JOURNAL OF INTERPROFESSIONAL CARE
https://doi.org/10.1080/13561820.2023.2287024

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

German translation and validation of the Interprofessional Facilitation Scale

Matthew J. Kerry[a], Adeline Paignon[b], Joanne Wiesner Conti[b], Michael Sy[a], and Marion Huber[a]

[a]Institute of Health Sciences, Zurich University of Applied Sciences (ZHAW), Winterthur, Switzerland; [b]Geneva School of Health Sciences and Centre for Interprofessional Simulation, University of Applied Sciences and Arts of Western Switzerland HES-SO, Geneva, Switzerland

ABSTRACT

We identified evidence from item response theory (IRT) to examine a German translation of the Interprofessional Facilitation Scale (IPFS). The IPFS was administered to $n = 130$ mixed-health profession participants in a post-interprofessional education practicum questionnaire. We used IRT analyses to examine the following three aspects of the IPFS: (a) general factor strength, (b) subscale usability, and (c) item bias. First, findings indicate a strong, general factor underlying the IPFS that supports unidimensional interpretations. Second, findings supported IPFS overall reliability, but failed to support subscale reliabilities. Third, item bias assessment using a comparator-French sample ($n = 89$) indicated insubstantial differences across German and French samples. Taken together, we find sufficient evidence to support the IPFS-German translation's application in IPE contexts and unidimensional interpretations. Subscores are not advisable for interpretation, and future researchers should aim to further inspect potential item bias.

ARTICLE HISTORY
Received 21 August 2023
Revised 14 November 2023
Accepted 14 November 2023

KEYWORDS
Interprofessional education; Interprofessional Facilitation Scale; IPFS; item response theory; teamwork

Introduction

Although interprofessional education (IPE) research is flush with student assessments (Guitar & Connelly, 2021), comparatively fewer assessments are available for facilitators (Sargeant et al., 2010). Furthermore, one region that has seen a particular uptick in IPE programs are the so-called German-speaking "DACH" countries, comprised of Germany, Austria, and Switzerland (Kaap-Fröhlich et al., 2022). Taken together, we aimed to examine a German-translated instrument for assessing facilitators in IPE contexts.

Background

The Interprofessional Facilitation Scale (IPFS) was constructed in Canada in 2010 by Sargeant et al. (2010). It was originally developed for administration in English. Noting IPE's "constructivist" nature, the IPFS's conceptualization is rooted in theories of social learning (Wenger et al., 2002) and professional identification. Its initial validation included a three-phase study including: (a) Competency-domain identification and content generation via literature review and subject matter expert advisory, resulting in a reduction from 27 to 18 items; (b) Feasibility piloting and content validity via a 5-day interactive facilitator development workshop ($n = 34$); and (c) Psychometric testing with $n = 311$ professionals representing 15 different professions: primarily nurses (58%), pharmacists (18%), and physicians (13%). It was found to have good reliability with a two-correlated subfactor structure, labeled: 1) Encourage IP Interaction, and 2) Contextualize IPE. The unique value of the IPFS to IPE research is exhibited by its expanded translational adaptation to, for example, Japanese (Haruta et al., 2018) and French (Paignon et al., 2021). Given the rapid expansion of IP programs in German-speaking countries, we aimed to examine the IPFS's translation for validated use in German-language IP settings.

Methods

A cross-sectional design with surveys was used to collect IPFS response data. Students and faculty from six health professions completing a 3- to 4-week long, teaching-hospital based IP clinical practicum (Zurich interprofessional clinical training wards, ZIPAS) in the German-speaking part of Switzerland were invited to participate in a post-training, online survey. The online-survey included demographics and three substantive variables, of which one, the IPFS, is the focus of the current study. Surveys were administered on the last day of training, and they were timed at approximately 17 minutes for completion.

Sample

A total of $n = 130$ responses were collected from second–, third-, and fourth-year students ($n = 85$) and faculty ($n = 45$) between 2018 and 2022. The professional representation comprised nursing (39%), medicine (27%), physical therapy (20%), and mixed-allied health professions (14%), such as dieticians and ergotherapists. The sample was primarily female (77%), with a mean age of approximately 30 years. Additionally, item-bias assessment was conducted using raw data ($n = 89$) from a previously validated French-translation of the IPFS (Paignon

CONTACT Matthew J. Kerry kerr@zhaw.ch Health Department, Zürich University of Applied Sciences (ZHAW), Katharina-Sulzer-Platz 9, Winterthur 8400, Switzerland

et al., 2021). The French sample comprised 89 facilitators from the University of Applied Sciences and at the Faculty of Medicine of Geneva. The evaluation took place during the debriefing of interprofessional simulations involving undergraduate students from six different curricula (nursing, medicine, nutrition, physiotherapy, medical radiology technology, and midwifery).

Measures

The IPFS comprises 18 items rated on a 4-point Likert-type scale with unique descriptors anchoring the poles for each item from 1 (*poor*) to 4 (*excellent*). A forward-back translation procedure was followed with native-speaking authors of each language (English and German). A forward-translation was made by a native-German speaker (author: MH) with fluent English proficiency. A back-translation was made by a native-English speaker (author: MK) with advanced German proficiency (level C2). The original English text was compared for content compatibility and substantive meaning. One item required iterative discussion to identify the appropriate noun: Item 4's use of "environment" was seen as too general for the German training setting and was adapted to "Context." An analogous procedure was conducted for the French translation by authors AP and JC, respectively, with consensus reach without need for further iterative discussions. The full-length translated questionnaire is provided in the supplementary appendix.

Analyses

Data cleaning and classical analyses were conducted in software IBM SPSS v29. Item Response Theory (IRT) analyses were conducted in software IRTPRO v6. Exploratory and confirmatory factor analyses were used to determine the dimensionality of the IPFS. Subscale reliabilities were computed with a bifactor indices calculator based in MS Excel (Dueber, 2017). Item bias was conducted with differential item functioning where equality constraints are imposed on items (intercepts and slopes) and compared to unconstrained versions for model-data fit. McFadden's pseudo R^2 was computed to assess the magnitude of differential item functioning (DIF) with the following rubric. We applied a conventional criteria guideline to evaluate R^2 as follows: < .13= negligible bias, .13–.26 = moderate bias, and > .26 = large bias (Zumbo & Thomas, 1996).

Ethics statement

Ethics consideration was deemed exempt status according to the bylaws of the Zürich Cantonal Ethics Committee due to the full-anonymization of our dataset and its nonsensitive population.

Data availability

All anonymized data are available to interested researchers upon written request.

Results

General factor strength

As shown in Figure 1, the first eigenvalues from EFA indicated the presence of a strong general factor. The first and second eigenvalue ratio was 15.25/.64 = 23.83, suggesting negligible multidimensionality (Embretson & Reise, 2000). This was supported by a high estimated common variance (ECV=.96), indicating that approximately 96% of all IPFS variance was explained by the general factor. Finally, comparing factor loadings across unidimensional and multidimensional models using confirmatory factor analysis (CFA) revealed a relative-bias from fitting multidimensional data to a unidimensional model. The average-relative parameter bias (.01) value indicates that the impact of ignoring the multidimensionality by using unidimensional IPFS scores was negligible (Muthén et al., 1987). Finally, because the original IPFS was mentioned as being viable for both other- and facilitator self-report, we repeated the parallel analysis separately for students ($n = 85$) and facilitators ($n = 45$). Findings were consistent in indicating a strong general factor according to eigenvalue ratios of 6.56 for student ratings and 34.74 for facilitator self-ratings. Overall reliability according to McDonald (2013) omega coefficient (ω) was .99, .94 for students, and .99 for facilitators.

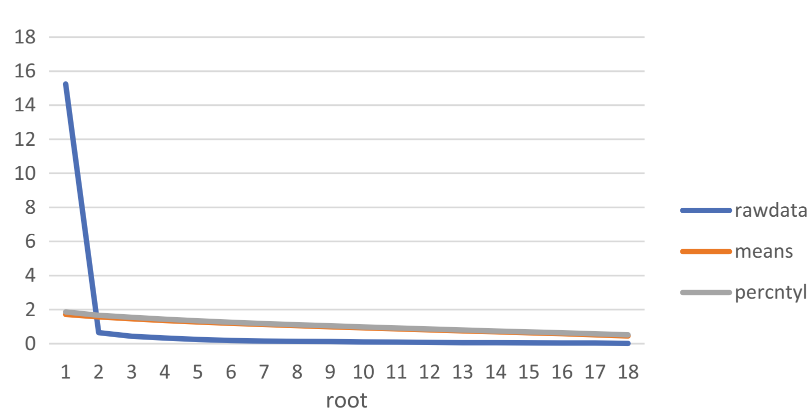

Figure 1. IPFS parallel analysis of eigenvalues for general factor strength.

Subscale reliability

The original IPFS was reported as having an "unexpected" second factor (Sargeant et al., 2010). The two factors were labeled: (a) "Encourage IP Interaction" (15 items), and (b) "Contextualize IPE" (3 items). Usability of these factors as subscales was, therefore, examined in its translation. Omega coefficients (ω) are analogous to Cronbach's alpha (α) when multidimensionality's impact is unknown (Cho, 2016). We computed ω reliabilities and compared it to hierarchical-omegas (ω_H), which indexes "pure" subscale reliability after removing the general factor. For clarity, subscale-specific reliabilities are labeled ω(Con) and ω(Enc) to denote original subscale labels "Context" and "Encourage," respectively.

First, IRT estimates were used to compute subscale omegas as ω(Con) = .99 and ω(Enc) = .99. Second, hierarchical-omegas were computed for each subscale as ω_H(Con) = .07 and ω_H(Enc) = .01. Third, dividing the subscales' ω_H by their respective ω coefficients illustrates the percentage of reliable variance in subscales that are independent of the general IPFS. Calculating for Context subscale (.07/.99 = .07) and Encourage subscale (.01/.99 = .01) indicates that 7% and 1% of reliable variance in the subscales is independent of the general IPFS. This should be interpreted as insufficient reliability for using IPFS subscores in IPE (Nunnally & Bernstein, 1994). These analyses were replicated for subsets of students and facilitators with no substantive change in findings (unsuitability of subscales).

Item bias

A two-step procedure was implemented to assess potential item bias in the IPFS. First, traditional statistical criteria were used to detect potential bias. Chi-square (X^2) values that were significant for each item were further examined for significant differences on their slope parameter. The IRT slope parameters are analogous to item loadings in factor analysis, and significant differences suggest the potential for measurement differences across groups (students and faculty). Any items displaying significant differences in slopes were further evaluated for their magnitude using calculations of McFadden's pseudo R^2 statistic. Two items (6 & 7) were necessarily omitted from this procedure due to their omission in the original administration of the French-version of the IPFS.

Results displayed in Table 1 indicate two items detected for potential bias: Item 9 – "Acknowledged and respected others' experiences and perceptions," and Item 11 – "Asked participants to share their professional opinions . . . relative to patient care and collaborative practice." McFadden's pseudo R^2 was calculated to evaluate bias magnitude to determine the appropriateness of retaining the item for use in a German-IPE context. As shown in Table 1's right column, R^2*s* were below the .13 threshold, suggesting negligible bias and appropriateness of the items to remain in the German-IPFS.

Finally, a more easily intuited index of item bias that is interpretable according to Cohen's *d* recommendations for small (.2), medium (.5), and large (.8) effect sizes was computed. Computations indicated that only one item even approached a medium effect with Item18 "Discussed issues related to hidden power structures . . . " = .47 (favoring German sample). Taking all items together, however, the expected test score standardized difference was only .06, representing a very small effect of item bias on total IPFS scores. Importantly, a sample-based stat indicating Item17 was low in our German sample prompted content review. This review revealed the ambiguous content formulation of Item17 – "Used effective skills to clarify and resolve misunderstanding and conflict, *if applicable*." This raises the question of whether low scores endorse lack of skills or inapplicability of the item. This item may be simplified by eliminating the loaded "if applicable" condition in future IPFS uses. The Cohen's *d*-interpretable item (largest) and overall IPFS bias are displayed in Figure 2.

Discussion

In this study, three analyses were conducted to examine the German translation and validation of the IPFS. IRT was used

Table 1. IPFS item bias detection and magnitude evaluation.

	Detection						Magnitude
Item #	Total X^2	*d.f.*	*p*	Slope X^2	*d.f.*	*p*	*McFadden's pseudo R^2*
1	11.4	4	.02	.1	1	.74	
2	0.9	4	.92	4.8	1	.62	
3	5.7	4	.23	.4	1	.03	
4	2.8	4	.59	.3	1	.71	
5	2.9	4	.57	.1	1	.43	
6	Omitted						
7	Omitted						
8	6.9	4	.14	1.8	1	.99	
9	7.1	4	.13	2.6	1	.04	<.01
10	7.5	4	.11	0	1	.94	
11	7.5	4	.11	1.1	1	.04	<.01
12	1	4	.92	.2	1	.46	
13	7.2	4	.13	5.2	1	.17	
14	7.6	4	.11	.2	1	.09	
15	2.8	4	.59	.2	1	.14	
16	10	4	.04	1.7	1	.38	
17	19.8	4	.00	1.9	1	.22	
18	13.1	4	.01	.8	1	.71	

N = 219. $X^2 = -2$ log likelihood. d.f. = degrees of freedom. McFadden's pseudo R^2 computed as ratio of restricted model/free model.

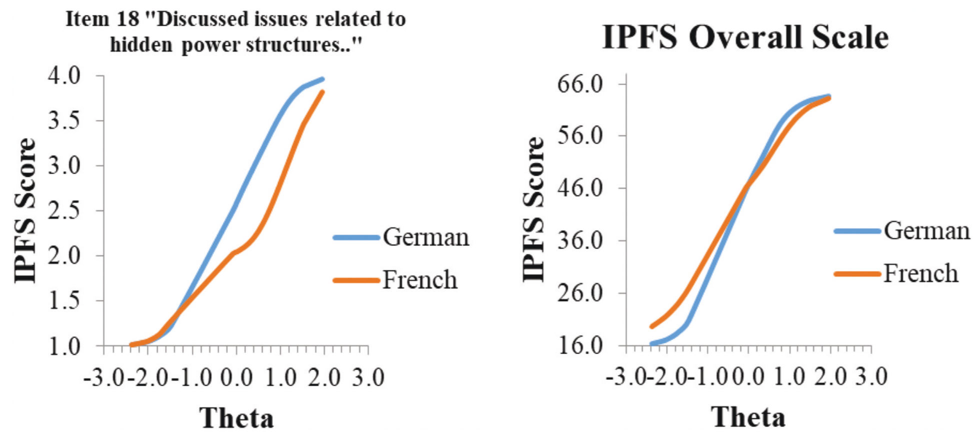

Figure 2. Largest item-bias (approaching medium effect) and overall IPFS functioning across German and French samples.

to examine the IPFS's: (a) General factor strength, (b) Subscale viability, and (c) Item bias. First, factor-analytic findings indicated the presence of a strong general factor, supporting the IPFS scale's unidimensional interpretation. Second, subscore omega coefficients indicated insufficiently weak reliabilities, thereby failing to support the use of IPFS subscales in IPE settings. Third, item bias assessment indicated one potentially biased item across French and German language samples. Magnitude evaluations, however, suggested negligible impact and supported the appropriateness for retaining the IPFS items in the German-translated instrument. Taken together, these results suggest that researchers interested in using a general IPFS score for prediction or structural equation modeling may do so simply by using an IPFS unidimensional score.

Several limitations to the current study bear mentioning. Firstly, our sample comprised a mix of students and facilitators, whereas the French-dataset comprised only facilitators. Future researchers may examine student versus professional samples in future research as an observer-rated instrument. Second, items six and seven were omitted after content analysis from the French-dataset due to inapplicability in a simulation setting. The items may still be retained for future simulation contexts, but they were deemed incompatible to the clinical simulation debriefings in the French context. For practicality, however, we advise that the specific examples given in Item 6 (g, icebreaker games, case studies, group discussions) may be adapted to align with the specific pedagogy or learning activities undertaken in a given study. Further DIF-inspection of these items, in particular, are therefore encouraged in future research.

Because the advancement of the IPE field depends critically on its quantitative methodologies, adapted measures should undergo comprehensive evaluation for appropriate use *via* advanced psychometrics (Kerry & Huber, 2018). The current, three-step approach aims for a heuristical illustration toward facilitating IP measurement examination. IP researchers should carefully consider: (a) General factor strength, (b) Subscale reliability, and (c) Item bias. By following the three-step, pseudo-tutorial approach presented here, researchers may be able to expedite evidence for linguistic validations of IPE instruments and their expanded applications.

Conclusion

The German-translated IPFS ($IPFS_{Ger}$) was found to be psychometrically valid for future use in German-speaking IP training settings. Only the overall instrument's usage is recommendable (not subscales). Item modifications such as item 6's customization and item 17's simplification may improve measurement properties. Currently presented evidence limits our recommendations for usage in cross-sectional or comparative designs (between-groups difference detection), whereas further testing of the $IPFS_{Ger}$'s responsiveness over time is required for future repeated-measures designs.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

The author(s) reported that there is no funding associated with the work featured in this article.

Notes on contributors

Matthew Kerry is a Scientific Research Associate at the Zürich University of Applied SciencesDepartment of Health. He applies substantive methods to a broad array of health-sciences education and health services research.

Adeline Paignon is a Scientific Assistant at the Geneva School of Health Sciences and University of Applied Sciences and Arts of Western Switzerland conducting health services research.

Joanne Wiesner Conti is a Scientific Assistant at the University of Geneva`s Standardized Patient Program.

Michael Sy is a Scientific Research Associate at the Zürich University of Applied Sciences Department of Health.

Marion Huber is Head of the Interprofessional Teaching and Practice Unit at the Zürich Universiy of Applied Sciences Department of Health.

ORCID

Matthew J. Kerry http://orcid.org/0000-0002-1339-899X
Adeline Paignon http://orcid.org/0000-0003-0634-8334

References

Cho, E. (2016). Making reliability reliable: A systematic approach to reliability coefficients. *Organizational Research Methods*, *19*(4), 651–682. https://doi.org/10.1177/1094428116656239

Dueber, D. M. (2017). Bifactor Indices Calculator: A Microsoft Excel-Based Tool to Calculate Various Indices Relevant to Bifactor CFA Models. https://doi.org/10.13023/edp.tool.01

Embretson, S., & Reise, S. (2000). *Item response theory for psychologists*. Psychology Press. https://doi.org/10.4324/9781410605269

Guitar, N. A., & Connelly, D. M. (2021). A systematic review of the outcome measures used to evaluate interprofessional learning by health care professional students during clinical experiences. *Evaluation & the Health Professions*, *44*(3), 293–311. https://doi.org/10.1177/0163278720978814

Haruta, J., Breugelmans, R., & Nishigori, H. (2018). Translation and cultural adaptation of the Japanese version of the interprofessional facilitation scale. *Journal of Interprofessional Care*, *32*(3), 321–328. https://doi.org/10.1080/13561820.2017.1398720

Kaap-Fröhlich, S., Ulrich, G., Wershofen, B., Ahles, J., Behrend, R., Handgraaf, M., Herinek, D., Mitzkat, A., Oberhauser, H., Scherer, T., Schlicker, A., Straub, C., Eichler, R. W., Wasselborg, B., Witti, M., Huber, M., & Bode, S. (2022). Position paper of the GMA Committee interprofessional education in the Health professions – Current status and outlook. *GMS Journal for Medical Education*, *39*(2). https://doi.org/10.3205/zma001538

Kerry, M. J., & Huber, M. (2018). Quantitative methods in interprofessional education research: Some critical reflections and ideas to improving rigor. *Journal of Interprofessional Care*, *32*(3), 254–256. https://doi.org/10.1080/13561820.2018.1426267

McDonald, R. P. (2013). *Test theory: A unified treatment*. psychology press. https://psycnet.apa.org/record/1999-02770-000

Muthén, B., Kaplan, D., & Hollis, M. (1987). On structural equation modeling with data that are not missing completely at random. *Psychometrika*, *52*(3), 431–462. https://doi.org/10.1007/BF02294365

Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory*. McGraw-Hill.

Paignon, A., Wiesner Conti, J., Cerutti, B., & Fassier, T. (2021). French translation and validation of the interprofessional facilitation scale for simulation. *Journal of Interprofessional Care*, *35*(5), 803–807. https://doi.org/10.1080/13561820.2021.1879750

Sargeant, J., Hill, T., & Breau, L. (2010). Development and testing of a scale to assess interprofessional education (IPE) facilitation skills. *Journal of Continuing Education in the Health Professions*, *30*(2), 126–131. https://doi.org/10.1002/chp.20069

Wenger, E., McDermott, R. A., & Snyder, W. (2002). *Cultivating communities of practice: A guide to managing knowledge*. Harvard Business School Press.

Zumbo, B. D., & Thomas, D. R. (1996). A measure of DIF effect size using logistic regression procedures. Paper presented at the *National Board of Medical Examiners*, http://www.sciepub.com/reference/304227