Perth: Curtin University; 2011. Zugänglich unter/available from http://healthsciences.curtin.edu.au/local/docs/IP_Capability_Framework_booklet.pdf
5. Freeth D, Hammick M, Koppel I, Reeves S, Barr H. A critical review of evaluations of interprofessional education. London, UK: LTSN Health Science and Practice; 2002. S.63
6. Gillan C, Lovrics E, Halpern E, Wiljer D, Harnett N. The evaluation of learner outcomes in interprofessional continuing education: a literature review and an analysis of survey instruments. *Med Teach*. 2011;33(9):461-470. DOI: 10.3109/0142159X.2011.587915
7. Gordon F. Combined Universities Interprofessional Learning Unit: Final Report. Sheffield, UK: Sheffield Hallam University; 2006.
8. Hammick M, Freeth D, Koppel I, Reeves S, Barr H. A best evidence systematic review of interprofessional education: BEME Guide no. 9. *Med Teach*. 2007;29(8): 35-751. DOI: 10.1080/01421590701682576
9. Havyer RD, Nelson DR, Wingo MT, Comfere NI, Halvorsen AJ, McDonald FS, Reed DA. Addressing the Interprofessional Collaboration Competencies of the Association of American Medical Colleges: A Systematic Review of Assessment Instruments in Undergraduate Medical Education. *Acad Med*. 2016;91(6):865-888. DOI: 0.1097/ACM.0000000000001053
10. Havyer RD, Wingo MT, Comfere NI, Nelson DR, Halvorsen AJ, McDonald FS, Reed DA. Teamwork assessment in internal medicine: a systematic review of validity evidence and outcomes. *J Gen Intern Med*. 2014;29(6):894-910. DOI: 10.1007/s11606-013-2686-8
11. Interprofessional Education Collaborative Expert Panel. Core competencies for interprofessional collaborative practice. Washington, D.C.: Interprofessional Education Collaborative; 2011. Zugänglich unter/available from: <http://www.aacn.nche.edu/education-resources/ipereport.pdf>
12. Klapper B, Schirlo C. Themenheft Interprofessionelle Ausbildung – herausgegeben von der Robert Bosch Stiftung und der Gesellschaft für Medizinische Ausbildung. *GMS J Med Educ*. 2016;33(2):Doc38. DOI: 10.3205/zma001037
13. Mahler C, Berger S, Pollard K, Krisam J, Karstens S, Szecsenyi J, Krug K. Translation and psychometric properties of the German version of the University of the West of England Interprofessional Questionnaire (UWE-IP). *J Interprof Care*. 2016 (in press)
14. Oates M, Davidson M. A critical appraisal of instruments to measure outcomes of interprofessional education. *Med Educ*. 2015;49(4):386-398. DOI: 10.1111/medu.12681
15. Orchard C, Bainbridge L, Bassendowski S, Stevenson K, Wagner SJ, Weinberg L, Curran V, Di Loreto L, Sawatsky-Girling B. A national interprofessional competency framework. Vancouver: Canadian Interprofessional Health Collaborative (CIHC); 2010.

16. Reeves S, Boet S, Zierler B, Kitto S. Interprofessional Education and Practice Guide No. 3: Evaluating interprofessional education. *J Interprof Care*. 2015;29(4):305-312. DOI: 10.3109/13561820.2014.1003637
17. Reeves S, Goldman J, Burton A, Sawatzky-Girling B. Synthesis of systematic review evidence of interprofessional education. *J All Health*. 2010;39(Supplement 1):198-203.
18. Schmitz CC, Brandt BF. The Readiness for Interprofessional Learning Scale: To RIPLS or not to RIPLS? That is only part of the question. *J Interprof Care*. 2015;29(6):525-526 DOI: 10.3109/13561820.2015.1108719
19. Singh J. Critical appraisal skills programme. *J Pharmacol Pharmacother*. 2013;4(1):76. DOI: 10.4103/0976-500X.107697
20. Thistlethwaite J, Kumar K, Moran M, Saunders R, Carr S. An exploratory review of pre-qualification interprofessional education evaluations. *J Interprof Care*. 2015;29(4):292-297. DOI: 10.3109/13561820.2014.985292

Bitte zitieren als

Ehlers JP, Kaap-Fröhlich S, Mahler C, Scherer T, Huber M. Analysis of Six Reviews on the Quality of Instruments for the Evaluation of Interprofessional Education in German-Speaking Countries. *GMS J Med Educ*. 2017;34(3):Doc36.
DOI: 10.3205/zma001113, URN: urn:nbn:de:0183-zma0011137

Artikel online frei zugänglich unter

<http://www.egms.de/en/journals/zma/2017-34/zma001113.shtml>

Eingereicht: 17.06.2016

Überarbeitet: 16.01.2017

Angenommen: 04.05.2017

Veröffentlicht: 15.08.2017

Copyright

©2017 Ehlers et al. Dieser Artikel ist ein Open-Access-Artikel und steht unter den Lizenzbedingungen der Creative Commons Attribution 4.0 License (Namensnennung). Lizenz-Angaben siehe <http://creativecommons.org/licenses/by/4.0/>.

Korrespondenzadresse:

Prof. Dr. Jan P. Ehlers, M.A.
Universität Witten/Herdecke, Fakultät für Gesundheit,
Didaktik und Bildungsforschung im Gesundheitswesen,
Alfred-Herrhausen-Str. 50, 58448 Witten, Deutschland,
Tel.: +49 (0)2302/92-7786
jan.ehlers@uni-wh.de