



Psychometric Testing of the German Versions of Three Implementation Outcome Measures

Christina Kien¹ · Ursula Griebler¹ · Marie-Therese Schultes² · Kylie J. Thaler³ · Tanja Stamm⁴

Received: 22 April 2021 / Accepted: 30 July 2021 / Published online: 23 August 2021
© The Author(s) 2021

Abstract

Internationally comparable assessments of implementation outcomes are a prerequisite for the accurate and useful exchange of implementation science results. We translated the English versions of the acceptability of intervention measure (AIM), the intervention appropriateness measure (IAM), and the feasibility of intervention measure (FIM) into German following recent guidelines. We tested the structural validity, the internal consistency (Cronbach's α), and the construct validity (known-groups validity) in a hypothetical situation study of secondary school teachers ($n_1 = 142$) and the structural validity and the internal consistency in a pilot field study of primary school teachers participating in a school health promotion project ($n_2 = 61$). The hypothesized three-factor solution (acceptability, appropriateness, and feasibility) revealed the best model fit in the hypothetical situation study (CFI = 0.97; RMSEA = 0.08). The second study tended to support these results (CFI = 0.99; RMSEA = 0.05) only when the error terms of two items from different constructs were allowed to correlate with each other ("seems like a good match" [IAM] and "seems possible" [FIM]). Furthermore, the intercorrelations between the three factors were higher in the pilot field study than in the hypothetical vignette study (0.71–0.86 vs. 0.06–0.35). The instruments' internal consistency proved to be high in both studies (Cronbach's α : 0.91–0.97). We gathered positive evidence for the known-groups validity, as participants who received different stimuli before answering the instruments rated them statistically significantly different. Our study showed that the German versions of the instruments could successfully be applied in health promotion research. We suggest using the four items per instrument rather than an abbreviated version and discuss recommendations for the instruments' application.

Keywords Psychometric properties · Test-theoretical criteria · Instrument · Questionnaires · Implementation outcomes

These data were partly presented as a poster at the 5th biennial Global Implementation Conference; Glasgow, Scotland; September 2019. Full results were presented as a conference presentation at the European Implementation Event 2021.

✉ Christina Kien
christina.kien@donau-uni.ac.at

- ¹ Department for Evidence-based Medicine and Evaluation, Danube-University Krems, Dr.-Karl-Dorrek Strasse 30, 3500 Krems an der Donau, Austria
- ² Institute for Implementation Science in Health Care, University of Zurich, Universitaetstrasse 84, 8006 Zurich, Switzerland
- ³ 1st Medical Department, Hanusch Krankenhaus, Heinrich-Collin-Straße 30, 1140 Vienna, Austria
- ⁴ Section for Outcomes Research, Center for Medical Statistics, Informatics, and Intelligent Systems, Medical University of Vienna, Spitalgasse 23, 1090 Vienna, Austria

Abbreviations

AIM	Acceptability of intervention measure
CFA	Confirmatory factor analysis: is a multivariate statistical procedure to investigate whether the hypothesized constructs can be represented by the instruments used. The researchers can specify the number of essential constructs and which measured item is related to which construct
CFI	Comparative fit index: analyses the model fit by investigating the difference between the hypothesized model and the given data. CFI values can range from 0 to 1; larger values indicate a better fit
EFA	Exploratory factor analysis: is a multivariate statistical procedure to reduce the complexity of the data and identify the factors and their relationships with the individual items
FIM	Feasibility of intervention measure

IAM	Intervention appropriateness measure
IOF	Implementation outcomes framework: framework describing implementation outcomes developed by Proctor et al. (2011)
KMO	Kaiser–Meyer–Olkin: this criterion indicates how suited the data are for factor analysis. KMO values can range from 0 to 1; higher values indicate a higher suitability
M	Mean
Max	Maximum
Md	Median
Min	Minimum
<i>n</i>	Number
RMSEA	Root mean square error of approximation: this criterion analyses the discrepancy between the hypothesized model and the population covariance matrix. RMSEA ranges from 0 to 1; smaller values indicate a better model fit
TRAPD	Translation, review, adjudication, pretesting, and documentation: guidelines for translating questionnaires and instruments
%	Percentage

Background

Implementing an intervention or a new evidence-based practice in a specific setting is a process that requires expertise as well as accurate and precise instruments to measure the outcomes (Greenhalgh et al., 2004; Nilsen, 2015). In addition to assessing the actual effectiveness of an intervention, researchers need to take the direct effects of the implementation endeavor (i.e., implementation outcomes) into consideration to advance the understanding of implementation processes or to compare implementation strategies. Implementation science is growing globally, and the need to compare research results across borders is important. To generate internationally comparable research results and facilitate multi-country studies and cross-border initiatives (Bührmann et al., 2020), translating reliable and valid instruments into languages other than English is essential (Schweizer, 2010).

Investigated Framework

The Implementation Outcomes Framework (IOF, a list of explained abbreviations can be found at the end of the article) covers eight conceptually different implementation outcomes (Proctor et al., 2011): acceptability, appropriateness, feasibility, cost, adoption, penetration, and sustainability. The first three outcomes are commonly used for the evaluation of implementation outcomes (Proctor et al., 2011; Weiner et al., 2017) in different fields, although this

framework was originally developed for the mental health sector.

“*Acceptability* is the perception among implementation stakeholders that a given treatment, service, practice, or innovation is agreeable, palatable, or satisfactory” (Proctor et al., 2011, p. 67). “*Appropriateness* is the perceived fit, relevance, or compatibility of the innovation or evidence based practice for a given practice setting, provider, or consumer; and/or perceived fit of the innovation to address a particular issue or problem” (Proctor et al., 2011, p. 69). Although these two constructs show semantic similarities, they can be distinguished on a theoretical level. For example, a new treatment can be considered by a provider as a good fit to address patients’ diseases (high appropriateness); nevertheless, the provider may be reluctant to use it, as he or she disapproves a certain feature of the intervention (low acceptability). “*Feasibility* is defined as the extent to which a new treatment, or an innovation, can be successfully used or carried out within a given agency or setting” (Proctor et al., 2011, p. 69). Even though appropriateness and feasibility can correlate, they address differing aspects on a conceptual level. For example, a program may fit to the culture and mission of a certain setting (high appropriateness), but may take up extraordinary personal costs and other resources (low feasibility).

Often existing instruments lack a sound development process or provide only limited information on psychometric criteria, other than internal consistency (Lewis et al., 2015; Mettert et al., 2020). However, a research group (Weiner et al., 2017) undertook a thorough development process and psychometrically tested self-reported instruments that aim to measure acceptability, appropriateness, and feasibility.

Measures

The three instruments, the Acceptability of Intervention Measure (AIM), the Intervention Appropriateness Measure (IAM) and the Feasibility of Intervention Measure (FIM), each consist of four items with scale values ranging from 1—*completely disagree* to 5—*completely agree*. The source English instruments are publicly available and free of charge (Weiner et al., 2017). Due to their generic nature, the measures can be applied when investigating psychosocial, technical, health promotion or other interventions in the hospital and health care setting, in the education system, or at the workplace. They can also be used within a wide range of stakeholder groups such as administrators, direct service providers, or family members.

The development and testing process of the instrument followed several steps described in three different studies included in one article. First, based on the definitions of the investigated constructs and on already existing instruments, the research team developed 31 items. In an online

study, 103 members of the relevant stakeholder group (i.e., implementation scientists and mental health practitioners) allocated the unsorted items to the constructs and judged their confidence in their allocation procedure (Weiner et al., 2017). The results of a factor analysis led to a refinement and shortening of the scales (overall 15 items). In a second study based on the answers of 326 members of the American Mental Health Counselors Association, a confirmatory factor analysis revealed the hypothesized three-factor structure (Lewis et al., 2018; Weiner et al., 2017) showing a good model fit (Comparative Fit Index, CFI=0.96; Root Mean Square Error of Approximation, RMSEA=0.079). The Cronbach α values for the revised 4-item scales were 0.85 for acceptability, 0.91 for appropriateness, and 0.89 for feasibility. A third study investigated test–retest reliability after seven weeks (r ranged from 0.73 to 0.88) and verified the instruments' sensitivity to change in both directions. Currently, comprehensive results on the psychometrical performance of the instruments based on other samples are lacking, but a range of studies applying these instruments have recently been published (Adrian et al., 2020; Damush et al., 2021; Swindle et al., 2021; Taboada et al., 2021; Wasil et al., 2021).

Aims of the Present Investigation

To the best of our knowledge, there is currently no generic, reliable, and valid instrument available in the German language that can be used to assess the implementation outcomes acceptability, appropriateness, and feasibility (Kien et al., 2018). We chose to translate the English versions of the measures AIM, IAM, and FIM (Weiner et al., 2017) into German as they are promising, reliable, valid and generic in nature and can therefore be applied in varying contexts (Finch et al., 2012). In addition, these instruments enable the assessment of the three implementation outcomes in a similar way, therefore enabling a direct comparison of the three different constructs.

These instruments have only been applied and extensively tested in the mental health sector so far and we aimed to extend their usage to health promotion interventions in the school setting. In general, the aim of the present investigation was to assess the structural validity, reliability, and known-groups validity of the German language instruments. Structural validity is the extent to which a scale shows the internal structure of its items as theoretically hypothesized (McDowell, 2006). Internal consistency refers to the extent the items of a scale are correlated (McDowell, 2006). Known-groups validity aims to conclude whether groups with distinct features can be differentiated as assumed by their answers to a measure (Lewis et al., 2018; Souza et al., 2017).

Therefore, our approach followed three distinct steps: (i) We translated the instruments into German. (ii) We assessed the instruments' structural validity, internal consistency, and known-groups validity based on the answers of study participants who read hypothetical situations and filled in the instruments from the perspective of the protagonist of the hypothetical situation (i.e., Study 1). (iii) We assessed the instruments' structural validity and internal consistency within the evaluation of a health promotion project in a real-world setting. We expected less variation in the participants' answers as they have already decided on adopting the program (i.e., Study 2).

Methods

Measures—Translation Process

Different methods for the cross-cultural adaptation of instruments exist (Epstein et al., 2015). Within the health sciences, the forward- and backward-translation design is the most common technique used for cross-cultural adaptations of self-reported instruments. Therefore, we followed the guidelines depicted by Guillemin and colleagues (Guillemin et al., 1993). This process also resembled the Translation, Review, Adjudication, Pretesting, and Documentation (TRAPD)-team translation model (Cross-Cultural Survey Guidelines, 2016):

1. Translation: Two German native speakers (qualified implementation scientist, evaluation researcher) independently translated the English source instruments including the description of the scale values and the introduction explaining the usage of the instruments to researchers and practitioners into German.
2. Back-translation: Two English-native speakers fluent in German (epidemiologist, English language teacher) each back-translated one of the two different German versions into English.
3. Committee review: The multidisciplinary committee of the above mentioned persons and the principal investigator, all native in German or English and fluent in the second language ($n=5$) reviewed and discussed the results of the translation and back translations in an online meeting. Disagreements were resettled via consensus or if not reached, we empirically tested different options (see next step).
4. Cognitive interviews: We conducted seven cognitive interviews with representatives of the intended sample population of primary and secondary school teachers (5 women, 2 primary school teachers, 3 over 50 years). Specifically, we elicited the teachers' associations to different translated options for one item to choose the one

that best fit the English source item. According to the results of the cognitive pre-tests the principal investigator adapted the instruments.

5. The Committee discussed and approved the adaptations. No further changes were considered necessary.

The translated German versions of the instruments (12 items) are available in Additional File 1. The scale names were translated in alignment with the results of the systematic translation and cross-validation (Gutt et al., 2018) of the implementation outcomes defined by Proctor et al. (2011).

Hypothetical Situations Study

Procedure, Material, and Participants

Following the study design developed by Weiner and colleagues we created eight different hypothetical situations (i.e., vignettes). In general, they presented the discussion of two fictitious secondary school teachers about the implementation of a physical activity health promotion program. Within these vignettes, the antecedents of each implementation outcome were either pronounced as very high or very low. This manipulated maximum variation within the vignettes was intended to lead to maximum variability in assessing the implementation outcomes *acceptability*, *appropriateness*, and *feasibility*. We manipulated the following antecedents: an individual's openness to new experiences and the possibility to try out and undo the implementation of an intervention (i.e., trialability) should affect the implementation outcome *acceptability* (Weiner et al., 2017). Furthermore, social norms within a professional group and the target groups' needs for an intervention should influence the perception of *appropriateness*. In addition, the available resources for the implementation endeavor and the perceived complexity of the intervention should affect the outcome *feasibility*.

To make sure that the different vignettes and the perceived variation of the antecedents were plausible, we tested them within the seven cognitive interviews. Afterward, we revised the vignettes to improve plausibility (e.g., age of pupils participating in the program, program costs) and to increase variation between antecedents (e.g., trialability and resources). The variation of the antecedents in the different vignettes is entailed in Table 1.

Secondary school teachers were invited to rate the measures (AIM, IAM, FIM) from the perspective of the illustrated character in the vignette without being influenced by their own opinion about the hypothetical program in the vignette. The eight different vignettes were randomly assigned. To avoid sequence effects, we also randomly ordered the items of the paper–pencil instrument (15 different versions). Additionally, we collected demographic data (e.g., gender, work experience, place of work—type of school, experience with implementation of health promotion projects) for the sample description.

We used GPower (Faul et al., 2007) to calculate the necessary sample size for detecting a difference in the implementation outcomes ratings defined as a medium effect size ($f=0.25$) in main and interaction effects (considering 8 different vignette types) with a power of 80% at a statistical significance level of $\alpha \leq 0.05$. We arrived at a calculated sample size of 128 participants. This sample size also conforms to the recommendation that the sample size for the confirmatory factor analysis (CFA) should be at least five times the number of items (Bentler & Chou, 1987), resulting in a recommended sample size of $n=60$.

Overall, a convenience sample of 147 secondary school teachers participated in the vignette study during April and July 2019. We originally intended to invite teachers to participate in the study who were attending further education courses at the University College of Teacher Education in Vienna (Austria), as we strived for a diverse sample population. However, in the last months of the school year, we could only manage to invite the participants of three further education courses ($n=76 / 51.7\%$). Therefore, we also recruited teachers from two higher secondary schools in Vienna ($n=65 / 48.3\%$) to fulfill our target sample size. The participation rates at the further education courses ranged from 73.3% to 93.3% and in the schools from 11.9% to 68.7%. Women prevailed in the sample (78.8%); reflecting a similar gender distribution as in the Austrian education system (Statistik Austria, 2019). The participants reported to have worked on average for 10.3 years, mainly at higher secondary schools (91.8%). The remaining participants worked at lower secondary schools (5%) or vocational schools (2.8%). Approximately a fifth (18.4%) and more than half of the participants (59.9%) indicated that they had implemented a health promotion project or another project in their school during the last five years. Five teachers did

Table 1 Overview of the characteristics of the different vignettes

	V-1	V-2	V-3	V-4	V-5	V-6	V-7	V-8
Acceptability (openness, trialability)	High	High	High	High	Low	Low	Low	Low
Appropriateness (social norms, target groups' needs)	High	High	Low	Low	High	Low	Low	High
Feasibility (resources, complexity of the intervention)	High	Low	Low	High	High	High	Low	Low

not answer all items of the AIM, IAM and FIM measure. Since a CFA cannot be performed with missing data (Weiber & Mühlhaus, 2014), we omitted these five cases from the analysis, resulting in a final sample size of 142.

The participating teachers were informed about the aim, the duration, and the demands of the study a priori in a leaflet. By filling in the paper–pencil instruments, teachers provided informed consent. The study protocol and the instruments were submitted to the ethics committee of the Medical University of Vienna that waived responsibility, since survey studies with experts/professionals do not need to be submitted to an ethics committee in Austria.

Data Analyses

First, we calculated descriptive statistics for the description of the study sample and to become familiar with the data. Second, to assess unidimensionality for each instrument separately, we performed an exploratory factor analysis (EFA) for each instrument (principal component analysis with promax rotation). We applied the following criteria for good model fit: communalities ≥ 0.5 ; Kaiser–Kriterium eigenvalues > 1 (Weiber & Mühlhaus, 2014); Kaiser–Meyer–Olkin (KMO) Measure of Sampling Adequacy criterion indicating correlation of variables ≥ 0.6 (Kaiser, 1974) and a statistical significant Bartlett’s Test of Sphericity (Dziuban & Shirkey, 1974).

Third, we calculated a three-factor CFA using maximum likelihood estimation. We applied the following guidelines for defining good model fit: CFI ≥ 0.95 (Schreiber et al., 2006), and RMSEA ≤ 0.08 (Browne & Cudeck, 1993). For comparison reasons we calculated two alternative models: a two-factor model with appropriateness and acceptability items loading on one factor, as it is hypothesized that these factors would show the highest correlations (Weiner et al., 2017), and a general factor model including all items.

Fourth, based on the results of the CFA, for assessing the internal consistency of the items of the three different instruments, we calculated Cronbach’s α , considering α values ≥ 0.8 as good (Rossiter, 2002).

Fifth, for assessing the known-groups validity we performed three 3×2 ANOVAs with type three sum of square methodology to assess the assumed differences in the scales’ ratings due to the assignment to the diverse vignettes. If interaction effects were not statistically significant, we omitted them from the model and recalculated it. For further analyses of two-way interactions, we applied the COMPARE approach (Howell & Lacroix, 2012).

We used SPSS AMOS Graphics 26 (IBM Knowledge Center, 2019a) for performing the CFA and carried out all other analyses with SPSS 26 (IBM Knowledge Center, 2019b).

Pilot Field Study

Procedure, Material, and Participants

Overall, 61 primary school teachers working in 20 schools out of 261 teachers (response rate: 23.4%) working in 22 schools participated in an online survey at the end of a health promotion program in June 2019. The “Water Schools in Lower Austria” program intended to encourage pupils (6 – 10 years) to drink more water in schools instead of other liquids and to increase the pupils’, the parents’, and the teachers’ positive attitudes toward drinking water. The intervention included a 2-h-workshop for teachers at the beginning of the intervention in October 2018, the implementation of drinking rules, and the provision of free water bottles for each pupil. Within a process and outcome evaluation of the health promotion project, we administered the three instruments. Due to practical constraints, the items in the instruments were not randomly ordered.

Fifty-nine teachers were women, one was a man, and one person did not indicate her/his gender. This gender distribution reflects the distribution in the Austrian education system (Statistik Austria, 2019). Overall, the participants reported a mean age of 40.4 years, and a mean work experience of 17.1 years. The participating teachers were informed about the aim, the duration, and the demands of the study. By completing the online instruments, teachers provided informed consent. Approval from the ethics committee at the Danube University Krems was obtained on July 12th, 2018 (GZ EK 31/2015-2018) and from the School Board of the state of Lower Austria on August 20th, 2018 (Praes.-420/2410-2018). This was necessary because pupils were involved in the larger and separate evaluation of the health promotion project.

Data Analyses

Similar to study 1, for assessing structural validity we ran EFAs and CFAs based on the same model fit guidelines. Additionally, we calculated Cronbach’s α to estimate internal consistency of the scales.

Results

Hypothetical Situations Study

Structural Validity

Based on the results of 142 respondents, we conducted an EFA for each hypothesized construct separately to assess unidimensionality of the scales. For all three EFAs the KMO criteria and Bartlett’s Tests of Sphericity showed satisfactory

results (i.e., $KMO \geq 0.6$; p -value for Bartlett’s Test of Sphericity ≤ 0.05 ; see Table 2) proving the correlation of the four items in each instrument. The instrument specific analyses yielded just one factor including all four items of each instrument (eigenvalues > 1). The explained variance for each one ranged from 79.35% for FIM to 85.90% for IAM.

Based on the answers of the same sample ($n = 142$) we calculated a three-factor CFA with maximum likelihood estimation. The factor loadings ranged from 0.81 to 0.92 (see Fig. 1) and the overall model fit was adequate with $CFI = 0.97$ and $RMSEA = 0.08$ [90% CI 0.06–0.11]. The correlations between AIM and IAM ($r = 0.35$) and AIM and FIM ($r = 0.26$) were moderate. No relevant correlation between IAM and FIM could be observed ($r = 0.06$). In general, the three-factor CFA fitted the data better than

the one-factor and the two-factor CFA. The two-factor CFA showed factor loadings ranging between 0.33 and 0.94. The CFI value was 0.65 and the RMSEA value was 0.27 [90% CI 0.25–0.29]. In the one-factor model the factor loadings varied from 0.04 to 0.93, and the model fit parameters were dissatisfying: $CFI = 0.40$; $RMSEA = 0.35$ [90% CI 0.33–0.37].

Internal Consistency

We calculated Cronbach’s α as estimates for the internal consistency of the scales. The Cronbach’s α values were 0.94 each for AIM and IAM, and 0.91 for FIM. Detailed item scale statistics are depicted in Additional File 2.

Table 2 EFAs for assessing unidimensionality of each instrument separately (Study 1: $n = 142$)

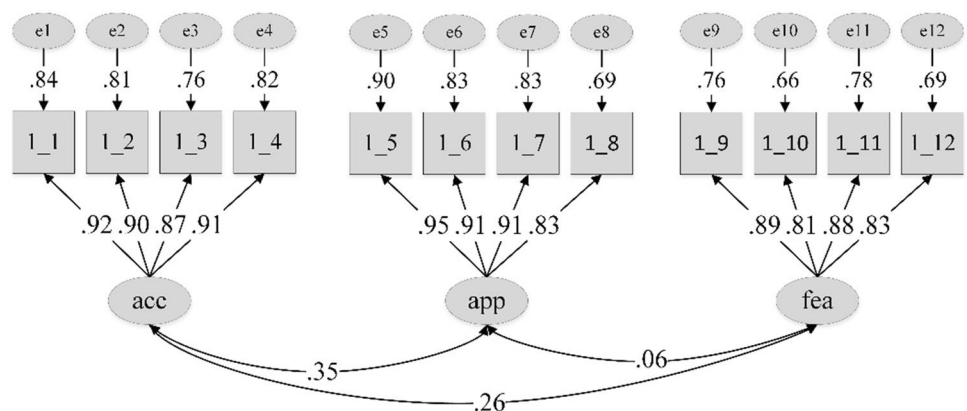
	Communalities	Component	Eigenvalue	Explained variance	KMO	χ^2	df	p
AIM								
1. “X” ^a meets my approval	0.867	0.931						
2. “X” is appealing to me	0.863	0.929						
3. I like “X”	0.832	0.912						
4. I welcome “X”	0.866	0.930	3.428	85.69%	0.87	523.89	6	<0.000
IAM								
5. “X” seems fitting	0.908	0.953						
6. “X” seems suitable	0.866	0.931						
7. “X” seems applicable	0.874	0.935						
8. “X” seems like a good match	0.788	0.888	3.436	85.90%	0.86	551.96	6	<0.000
FIM								
9. “X” seems implementable	0.831	0.911						
10. “X” seems possible	0.747	0.864						
11. “X” seems doable	0.829	0.910						
12. “X” seems easy to use	0.767	0.876	3.174	79.35%	0.84	390.93	6	<0.000

Extraction method: principal component analysis; Rotation: Promax with Kaiser Normalization. These results relate to the German version of the measures AIM, IAM, and FIM

df degrees of freedom, KMO Kaiser-Maier-Olkin measure of sampling adequacy criterion, p probability value, χ^2 Chi²

^aInstead of “X”, the name of the intervention should be included

Fig. 1 Confirmatory factor analysis (Study 1: $n = 142$)



Construct Validity—Known-Groups Validity

We investigated the known-groups validity of the instruments applying 3 × 2 ANOVAs (see Table 3). The largest effects on the respective instruments (η^2 ranging from 0.630–0.805) were observed according to the manipulation of the antecedents (see *Procedure, Material, and Participants* sections). However, to a smaller extent, the manipulation of the respective antecedents of the acceptability and the appropriateness construct influenced all three scale ratings.

Specifically, the analysis for AIM showed a large statistically significant main effect for the manipulation of the antecedents of acceptability, $F(1, 138) = 245.81, p < 0.000, \eta^2 = 0.64$, and a moderate statistically significant main effect for the manipulation of the antecedents of appropriateness, $F(1, 138) = 9.80, p = 0.002, \eta^2 = 0.07$. The analysis for IAM revealed a small statistically significant interaction effect between the manipulation of the antecedents of acceptability and appropriateness, $F(1, 137) = 6.36, p = 0.013, \eta^2 = 0.04$. The interaction effect occurred because the differences between the groups “acceptability low” and “acceptability high” were larger in the “appropriateness low” than in the “appropriateness high” group (depicted in Fig. 2). In other words, the manipulation of appropriateness had a larger effect in the “low acceptability” groups (see Additional File 3 for details). The main effect for the manipulation of appropriateness’ antecedents was very large, $F(1, 137) = 566.75, p < 0.000, \eta^2 = 0.81$. Finally, the analysis for FIM detected a statistically significant interaction effect between the manipulation of the antecedents of acceptability and appropriateness,

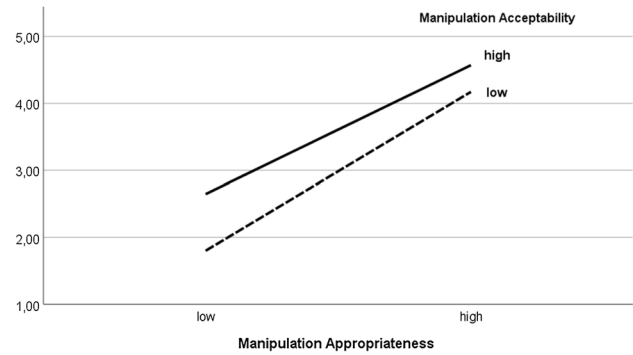


Fig. 2 Interaction effect between manipulation antecedents of acceptability and appropriateness on IAM (Study 1: $n = 142$)

$F(1, 137) = 16.94, p < 0.000, \eta^2 = 0.11$ and a large main effect for the expected manipulation of the antecedents of feasibility, $F(1, 137) = 232.83, p < 0.000, \eta^2 = 0.63$. The interaction effect was as follows: Within the “appropriateness low” group the manipulation of the acceptability’s antecedents did not affect the feasibility ratings. However, in the “appropriateness high” group, the “acceptability high” group rated the intervention as more feasible than the “acceptability low” group (see Fig. 3, and Additional File 3). The hypothetical situations consisting of the patterns “appropriateness high” and “acceptability low” may have confused the participants. Around four out of ten participants (41%) indicated that they doubted the plausibility of this hypothetical situation, while on average two out of ten people (23%) doubted the plausibility of the vignettes in general. The statistical models explained 64%, 81%, and 65% of the variance (adjusted R^2 reported).

Table 3 ANOVAs for assessing known-groups validity

Source	df	Dependent variables								
		Acceptability (AIM)			Appropriateness (IAM)			Feasibility (FIM)		
		F	p	η^2	F	p	η^2	F	p	η^2
ACC	1	245.81	< 0.000	0.640	47.80	< 0.000	0.259	19.52	< 0.000	0.125
APP	1	9.80	0.002	0.066	566.75	< 0.000	0.805	0.05	0.821	0.000
FEA	1	0.92	0.339	0.007	2.16	0.144	0.016	232.83	< 0.000	0.630
ACC x APP	–	–	–	–	6.36	0.013	0.044	16.94	< 0.000	0.110
ACC x FEA	–	–	–	–	–	–	–	–	–	–
APP x FEA	–	–	–	–	–	–	–	–	–	–
ACC x APP x FEA	–	–	–	–	–	–	–	–	–	–
		$R^2 = 0.65$ (Adj. $R^2 = 0.64$)			$R^2 = 0.82$ (Adj. $R^2 = 0.81$)			$R^2 = 0.66$ (Adj. $R^2 = 0.65$)		

ACC manipulation acceptability, APP manipulation appropriateness, FEA manipulation feasibility, df degree of freedoms, p probability value, η^2 effect estimate

“–” indicating that interaction effects were not statistically significant and therefore dropped from the analysis

“x” refers to the investigation of interaction effect

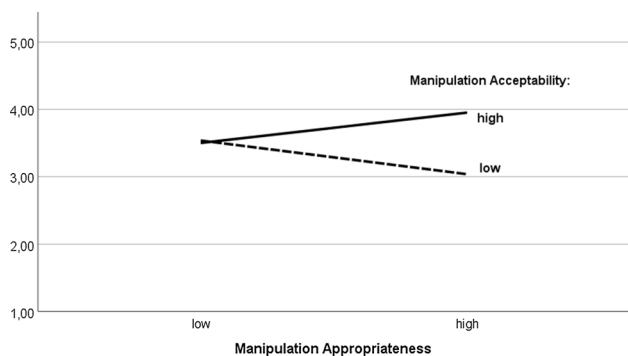


Fig. 3 Interaction effect between manipulation antecedents of acceptability and appropriateness on FIM (Study 1: $n = 142$)

Pilot Field Study

Structural Validity

After calculating descriptive statistics, we started to assess unidimensionality by applying EFAs separately for each instrument based on 61 respondents. For AIM the KMO criterion (0.84) and Bartlett’s Test of Sphericity ($\chi^2 = 232.53$; $df = 6$; $\alpha < 0.000$) showed satisfactory results. All items were loading high (ranging from 0.918 to 0.938) on one component with an eigenvalue > 1 (see Table 4). Two items within each of the measures IAM (#6 “suitable” and #8 “good match”) and FIM (#9 “implementable” and #10 “possible”) correlated perfectly ($r = 1$) with each other. Therefore, we

decided to omit the items #6 “suitable” and #9 “implementable” from further analyses, as these were each the first in the row. Due to the perfect correlation, this decision did not affect the further analyses. Subsequently, the preconditions for conducting EFAs were met (see Table 4). The EFAs of the trimmed versions of IAM and FIM yielded a one component solution in each case, explaining 93.67%, respectively, 88.92% of variance.

The three-factor CFA with maximum likelihood estimation failed to reveal a good model fit as indicated by the model fit parameters (CFI = 0.98; RMSEA = 0.10 [90% CI 0.044–0.150]). Factor loadings ranged from 0.83 to 0.99. The factors correlated highly with each other, appropriateness and feasibility showing the highest correlation with 0.86. We could only establish a good model fit after allowing the error terms of the items #8 “seems like a good match” (IAM) and #10 “seems possible” (FIM) to correlate, (CFI: 0.99; RMSEA: 0.05 [90% CI 0.00–0.11]), see Fig. 4 for details. As the items of IAM and FIM are similar in structure, all starting with the word “seems” (“scheint” in the German version), and the items of all three scales were not randomly ordered, the correlation of the error terms might account for the assumed sequence effect and the semantic similarities.

Other established models did not reveal a better model fit. We calculated a two-factor CFA with appropriateness and feasibility items loading on one factor and acceptability items as another factor. The model fit parameters were then: CFI = 0.85; RMSEA = 0.19 [90% CI 0.15–0.23]. We

Table 4 EFA for assessing unidimensionality of each instrument separately (Study 2: $n = 61$)

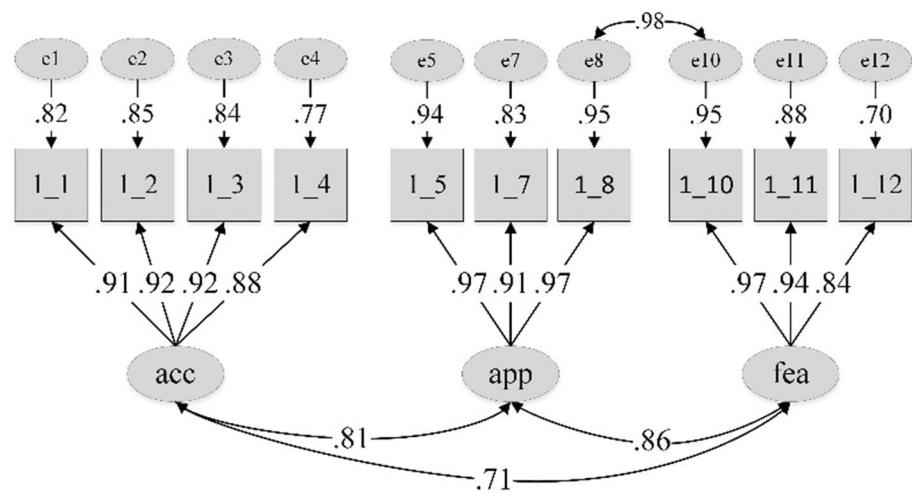
	Communalities	Component	Eigenvalue	Explained variance	KMO	χ^2	df	p
AIM								
1. “X” ^a meets my approval	0.867	0.935						
2. “X” is appealing to me	0.863	0.929						
3. I like “X”	0.832	0.938						
4. I welcome “X”	0.866	0.918	3.460	86.49%	0.84	232.53	6	<0.000
IAM								
5. “X” seems fitting	0.950	0.975						
6. “X” seems suitable	–	–						
7. “X” seems applicable	0.910	0.954						
8. “X” seems like a good match	0.950	0.975	2.810	93.67%	0.77	224.30	3	<0.000
FIM								
9. “X” seems implementable	–	–						
10. “X” seems possible	0.936	0.968						
11. “X” seems doable	0.900	0.948						
12. “X” seems easy to use	0.832	0.912	2.688	88.92%	0.72	169.37	3	<0.000

Extraction method: principal component analysis; Rotation: Promax with Kaiser Normalization. These results relate to the German version of the measures AIM, IAM, and FIM

df degrees of freedom, KMO Kaiser-Maier-Olkin Measure of Sampling Adequacy criterion, p probability value, $\chi^2 = \text{Chi}^2$

^aInstead of “X”, the name of the intervention should be included

Fig. 4 Confirmatory factor analysis (Study 2: $n=61$)



calculated another two-factor CFA with acceptability and appropriateness items loading on one factor and feasibility items as a single factor. The model fit parameters were then: CFI=0.91; RMSEA=0.24 [90% CI 0.20–0.28]. A single factor model also showed unsatisfying model fit with: CFI=0.78; RMSEA=0.28 [90% CI 0.25–0.32].

Internal Consistency

The analysis of study 2 based on 61 respondents revealed similar results to study 1. The Cronbach's α values were 0.95 for AIM (4 items), 0.97 for IAM (3 items), and 0.93 for FIM (3 items). Detailed item scale statistics are depicted in Additional File 2.

Discussion

The assessment of implementation outcomes can provide immediate feedback regarding the success or non-success of an implementation endeavor. The importance of this topic for both researchers and practitioners is also reflected in the Society for Implementation Research Collaboration Instrument Review Project (Society for Implementation Research Collaboration {SIRC}, 2021), and the newly developed and openly accessible Implementation Outcome Repository (Centre for Implementation Science—King's College London, 2019). It allows for searching for instruments to measure implementation outcomes in the English language.

To our best knowledge, our study provided the first instruments that are generic, reliable, and valid in the German language to assess the implementation outcomes acceptability, appropriateness, and feasibility in a comparable manner. We tested the reliability and validity of the German versions of AIM, IAM, and FIM in a vignette study, creating optimal conditions as variance was induced. This study found high

internal consistency of the scales and revealed the hypothesized three-factor solution. Where differences in hypothetical situations were pronounced, participants could differentiate between the three distinct scales. However, the variation of the antecedents of acceptability and appropriateness also influenced the ratings of the other measures, although to a lesser extent than the respective main effects.

In the pilot field study with primary school teachers participating in a health promotion project, the hypothesized three-factor solution of the CFA exhibited the best comparable model fit. Nonetheless, we only achieved a good model fit once we allowed the correlation of the error terms of the items #8 “seems like a good match” (IAM) and #10 “seems possible” (FIM). This accounted for the assumed sequence effect and the semantic similarities. Finally, the AIM, IAM, and FIM showed high internal consistency values in both studies.

Consistent with the theoretical assumptions and previous psychometric assessment, confirmatory factor analyses applied in the vignette study yielded three correlated factors. The correlations between the factors (r ranging from 0.06 to 0.35) were low to moderate in the vignette study. Furthermore, the results of the ANOVAs showed that the variations of the antecedents of acceptability (i.e., individual's openness, and possibility to undo the implementation of the intervention) influenced not only the respective AIM, but to a lesser extent also IAM and FIM. Weiner and colleagues' study showed higher correlations between different factors (r ranging from 0.36 to 0.77), which might be imposed due to a study design error (Weiner et al., 2017). However, the correlations between the different factors in the pilot field study ranged from 0.71 to 0.86, highlighting the conceptually distinguishable implementation outcomes being hard to distinguish empirically (Proctor et al., 2011).

When programs are designed to be implemented in different settings, in our case, the school setting, program

developers consider several assumed barriers of implementation (Pearson et al., 2015) and aim at developing an acceptable, appropriate, and feasible intervention to foster adoption and use of the program (Chor et al., 2015). Furthermore, when schools already participate in a health promotion project, the decision to adopt a program has already been made. Both influencing factors would lead to less variance in the data. This was reflected in the two data sets of our investigations. The results of the pilot field study compared with the vignette study showed that the item means (see Additional File 2) were higher in the pilot field study (ranging from 4.31 to 4.43 vs. 3.24 to 3.70) and the standard deviation was lower (ranging from 0.50 to 0.65 vs. ranging from 1.09 to 1.39). Additionally, the higher mean values in the pilot field study might indicate that only teachers being most satisfied with the intervention may have participated in the study (i.e., selection bias). The high intercorrelation of the implementation outcomes in the field study makes it difficult to disentangle the hypothesized factor structure.

A strength of this study is that we assessed the psychometric criteria of the measures by applying them in two different studies. The hypothetical situation study enabled us to assess the instruments under optimal conditions. We cannot exclude the possibility, however, that the artificial situation provokes the association of a test situation resulting in a more in-depth reading and then paying closer attention to the wording of the items. Therefore, the results may not be transferable to field studies. Unfortunately, we did not record the time respondents needed to fill in the instruments in either study to assess this hypothesis. The sample size of the health promotion study's evaluation was just above the recommended threshold of five times the number of items (Bentler & Chou, 1987). Although the teachers participating in the program were reminded several times, the response rate only reached 23.4%. Therefore, the results of this study should be interpreted cautiously.

Another limitation of our study is that we could not assess the psychometric equivalence of the constructs in question across the English original and the translated German version of the scales (i.e., *measurement invariance*). Such analyses are considered a requirement for cross-cultural comparisons (Boer et al., 2018; Fischer & Karl, 2019; Putnick & Bornstein, 2016); however, we did not have access to an English speaking population.

Currently, the AIM, IAM, and FIM have not been tested extensively except in one study (Weiner et al., 2017), yet other studies applying these instruments have recently been published (Adrian et al., 2020; Damush et al., 2021; Swindle et al., 2021; Taboada et al., 2021; Wasil et al., 2021). The first results are promising; however, further field studies are necessary to test the psychometric criteria, and especially further investigate the scales' construct validity and predictive validity. We also encourage other researchers using

these scales to report on psychometric criteria. Based on our experience, we propose the following recommendations for future applications of the scales: First, the instruments should be applied at an earlier stage, within the initial implementation phase of the process (Pearson et al., 2015). Also Proctor et al. (2011) hypothesized that acceptability, appropriateness, and feasibility are especially relevant at an early stage in the implementation process, influencing adoption (i.e., uptake) and penetration (i.e., continuing use) of an intervention. At this stage the providers' perception of the implementation outcomes may vary more, as people may experience more difficulties which may be already settled after approximately one school year. In addition, to avoid selection bias, a high participation rate in the study should be obtained. Second, because respondents seem to consider the items to be very similar, the optimal order of the questions should be investigated empirically. One possibility would be to administer the items of three scales randomly to avoid sequence effects. The other possibility would be to administer the scales and provide their definitions highlighting the differences to raise participants' awareness while answering. Third, once the hypothesized factor structure is proven by further field studies, the development of ultra-short scales, with fewer items per scale could be pursued.

Conclusions

The German versions of AIM, IAM, and FIM were shown to be promising, reliable, and valid instruments for the assessment of the implementation outcomes acceptability, appropriateness, and feasibility in a hypothetical testing situation. The pilot field study reinforced these results with some limitations. Nevertheless, further field studies, especially in the initial implementation phase, are necessary to provide additional evidence regarding the scales' structural validity, internal consistency, and predictive validity. Based on the current experience with these tools, we suggest using all four items per instrument instead of an abbreviated version; however, we believe that some items can be removed once further research and testing is completed.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s43477-021-00019-y>.

Acknowledgements We are thankful to all teachers who supported us with their valuable time and participated in the study. We thank Deborah Starkey for her support in translating the instruments and Bryan Weiner for his suggestions on constructing hypothetical situations. We also acknowledge Rahel Naef's and Gabi Schmid-Mohler's contribution to include a German abbreviation of the measures. We are thankful to Silke Kuske for the valuable discussion about the constructs.

Author Contributions CK developed the aims, the methodology, and drafted the protocol. UG, MTS, KJT, and TS provided feedback to the

protocol. CK and UG contributed to data collection. CK conducted the analysis and wrote the first draft of the manuscript. UG, MTS, KJT, and TS contributed to authorship and all approved the final manuscript.

Funding Open access funding provided by Danube University Krems University for Continuing Education. Study 1 did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. Study 2 was partly funded by the Lower Austrian Health and Social Fund. As this was a participatory evaluation of a health promotion project, the funding body was entitled to select relevant outcomes of the evaluation, but was not involved in the collection of the data or the analysis process.

Data Availability The datasets used and/or analyzed during the current study as well as the applied vignettes are available from the corresponding author on request.

Declarations

Conflict of interest The authors declare that they have no competing interests.

Ethical Approval The study protocol and the questionnaires were submitted to the ethics committee of the Medical University of Vienna that waived responsibility (personal e-mail communication, 12th February, 2019), since survey/questionnaire studies with experts/professionals do not need to be submitted to an ethics committee in Austria. Approval from the ethics committee at the Danube University Krems for the conduct of study 2 was obtained on the 12th July 2018 (GZ EK 31/2015-2018) and from the School Board of the state of Lower Austria on the 20th August 2018 (Präs.-420/2410-2018). This was necessary because pupils were involved in the larger and separate evaluation of the health promotion project.

Consent to Participate All participating teachers were informed about the aim, the duration, and the demands of the study a priori in a leaflet. By filling in the paper–pencil questionnaires, teachers provided informed consent.

Consent to Publish Not applicable as research results are presented on an aggregated level.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Adrian, M., Coifman, J., Pullmann, M. D., Blossom, J. B., Chandler, C., Coppersmith, G., Thompson, P., & Lyon, A. R. (2020). Implementation determinants and outcomes of a technology-enabled service targeting suicide risk in high schools: Mixed methods study. *JMIR Mental Health*, 7(7), e16338. <https://doi.org/10.2196/16338>
- Bentler, P. M., & Chou, C.-P. (1987). Practical issues in structural modeling. *Sociological Methods & Research*, 16(1), 78–117. <https://doi.org/10.1177/0049124187016001004>
- Boer, D., Hanke, K., & He, J. (2018). On detecting systematic measurement error in cross-cultural research: A review and critical reflection on equivalence and invariance tests. *Journal of Cross-Cultural Psychology*, 49(5), 713–734. <https://doi.org/10.1177/0022022117749042>
- Browne, M., & Cudeck, R. (1993). Alternative ways of assessing model fit. In K. Bollen & J. Long (Eds.), *Testing structural equation models* (pp. 136–162). Sage.
- Bührmann, L., Schuurmans, J., Ruwaard, J., Fleuren, M., Etzelmüller, A., Piera-Jiménez, J., Finch, T., Rapley, T., Potthoff, S., Aouizerate, B., & Batterham, P. J. (2020). Tailored implementation of internet-based cognitive behavioural therapy in the multinational context of the ImpleMentAll project: A study protocol for a stepped wedge cluster randomized trial. *Trials*. <https://doi.org/10.1186/s13063-020-04686-4>
- Centre for Implementation Science - King's College London. (2021, July 13). *Implementation Outcome Repository*. <https://implementationoutcomerepository.org/>
- Chor, K. H., Wisdom, J. P., Olin, S. C., Hoagwood, K. E., & Horwitz, S. M. (2015). Measures for predictors of innovation adoption. *Administration and Policy in Mental Health*, 42(5), 545–573. <https://doi.org/10.1007/s10488-014-0551-7>
- Cross-Cultural Survey Guidelines. (2016). Chapter: Translation. In *Guidelines for best practice in cross-cultural surveys. Full guidelines* (pp. 853). University of Michigan.
- Damush, T. M., Penney, L. S., Miech, E. J., Rattray, N. A., Baird, S. A., Cheatham, A. J., Austin, C., Sexson, A., Myers, L. J., & Bravata, D. M. (2021). Acceptability of a complex team-based quality improvement intervention for transient ischemic attack: A mixed-methods study. *BMC Health Services Research*, 21(1), 453. <https://doi.org/10.1186/s12913-021-06318-2>
- Dziuban, C. D., & Shirkey, E. C. (1974). When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological Bulletin*, 81(6), 358–361. <https://doi.org/10.1037/h0036316>
- Epstein, J., Santo, R. M., & Guillemin, F. (2015). A review of guidelines for cross-cultural adaptation of questionnaires could not bring out a consensus. *Journal of Clinical Epidemiology*, 68(4), 435–441. <https://doi.org/10.1016/j.jclinepi.2014.11.021>
- Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods*, 39(2), 175–191. <https://doi.org/10.3758/BF03193146>
- Finch, T. L., Mair, F. S., O'Donnell, C., Murray, E., & May, C. R. (2012). From theory to “measurement” in complex interventions: Methodological lessons from the development of an e-health normalisation instrument. *BMC Medical Research Methodology*, 12, 69. <https://doi.org/10.1186/1471-2288-12-69>
- Fischer, R., & Karl, J. A. (2019). A primer to (cross-cultural) multi-group invariance testing possibilities in R. *Frontiers in Psychology*. <https://doi.org/10.3389/fpsyg.2019.01507>
- Greenhalgh, T., Robert, G., Macfarlane, F., Bate, P., & Kyriakidou, O. (2004). Diffusion of innovations in service organizations: Systematic review and recommendations. *Milbank Quarterly*, 82(4), 581–629. <https://doi.org/10.1111/j.0887-378X.2004.00325.x>
- Guillemin, F., Bombardier, C., & Beaton, D. (1993). Cross-cultural adaptation of health-related quality of life measures: Literature review and proposed guidelines. *Journal of Clinical Epidemiology*, 46(12), 1417–1432. [https://doi.org/10.1016/0895-4356\(93\)90142-n](https://doi.org/10.1016/0895-4356(93)90142-n)

- Gutt, A.-K., Hoben, M., Roes, M., Willmeroth, T., Wesselborg, B., & Kuske, S. (2018). Systematische Übersetzung und Cross-Validierung definierter Implementierungsergebnisse der Gesundheitsversorgung. *Zeitschrift Für Evidenz, Fortbildung Und Qualität Im Gesundheitswesen*, 135–136, 72–80. <https://doi.org/10.1016/j.zefq.2018.06.005>
- Howell, G. T., & Lacroix, G. L. (2012). Decomposing interactions using GLM in combination with the COMPARE, LMATRIX and MMATRIX subcommands in SPSS. *Tutorials in Quantitative Methods for Psychology*. <https://doi.org/10.20982/tqmp.08.1.p001>
- IBM Knowledge Center. (2019, July 17a). *SPSS Amos 26 documentation*. <https://www-01.ibm.com/support/docview.wss?uid=ibm10878803>
- IBM Knowledge Center. (2019, July 17b). *SPSS statistics documentation*. <https://www.ibm.com/support/knowledgecenter/SSLVMB>
- Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika*, 39(1), 31–36. <https://doi.org/10.1007/BF02291575>
- Kien, C., Schultes, M. T., Szelag, M., Schoberberger, R., & Gartlehner, G. (2018). German language questionnaires for assessing implementation constructs and outcomes of psychosocial and health-related interventions: A systematic review. *Implementation Science*, 13(1), 150. <https://doi.org/10.1186/s13012-018-0837-3>
- Lewis, C. C., Fischer, S., Weiner, B. J., Stanick, C., Kim, M., & Martinez, R. G. (2015). Outcomes for implementation science: An enhanced systematic review of instruments using evidence-based rating criteria. *Implementation Science*, 10, 155. <https://doi.org/10.1186/s13012-015-0342-x>
- Lewis, C. C., Mettert, K. D., Dorsey, C. N., Martinez, R. G., Weiner, B. J., Nolen, E., Stanick, C., Halko, H., & Powell, B. J. (2018). An updated protocol for a systematic review of implementation-related measures. *Systematic Review*, 7(1), 66. <https://doi.org/10.1186/s13643-018-0728-3>
- McDowell, I. (2006). *Measuring health: A guide to rating scales and questionnaires* (3rd ed.). Oxford University Press.
- Mettert, K., Lewis, C., Dorsey, C., Halko, H., & Weiner, B. (2020). Measuring implementation outcomes: An updated systematic review of measures' psychometric properties. *Implementation Research and Practice*, 1, 263348952093664. <https://doi.org/10.1177/2633489520936644>
- Nilsen, P. (2015). Making sense of implementation theories, models and frameworks. *Implementation Science*, 10, 53. <https://doi.org/10.1186/s13012-015-0242-0>
- Pearson, M., Chilton, R., Wyatt, K., Abraham, C., Ford, T., Woods, H. B., & Anderson, R. (2015). Implementing health promotion programmes in schools: A realist systematic review of research and experience in the United Kingdom. *Implementation Science*, 10, 149. <https://doi.org/10.1186/s13012-015-0338-6>
- Proctor, E., Silmere, H., Raghavan, R., Hovmand, P., Aarons, G., Bunger, A., Griffey, R., & Hensley, M. (2011). Outcomes for implementation research: Conceptual distinctions, measurement challenges, and research agenda. *Administration and Policy in Mental Health*, 38(2), 65–76. <https://doi.org/10.1007/s10488-010-0319-7>
- Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions for psychological research. *Developmental Review*, 41, 71–90. <https://doi.org/10.1016/j.dr.2016.06.004>
- Rossiter, J. R. (2002). The C-OAR-SE procedure for scale development in marketing. *International Journal of Research in Marketing*, 19(4), 305–335. [https://doi.org/10.1016/S0167-8116\(02\)00097-6](https://doi.org/10.1016/S0167-8116(02)00097-6)
- Schreiber, J. B., Nora, A., Stage, F. K., Barlow, E. A., & King, J. (2006). Reporting structural equation modeling and confirmatory factor analysis results: A review. *The Journal of Educational Research*, 99(6), 323–338. <https://doi.org/10.3200/JOER.99.6.323-338>
- Schweizer, K. (2010). The adaptation of assessment instruments to the various European languages. *European Journal of Psychological Assessment*, 26(2), 75–76. <https://doi.org/10.1027/1015-5759/a000011>
- Society for Implementation Research Collaboration (SIRC). (2021, July 13). *Instrument review project*. <https://societyforimplementationresearchcollaboration.org/sirc-instrument-project/>
- Souza, A. C., Alexandre, N. M. C., & Guirardello, E. B. (2017). Psychometric properties in instruments evaluation of reliability and validity. *Epidemiologia e Serviços De Saúde*, 26(3), 649–659. <https://doi.org/10.5123/s1679-49742017000300022> (Propriedad espsicometricasnaavaliacaodeinstrumentos:avaliacaodaconfiabilidadeevalidade.)
- Statistik Austria. (2019, June 15). *Teachers including teachers currently on leave of absence in the school year 2019/20* https://pic.statistik.at/web_de/statistiken/menschen_und_gesellschaft/bildung/schulen/lehrpersonen/125683.html
- Swindle, T., McBride, N. M., Selig, J. P., Johnson, S. L., Whiteside-Mansell, L., Martin, J., Staley, A., & Curran, G. M. (2021). Stakeholder selected strategies for obesity prevention in child-care: Results from a small-scale cluster randomized hybrid type III trial. *Implementation Science*. <https://doi.org/10.1186/s13012-021-01119-x>
- Taboada, A., Ly, E., Ramo, D., Dillon, F., Chang, Y. J., Hooper, C., Yost, E., & Haritatos, J. (2021). Implementing goal mama: Barriers and facilitators to introducing mobile health technology in a public health nurse home-visiting program. *Global Qualitative Nursing Research*, 8, 23333936211014496. <https://doi.org/10.1177/23333936211014497>
- Wasil, A. R., Taylor, M. E., Franzen, R. E., Steinberg, J. S., & DeRubeis, R. J. (2021). Promoting graduate student mental health during COVID-19: Acceptability, feasibility, and perceived utility of an online single-session intervention. *Frontiers in Psychology*, 12, 569785. <https://doi.org/10.3389/fpsyg.2021.569785>
- Weiber, R., & Mühlhaus, D. (2014). *Strukturgleichungsmodellierung. Eine anwendungsorientierte Einführung in die Kausalanalyse mit Hilfe von AMOS, SmartPLS und SPSS*. Heidelberg: Springer.
- Weiner, B. J., Lewis, C. C., Stanick, C., Powell, B. J., Dorsey, C. N., Clary, A. S., Boynton, M. H., & Halko, H. (2017). Psychometric assessment of three newly developed implementation outcome measures. *Implementation Science*, 12(1), 108. <https://doi.org/10.1186/s13012-017-0635-3>