BMJ Open

Multimethods study to develop tools for competency-based assessments of implementation research training programmes in low and middle-income countries

Olakunle Alonge ,[1] Aditi Rao,[2] Anna Kalbarczyk ,[2] Latifat Ibisomi,[3] Phyllis Dako-Gyeke ,[4] Yodi Mahendradhata,[5] Carlos Alberto Rojas,[6] Choolwe Jacobs,[7] Kwasi Torpey,[8] Edwin Rolando Gonzalez Marulanda,[6] Pascal Launois,[9] Mahnaz Vahedi[9]

To cite: Alonge O, Rao A, Kalbarczyk A, *et al*. Multimethods study to develop tools for competency-based assessments of implementation research training programmes in low and middle-income countries. *BMJ Open* 2024;**14**:e082250. doi:10.1136/bmjopen-2023-082250

► Prepublication history and additional supplemental material for this paper are available online. To view these files, please visit the journal online (https://doi.org/10.1136/bmjopen-2023-082250).

Received 17 November 2023
Accepted 01 July 2024

For numbered affiliations see end of article.

Correspondence to
Dr Olakunle Alonge;
oalonge@uab.edu

ABSTRACT

Introduction Methods and validated tools for evaluating the performance of competency-based implementation research (IR) training programmes in low–middle-income countries (LMICs) are lacking. In this study, we developed tools for assessing the performance of IR training programmes based on a framework of IR core competencies in LMICs.

Methods We developed self-assessment and objective-assessment tools drawing on the IR competency framework. We used exploratory factor analyses and a one-parameter logistic model to establish construct validity and internal consistency of the tools drawing on a survey conducted in 2020 with 166 trainees before and after an index IR course across five universities and LMICs under the Special Program for Research and Training in Tropical Diseases postgraduate IR training scheme. We conducted key informant interviews (KII) with 20 trainees and five trainers to reflect on the usefulness of the tools and framework for guiding IR training in LMICs.

Results Two 16-item tools for self-assessment of IR knowledge and self-efficacy and a 40-item objective assessment tool were developed. The factor loadings of items in the self-assessment tools were 0.65–0.87 with Cronbach's alpha (α) of 0.97, and 0.77–0.91 with α of 0.98 for the IR knowledge and self-efficacy tools, respectively. The distribution of item difficulty in the objective-assessment tool was consistent before and after the index IR course. Pearson correlation coefficient (r) between the self-assessed and objectively assessed IR knowledge before the index IR course was low, r=0.27 (p value: <0.01), with slight improvements after the index IR course, r=0.43 (p value: <0.01). All KII respondents reported the assessment tools and framework were valid for assessing IR competencies.

Conclusion The IR competency framework and tools developed for assessing IR competencies and performance of IR training programmes in LMICs are reliable and valid. Self-assessment methods alone may not suffice to yield a valid assessment of performance in these settings.

STRENGTH AND LIMITATIONS OF THIS STUDY

⇒ The study protocol was implemented across multiple academic institutions and countries—which adds strength to the external validity of the study and to the generalisability of the developed tools for evaluating implementation research (IR) trainings across diverse settings.
⇒ The integration of multiple methods including quantitative surveys, psychometric statistical methods and qualitative methods provides a rigorous and robust approach and adds strength to the internal validity of the study.
⇒ This study draws upon both classical test theory and item response theory in generating a valid approach for assessing IR competencies in low and middle-income countries (LMICs).
⇒ The recruitment strategy—respondents volunteering to participate in the assessment—along with COVID-19 disruptions to the data collection could have introduced a selection bias with respect to the study participants in the different countries.
⇒ Hence, the study findings should not be interpreted as a comparative country/institutional performance assessment of IR training in LMICs.

INTRODUCTION

Implementation research (IR) is increasingly recognised as essential for facilitating effective delivery of evidence-supported interventions and maximising health benefits of these interventions for improving population health, especially the health of vulnerable populations.[1–3] However, core competencies in IR had not been defined, especially in low and middle-income countries (LMICs), nor a consistent curriculum recognised across different IR training programmes globally.[4] To address this gap, a framework

of core competencies in IR was developed by an international consortium of academic partners, primarily based in LMICs, in collaboration with the Special Program for Research and Training in Tropical Diseases (TDR) and other global health agencies using a modified Delphi process involving IR scientists/trainers and students/trainees globally.[4] The framework comprises of 11 domains, 59 competencies and 52 subcompetencies and was developed to guide the objective assessment of training needs and effectiveness of IR training programmes in LMICs as well as the development of future capacity building programmes in IR. The framework highlights the salience of certain IR competencies for LMICs relative to high-income settings, for example, competencies in analysing health systems and contexts, applying ethical principles to IR and communication and advocacy skills.[4] The IR competency framework enables researchers to identify skills needed by teams to respond successfully to implementation challenges surrounding the effective delivery of healthcare programmes and services, in real-time and under real-life conditions, through research embedded in specific contexts.

The framework development exercise highlighted the need to develop methods and tools for evaluating IR training programmes in LMICs using the framework,[4] whereas numerous guidelines for competency-based education and assessment exist in the literature,[4–9] actual examples of a validated and comprehensive tool for assessing competencies in IR were lacking.[10–16] Additionally, there were a few examples of evaluation tools for assessing competencies in implementation practice, but not in IR.[9] The assessment of public health and biomedical training programmes, of which health-related IR training is a subset, has traditionally focused on what and how learners should 'know' and less on what and how they can use their learning 'to do'—that is, solve real-world problems, perform tasks, communicate effectively and make sound decisions.[5–16] Unlike other types of public health or biomedical trainings, IR training calls for a greater emphasis on the application of knowledge and skills from diverse disciplines to solve implementation problems.[17] Hence, competency evaluation tools that assess not only knowledge but also attitudes and demonstrable skills and capabilities in IR are needed. Such tools will place emphasis on learners' performance of tasks through integration of knowledge and skills.

Other studies suggest trainings that cover diverse competencies also require diverse assessment methods, ideally conducted at the beginning and at the end of learning.[18–31] Different assessment methods include objective test questions (which has high reliability and sampling validity, but often serve as a poor measure of performance because it might be affected by other factors outside of the learnt competencies, eg, testing anxiety)[32]; self-assessment (which has been shown to provide reliable and valid assessment of trainees' performance and can improve trainees' self-efficacy and practice if adequate measures are in place to minimise grade inflation among test-takers)[33]; direct observation (which may yield a valid and reliable assessment of performance, but like objective tests, is influenced by other factors outside of the learnt competencies, eg, testing anxiety and Hawthorne effect (trainees behaving differently than they would normally behave because they are being observed))[34 35]; case-based activity and interview (which has similar characteristics as direct observation) and multisource feedback,[36] among others.

To address the gap on how to assess IR competencies and performance of IR training programmes as well as the paucity of assessment tools and methods, this paper describes a multimethods study to develop assessment tools and approaches for assessing the performance of IR training programmes in LMICs based on the IR core competency framework developed for LMICs.[4] It is hoped that the study will contribute to efforts to develop comprehensive IR training programmes and build capacity of researchers and practitioners to conduct IR around effective delivery of life-saving interventions and health system strengthening activities in LMICs.

METHODS

We used multiple methods to achieve the objective of this study. First, we developed quantitative self-assessment and objective-assessment tools based on the IR core competency framework.[37] Second, we used quantitative methods, including exploratory factor analyses and item response theory (IRT) to establish the construct validity and internal consistency of the assessment tools, and reflect on the validity of the IR competency framework on which the tools were based. Finally, we used a quantitative survey and qualitative interviews with key informants to demonstrate the usefulness of the assessment tools and competency framework for guiding IR training and practice among IR trainers and trainees in LMICs.

Setting and study population

This study was conducted in five LMICs, with universities involved with the Joint WHO/UNICEF/UNDP/World Bank TDR in conducting graduate level training programmes in IR (box 1). TDR has been at the forefront of conducting postgraduate training in IR in LMICs, with these institutions representing geographically diverse contexts, at different stages of IR curriculum development

Box 1 Countries and selected universities involved in the study.

1. **Colombia**: National School of Public Health, Universidad de Antioquia.
2. **Ghana**: The School of Public Health, University of Ghana.
3. **Indonesia**: The Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta.
4. **South Africa**: School of Public Health, University of the Witwatersrand.
5. **Zambia**: School of Public Health, University of Zambia.

and training. Four of the universities also serve as centres for regional training in IR under the TDR programme.

All universities participated in the initial IR competency framework development phase and have been conducting graduate-level training in IR for 7–8 years (box 2).

The study was conducted during the peak of the first wave of the COVID-19 pandemic, between January and July 2020. In each institution, students enrolled in index IR courses offered during the study period were included, as well as 1–2 course instructors, with the goal of reaching both trainees and trainers that have first-hand experiences with IR training in LMICs. The index IR courses are 2–4 months long, and cover similar concepts in each institution, including the definitions and principles of IR, how to identify implementation challenges in various health settings, descriptions of implementation strategies and implementation outcomes, and how to specify IR questions that include implementation outcomes.

Quantitative methodology

Developing tools for assessing performance of IR training programmes based on the IR core competency framework

Building on guidelines and examples identified in the literature,[8 10] we used the IR competency framework to develop self-assessment scales for measuring *knowledge* (ie, knowledge of IR principles, concepts and methodologies) and *self-efficacy* (ie, perceived ability and confidence) in enacting IR competencies, adapting self-assessment tools for implementation practice from Moore *et al*.[10] We consolidated the 59 competencies and 52 subcompetencies from the framework (The content validity of the competency statements and the domains and themes they represent have been previously established through a modified Delphi method as part of the IR core competency framework development process[4]), merging statements that were similar and preserving those that were unique, to create an initial questionnaire of 16 items for self-assessing IR knowledge and 16 items for self-assessing IR self-efficacy. Both knowledge and self-efficacy have been recognised as key constructs that influence enactment of competencies, that is, practice and performance of trainees in public health training programmes.[11] The competency statements were consolidated to reduce the cognitive burden in completing the questionnaire. Studies have shown that lengthy questionnaires can contribute to cognitive burden prompting a higher drop-out rate (ie, potential respondents refusing to take or finish the survey) and diminished data quality.[38]

The 16-item self-assessment scale for IR knowledge evaluates individual knowledge across six IR themes: (1) working with stakeholders (competencies 3.1–3.7 in the *IR framework* in online supplemental appendix 1); (2) scientific inquiry (competencies 6.1–6.8, 7.1–7.5 and 8.1–8.5); (3) implementation strategies (competencies 2.1–2.3); (4) resources for IR (competencies 4.1–4.4 and 9.1–9.4); (5) communication and advocacy (competencies 10.1–10.7 and 11.1–11.5) and (6) cross-cutting

Box 2 Description of IR training programmes in selected universities involved in the study.

Colombia: Universidad de Antioquia (UdeA)—Héctor Abad Gómez National School of Public Health (Medellín, Colombia)
IR training programme at UdeA under the Special Programme for Research and Training in Tropical Diseasese (TDR) scheme started in 2016. It is offered as a specialisation under the Master in Epidemiology degree programme, which is a full-time 2-year (four semesters) research-oriented programme. Students must conduct one IR study and write one final report or submit one paper over the final two semesters. Prior to conducting IR, students are required to complete research courses I–IV and seminars I–IV which cover IR concepts and methods, and their application. During the first semester, these courses are complemented with the TDR MOOC in implementation research. Since inception, 15 postgraduate students have graduated as part of the first cohort (2016–2018). The second cohort (2018–2020) currently has nine students, enrolled in August 2018, and third Cohort (2019–2021) has 10 students, enrolled in September 2019.

Ghana: School of Public Health, University of Ghana (Accra, Ghana)
The University of Ghana School of Public Health (UGSPH) ran the phase 1 of the TDR Postgraduate Training Scheme from 2015 to 2019. During this 5-year period, UGSPH trained a total of 50 postgraduates (46 Master of Public Health (MPH)/Master of Science (MSc) and 4 PhD). In the current 2022–2023 academic year, which marks the beginning of the phase 2, 14 students are registered for MPH and MSc as the first cohort. There are about 50 faculty from across six departments in the school who contribute towards the skill building of the students. In addition to implementation research courses, other courses taught for these programmes include Statistics for Social Science; Behavioral Sciences; Research Methods in Public Health; Community Mobilization for Health and Development; Social Theories and Public Health Practice; Theories and Models of Health Promotion; Social Epidemiology; Epidemiology; Health Systems; Essentials of Environment and Occupation on Human Health; Gender and Health; Health Psychology; Management and Report Writing; Applied Medical Anthropology; Population Studies and Reproductive health; Biostatistics. Students also participate in customised workshops and seminars which support the design, conduct and write-up of their implementation research focused dissertations.

Indonesia: Faculty of Medicine, Public Health and Nursing, Gadjah Mada University (Yogyakarta, Indonesia)
IR training at the Faculty of Medicine, Public Health, and Nursing, Universitas Gadjah Mada, Indonesia is conducted as a 2-year MPH degree programme. The programme was established in 2016. As of March 2023, 51 MPH students have graduated from the IR training scheme and 32 are still studying (12 registered in 2022). There are 54 faculty that contribute to the students' learning, with five of them teaching the Implementation Research course. Beside Implementation Research, the postgraduate programme offers courses on Biostatistics; Epidemiology; Social Science and Health Behaviors; Theory and Practice of Public Health; Health Policy and Management; Epidemiology; Control and Management of Tropical Diseases; Public Health Program Management; Research Method; Infection Control; Drugs and Vaccines for Tropical Diseases; Public Health Advocacy; Environmental Health; Public Health Informatics; Zoonosis and Vector Control and Global Health. Students must submit a thesis based on an IR project carried out in their home country to complete requirements for obtaining an MPH degree. Students are also encouraged to publish a manuscript based on their thesis project.

South Africa: School of Public Health, University of the Witwatersrand (Johannesburg, South Africa)

Continued

Box 2 Continued

IR training started at the University of the Witwatersrand in 2016. The master's training is run as a field of study within the MSc Epidemiology degree programme, while the PhD training is housed within the school's interdisciplinary PhD Programme. The requirements for completing the master's degree are spread over 18 months and the PhD typically takes 36–48 months. As of March 2023, 39 (36 MSc and 3 PhDs) students have graduated from the IR training scheme, while 11 are still studying (10 of them registered in 2022). There are over 40 faculties who contribute to the learning of the students with nine of them teaching on IR-focused courses. The courses taught include Implementation Science (the TDR MOOC in IR is a component of this course); Quality Improvement Science; Monitoring and Evaluation; Research Protocol Development; Integration of Qualitative and Quantitative Research Methods; Computing; Biostatistics for Health Researchers; Epidemiology for Health Researchers; Adapting, Implementing, and Evaluating Evidence-based Interventions; Introduction to Health Economics; Data Management in Clinical Research Studies; and Spatial Analysis & GIS in Public Health. The students also undertake other non-credit earning courses/workshops. These include Neglected Tropical Diseases (NTD) seminars; Health Policy Analysis; Project Management; Scientific Writing Workshop; and Grant Writing Workshop. All courses are undertaken in the first academic year while research on a TDR-focused area is conducted (typically in the fellows' home countries) in the last 6 months of training.

Zambia: School of Public Health, University of Zambia UNZA (Lusaka, Zambia)

At the University of Zambia, the IR training programme started in 2015 as a concentration area within the traditional master's degree programmes. The IR concentration is offered within the MSc in Epidemiology; MPH, Health Promotion and Education and MPH, Health Policy and Management degree programmes. The degree programme is for 2 years. Since its inception, the programme at UNZA has trained 30 candidates, 11 of whom are Zambian and 19 international students from other LMICs. Over 30 faculties contribute to the training of the students with six of them teaching IR-focused courses. All the students undertake the Introduction to Implementation Research courses integrated in the research methods course in term 1. In term 2, students undertake a comprehensive 2-week module in implementation research. In addition, students are expected to complete four core courses including: Introduction to Epidemiology; Introduction to Biostatistics; Fundamentals of Public Health and Research Methods and Development and two other electives which are based on the student's area of specialty, for instance, a student undertaking an MPH in Health Promotion and Education with implementation research will undertake two more health promotion and education specific courses. Students also undertake other non-credit earning courses/workshops, which include Scientific Writing Workshop and Monitoring and Evaluation courses.

themes, including contexts and ethical considerations (competencies 1.1–1.5 and 5.1–5.6).

The 16-item self-assessment scale for IR self-efficacy evaluates the level of confidence respondents have in applying the knowledge across the six IR themes described in the knowledge scale. Furthermore, we added 10 questions to assess participants' prior experience in applying or training others to enact IR competencies similar to Moore *et al*, and seven questions on participants' satisfaction with and perception of quality of the index IR course offered during the study period at their institution. Items in both scales were structured along the recommendation by DeVellis *et al*,[37] including examining the readability of items, avoiding double-barrel and double negative questions. Responses to statements assessing knowledge, self-efficacy and perception of quality with the index IR course were recorded on a 7-point Likert scale with 1 being 'strongly disagree' and 7 being 'strongly agree'. Negatively worded questions were reversed in numeric value, so the number 7 consistently reflected positive attitudes. Responses to experience in applying the competencies were recorded using binary 'yes' or 'no' options. See online supplemental appendix 2 for the *self-assessment questionnaire,* which include scales for IR knowledge and IR self-efficacy.

In addition, we prepared an *objective-assessment questionnaire* comprising 40 true–false statements (online supplemental appendix 3) assessing individual knowledge across the six IR themes based on questions derived from generalisable principles, concepts and methodologies in IR in published and highly cited IR literature (see Bibliography in online supplemental appendix 3). Both the self-assessment and objective-assessment questions were structured to ensure easy readability.

Data collection

The self-assessment and objective-assessment questionnaires were administered to students enrolled in the index IR courses offered during the pilot study period (January–July 2020). At the beginning of the index IR course, a short informational session about the assessment was conducted by a locally based study personnel in each institution. The objective of the assessment—to develop tools and approaches for assessing performance of IR trainings (and not to conduct individual evaluation of a particular student's performance)—were clarified during the informational session. The questionnaires were hosted on Google Forms, a link to the forms was distributed to all students attending the course—and they were invited to participate in the assessment on a voluntary basis. They could choose to participate or not participate in the assessment. Each student was assigned an anonymous identification number to complete the forms. Both forms were administered before (baseline; no earlier than 1 week prior) and after (end-line; no later than 1 week after) the course. The students were also asked about IR coursework they had completed before the index IR course, and the competencies covered by the index IR course (online supplemental table 1). Both questionnaires were available in English and Spanish. Where possible, the tools were also administered to other individuals at the school who did not take the index IR course but were interested in IR. These individuals were referred by their degree programme coordinator to the study personnel, who provided them with the informational session and a link to the survey to complete anonymously on a voluntary basis. For these individuals, the tools were administered at one time point only.

The respondents were not all at the same level (with respect to how far along they are in their degree programmes) at the different institutions. This does not constitute a limitation to the study given our goal of developing tools for assessing IR training performance as opposed to conducting an actual assessment of the performance of trainees in the different IR programmes.

Data analyses

Self-assessment tool

Baseline data, including data that were collected at one time point only, was used to determine the construct validity and reliability of the self-assessment scales using exploratory factor analyses. First, the baseline data were cleaned and assessed for factorability (ie, to confirm the data is suitable for factor analysis) based on the Bartlett test and sample adequacy based on the Kaiser-Meyer-Olkin test. Second, the correlation matrix for the 16 items measuring IR knowledge and 16 items measuring IR self-efficacy were independently assessed. Third, exploratory factor analyses were conducted using principal component analyses (PCA) method to assess the construct validity of the IR knowledge and self-efficacy scales. If a scale accurately measures what it intends to (ie, if a scale is valid), the number of principal components explaining most of the variance in the PCA should correspond to the number of constructs that the scale intends to measure.[39] A measurement scale with high construct validity will also have items with high factor loadings for the underlying construct (eg, factor loading >0.4–0.5).[40 41] Fourth, we estimated Cronbach's alpha, that is, a measure of internal consistency, to assess the reliability of the items measuring IR knowledge and self-efficacy. Cronbach's alpha (α) assesses the degree to which separate items on a scale relate to each other, if the scale is reliable (ie, the scale can consistently produce the same result when used under the same circumstances), its items will be strongly related to each other (ie, α>0.7–0.8).[42]

Objective-assessment tool

The objective assessment questionnaire comprising 40 true/false items was analysed using IRT methods. IRT allows us to establish a link between the properties of the items on a measurement scale and an underlying unobservable trait being measured by the scale,[43] that is, as one's level of the underlying trait increases, the probability of a correct response on an item also increases. Assuming that a single latent trait—IR competence—is sufficient to explain a person's response behaviour on the 40 objective items (ie, a unidimensional space) and that a person's responses to an item are independent of his or her responses to other items, we can estimate a one parameter IRT model to describe the difficulty of each of the items independent of the test and any group characteristics (eg, country, institution, breadth and depth of IR curriculum, and IR experience) as a measure of performance and to infer on the structural validity of the objective assessment questionnaire. Using a 1-Parameter Logistic model, we modelled the difficulty of each of the 40 items as a function of the IR competence of the study population while fixing the item discrimination (ie, the ability of each item to discriminate individuals with different levels of IR competence) for data collected before and after the index IR course. We expect if the objective tool is structurally valid, the distribution of the item difficulty should remain the same before and after the index IR course (ie, the same set of items should be identified as having higher difficulty) for the same population, though the fixed item discrimination may change due to the impact of the index IR course on the students.

Qualitative methodology

Data collection and analyses

We conducted key informant interviews (KII) with purposively selected students as well as 1–2 course instructors of the index IR course in each institution. We designed a KII guide (online supplemental appendix 4) to collect information on the usefulness of the core competency framework and assessment questionnaires and perceived facilitators and barriers to the effective delivery of a specific IR course. Three categories of students were selected: those who displayed high knowledge and self-efficacy, average knowledge and self-efficacy and low knowledge and self-efficacy based on results from the quantitative surveys. Each interview was conducted face-to-face in a private office/classroom or via a secured online platform. Verbal informed consent was obtained from each respondent prior to the interview, and participation in the interviews was voluntary. No financial incentives were given to participants. One locally based study personnel was trained to conduct the interviews with the application of the interview guide in each institution. The interviews were conducted in English and were audio-recorded with detailed notes taken. Deductive thematic analyses were performed on the interview data to explore respondents' perception of the relevance of the IR framework for training; ease of using the framework and tools for developing course content; knowledge and confidence of respondents to do IR and how the framework and tools help identify training aspects that need improvement.

Combining quantitative and qualitative results: performance of IR trainees

We estimated the performance score of the respondents on the objective-assessment and established the Pearson correlation of these scores with their self-assessment knowledge score at baseline and endline. We estimated and compared the frequency and proportion of respondents who self-reported high knowledge and high self-efficacy (Likert scale response '5 and above') for the different IR competencies as measured by the self-assessment questionnaire before and after the index IR course in each institution/programme. We also estimated the frequency of those who were satisfied with various aspects of the index IR course and used the KII data to reflect on the performance of the IR training and index

Table 1 Number of respondents from each country

Country	University	Self-assessment		Objective assessment		Self-assessment (one time point only)	Objective-assessment (one time point only)	Key Informant Interview (KII)
		Baseline	Endline	Baseline	Endline			
Colombia	Universidad de Antioquia (UdeA)—Héctor Abad Gómez National School of Public Health	8	9	6	5	16	16	–
Ghana P1	University of Ghana—The School of Public Health	18	17	18	17	–	–	6
Ghana P2	University of Ghana—The School of Public Health	19	19	19	19	–	–	1
Indonesia	Universitas Gadjah Mada, Yogyakarta—The Faculty of Medicine, Public Health and Nursing	18	18	17	16	16	13	8
South Africa	University of the Witwatersrand—School of Public Health	–	–	–	–	14	14	5
Zambia	University of Zambia—School of Public Health	53	45	30	38	–	–	5
Total		**116**	**108**	**90**	**95**	**46**	**43**	**25**

course and to understand the facilitators and barriers to implementing an effective IR training programme at the different institutions.

All analyses were completed by pooling data across all institutions. Data from the self-assessment and objective-assessment questionnaire were analysed using STATA statistical package, while the qualitative data were organised in Microsoft Excel.

The study was approved by the Institutional Review Board of the Johns Hopkins Bloomberg School of Public health and was deemed to be 'non-human subjects research' as no personal health information or any identifiers were collected from the participants (IRB No: 00011320).

Patient and public involvement

It was not appropriate to involve patients or the public in the design, or conduct or reporting of this study.

RESULTS

Study population

Table 1 shows the number of respondents included for the self-assessment, objective assessment and KIIs from each institution and country.

A total of 166 unique individuals participated in the survey across the five institutions. There were two major IR programmes at the University of Ghana, one was focused on IR for infectious diseases of poverty (Ghana P1) and the other was focused on IR for sexual and reproductive health and rights (Ghana P2). Out of the 166 unique individuals, 162 completed the self-assessment and 133 completed the objective-assessment at baseline or over one time point. A total of 108 completed the self-assessment and 90 completed the objective assessment at two time points (ie, before and after the index IR course). KIIs were conducted with 25 individuals, 20 of these were students participating in the index IR courses and five were faculty teaching the courses.

The COVID-19 lockdown differentially impacted our access to respondents in the different countries. In some places, schools were closed and scheduled IR courses were cancelled or moved online, which disrupted data collection activities. For example, in Colombia, the index IR course was moved online, limiting our access to respondents for the KIIs. In South Africa, we were only able to complete the survey for the self-assessment and objective-assessment at one time point (ie, prior the index IR course). In Ghana and Zambia, there were no other individuals identified at the schools who did not take the index IR course and were interested in IR.

Construct validity and reliability of the self-assessment questionnaires

For the pooled baseline dataset (n=162), results of Bartlett's test of sphericity, p<0.001, suggested the two scales for self-assessment of IR knowledge and self-efficacy were correlated and suitable for factor analysis. The KMO test was 0.929 (ie, >0.6) suggesting an adequate sample size for factor analysis. Our initial PCA showed that one component explained 86% of variance in the scale for measuring IR knowledge, and one component had eigenvalue >1. The scree plot and parallel analysis suggested one factor as well. The construct validity of the scale was confirmed by high average factor loading of items ranging from 0.65 to 0.87 and communality greater than 0.3 (0.6–0.82).[41] The final factor solution showed adequate reliability with Cronbach's alpha (α)=0.97. Similarly, for the scale for measuring IR self-efficacy, one component explained 86% of variance, with an eigenvalue >1, factor loadings from 0.77 to 0.91, communality greater than 0.3 (0.60–0.82) and Cronbach's alpha (α)=0.98.

Box 3 Objective questions with the highest difficulty characteristics

⇒ Q7, theme 2: psychological theories such as theory of planned behaviour are not useful in implementation research.
⇒ Q2, theme 4: implementation research teams should include only researchers with expertise in quantitative and qualitative research methods.
⇒ Q2, theme 3: identification of strategies for facilitating implementation of evidence-supported interventions and conduct of implementation research should be conducted independent of the stakeholders' consultation process.

Item characteristics for the objective assessment questionnaire

Based on data collected before and after the index IR course, question #7 under theme 2 on scientific inquiry, question #2 under theme 4 on resources for IR and question #2 under theme 3 on implementation strategies in the objective assessment tool (box 3) had the highest difficulty characteristics before the course (b=2.54, 2.36 and 1.12, respectively) and after the course (b=1.64, 1.42 and 0.79 respectively) (online supplemental figure 1). These questions were consistent with respect to their difficulty characteristics.

Furthermore, the overall item discrimination characteristic (ie, the ability of the items to discriminate individuals with different levels of IR competence) improved slightly after the course (ie, discrimination parameter, a=0.90 before the course and a=1.07 after the course), which would be expected if the index IR course was indeed improving the participants knowledge and self-efficacy.

Correlation of scores from the self-assessment and objective-assessment questionnaires

Table 2 shows the overall Pearson's correlation coefficient (r) between self-assessed and objectively assessed IR knowledge before the index IR course. This was largely positive but low at 0.27 (p value: <0.01) (r <0.35 is generally regarded as low, 0.36–0.67 moderate, 0.68 to 1.0 as high)[44]—which may suggest that the students' self-report of their IR knowledge is not strongly correlated with their performance on the objective assessment. This may be an indication of overinflation of students' self-assessment regarding their ability to enact a specific competency, which is a major drawback of the self-assessment method.[33]

Observing the breakdown of the correlation coefficient by institutions/countries, and for competencies grouped by themes, we see there were negative correlation coefficients between the self-assessment and objective-assessment of IR knowledge in Indonesia and South Africa, which suggests the students' self-report of their own IR knowledge in these institutions goes in the opposite direction of their objective performance.

Following the index IR course, the Pearson's correlation between the self-assessment and objective score increased overall, r=0.43 (p value: <0.01), and there was no negative correlation in the country-specific analyses (for countries where we were able to collect data after an index IR course) (table 3). This may be taken as an indication of overall improvements in learner's self-assessment following the index IR course.

IR competency-based assessment of trainees at baseline

At baseline, over 50% of respondents self-reported a high level of knowledge in IR (Likert scale 5 and above) and were confident in their ability to enact all the IR

Table 2 Correlation coefficient comparing self-assessed and objectively assessed IR knowledge, overall and by institution/country—before index IR course

	Sites (P value)					
	All sites	Colombia	Ghana	Indonesia	South Africa	Zambia
Overall	0.268 (0.005)	0.495 (0.010)	0.284 (0.088)	−0.118 (0.507)	−0.464 (0.095)	0.222 (0.104)
Theme 1: Working w/ stakeholders	0.214 (0.006)	0.476 (0.014)	0.032 (0.853)	−0.095 (0.595)	−0.415 (0.139)	0.123 (0.373)
Theme 2: Scientific inquiry	0.281 (0.003)	0.473 (0.015)	0.519 (0.001)	−0.048 (0.788)	−0.143 (0.625)	0.213 (0.119)
Theme 3: Implementation strategies	0.213 (0.006)	0.403 (0.041)	0.166 (0.327)	−0.129 (0.468)	−0.238 (0.413)	0.298 (0.027)
Theme 4: Resources for IR	0.178 (0.022)	0.405 (0.040)	−0.229 (0.172)	0.010 (0.955)	−0.235 (0.418)	0.203 (0.136)
Theme 5: Communication and advocacy	0.244 (0.002)	0.438 (0.025)	0.161 (0.342)	−0.159 (0.370)	−0.195 (0.504	0.292 (0.031)
Theme 6: Context and ethics	0.277 (0.003)	0.480 (0.013)	0.207 (0.219)	−0.138 (0.438)	0.218 (0.453)	0.238 (0.079)

IR, implementation research.

Table 3 Correlation coefficient comparing self-assessed and objectively assessed IR knowledge, overall and by institution/country—after index IR course*

	Sites (P value)			
	All sites	Ghana	Indonesia	Zambia
Overall	0.428 (0.000)	0.496 (0.002)	0.249 (0.319)	0.470 (0.000)
Theme 1: Working w/ stakeholders	0.341 (0.001)	0.237 (0.157)	0.087 (0.731)	0.409 (0.002)
Theme 2: Scientific inquiry	0.389 (0.000)	0.245 (0.143)	0.048 (0.849)	0.472 (0.000)
Theme 3: Implementation strategies	0.413 (0.001)	0.428 (0.008)	0.438 (0.069)	0.414 (0.001)
Theme 4: Resources	0.416 (0.000)	0.530 (0.001)	0.194 (0.439)	0.433 (0.001)
Theme 5: Communication and advocacy	0.302 (0.000)	0.219 (0.193)	0.089 (0.726)	0.344 (0.010)
Theme 6: Context and ethics	0.366 (0.000)	0.186 (0.271)	0.143 (0.570)	0.392 (0.003)

*Columbia was not included because no assessment was done after the index IR course due to COVID-19 lockdown.
IR, implementation research.

competencies, that is, self-efficacy (table 4), except for competencies related to *conducting IR in a robust and rigorous manner* (competencies 8.1–8.5 in online supplemental appendix 1).

At baseline, the mean raw score on the objective-assessment was 65.26% (95% CI 63.6% to 66.9%) and the median was 65% (95% CI 62.5% to 67.5%) for trainees without any missing information (n=133).

Online supplemental tables 2 and 3 show a breakdown of these self-assessment and objective-assessment scores by institutions and countries and can serve as an initial guide for countries/institutions in identifying IR competency areas that need to be strengthened.

Over 50% of participants reported *using IR theories, methods and tools in a new or existing project*, and less than 50% reported *using systematic approaches to work with different stakeholders or to disseminate and communicate research findings from IR projects* (table 5). Broadly, most participants had limited experience on training others in IR across all countries (which is expected given that most respondents were fresh postgraduate students with limited field experience).

Overall, most participants were satisfied with the index IR course and thought the course was implemented with high quality.

Usefulness and relevance of the IR competency framework and assessment tools

All respondents from the KIIs reported that the framework and assessment tools were valid for assessing trainees' performance in IR. Many reflected specifically on the importance of stakeholders as central to IR and the mediating role of context in IR. Online supplemental table 4 provides a summary of the usefulness of the framework and assessment tools to guide IR training and practice, and potential challenges based on the KIIs with the students.

Most students felt more knowledgeable about IR and more confident in their IR skills following the index IR course but wanted a mentor to help guide them through the next steps including proposal development and integration of IR in their work outside of the academic programme. Most students had plans to apply IR in their thesis and future study. However, they noted that there were limited opportunities to apply IR within their workplaces across all contexts.

Both students and faculty reflected on the value of field work for learning IR competencies, particularly those related to stakeholder engagement. Students specifically noted that opportunities to learn in the field and apply their knowledge would increase their confidence to apply IR competencies. Some also mentioned the benefits of having both an online and in-person course to reinforce concepts (ie, taking online modules first can enhance in-class discussions). The value of a dual approach is an important finding because time was consistently noted as a barrier—that is, not having enough time with the material, to apply concepts or to network and explore IR resources.

DISCUSSION

A framework of IR core competencies for teams focused in LMICs was previously developed through a modified Delphi process with global stakeholders, including trainers and trainees involved in IR training in LMICs.[4] To operationalise the framework and validate it for evaluating IR training, we developed self-assessment and objective-assessment tools for assessing IR knowledge and self-efficacy based on the framework. In this paper, we showed that the two 16-item self-assessment scales developed for assessing IR knowledge and self-efficacy have high internal consistency and construct validity, and item characteristics of the 40-item objective assessment tool were consistent for assessing IR knowledge described by the framework. The KII responses suggested that both the framework and assessment tools are valid and highly relevant for guiding IR training in LMICs and identifies important competencies for enacting IR from the perspectives of IR trainers and trainees in LMICs, for example, the salience of stakeholders' engagement and contexts in IR.

However, the KII also pointed out important challenges in operationalising the framework, for example, extent of time needed for building IR competencies, which requires both practice and research, and limited opportunities for enacting the competencies in practice, which suggests that to do IR in LMICs, proficiency in both research-based IR competencies (eg, conducting IR in a rigorous and robust manner) and practice-based

Table 4 Self-assessment before index IR course, overall

Items	Self-assessment of IR knowledge (N=162) n (%)	Self-assessment of IR self-efficacy (N=162) n (%)
Theme 1: Working with stakeholders		
Identifying relevant stakeholders for the implementation of evidence-supported interventions and IR. (*Competencies 3.1–3.7*)	114 (70.4)	106 (65.4)
Engaging relevant stakeholders for the implementation of evidence-supported interventions and IR. (*Competencies 3.1–3.7*)	108 (66.7)	103 (63.6)
Theme 2: Scientific inquiry		
Formulating appropriate IR questions. (*Competencies 6.1–6.8*)	102 (62.9)	100 (61.7)
Determining applicable measures (or variables) for conducting IR. (*Competencies 7.1–7.5*)	98 (60.5)	92 (56.8)
Determining applicable study designs and methods for conducting IR. (*Competencies 7.1–7.5*)	95 (58.6)	94 (58.0)
Conducting IR in a robust and rigorous manner. (*Competencies 8.1–8.5*)	**75 (46.3)**	**71 (43.8)**
Theme 3: Implementation strategies		
Synthesising evidence to support implementation of a given intervention(s). (*Competencies 2.1–2.3*)	92 (56.8)	86 (53.1)
Analysing facilitators and barriers to the implementation of evidence-supported interventions. (*Competencies 2.1–2.3*)	106 (65.4)	92 (56.8)
Developing implementation strategies to address barriers to implementation of evidence-supported interventions and IR. (*Competencies 2.1–2.3*)	88 (54.3)	84 (51.9)
Analysing implementation strategies. (*Competencies 2.1–2.3*)	90 (55.5)	83 (51.2)
Theme 4: Resources for IR		
Building an IR team. (*Competencies 4.1–4.4*)	84 (51.9)	88 (54.3)
Leveraging required resources for conducting IR. (*Competencies 9.1–9.4*)	85 (52.5)	80 (49.4)
Theme 5: Communication and advocacy		
How to use information from IR. (*Competencies 10.1–10.7*)	102 (62.9)	100 (61.7)
Communicating and advocating effectively throughout the IR process. (*Competencies 11.1–11.5*)	100 (61.7)	98 (60.5)
Theme 6: Context and ethics		
Analysing contexts (health systems, implementation organisation and community) affecting the implementation of evidence-supported interventions (*Competencies 1.1–1.5*)	98 (60.5)	90 (55.5)
Applying ethical principles in conducting IR (*Competencies 5.1–5.6*)	114 (70.4)	111 (68.5)

IR, implementation research.

IR competencies (eg, engaging and partnering with different stakeholders) are needed.[2 45] Indeed, it is nearly impossible to sufficiently train individuals and/or teams on all the competencies needed to carry out IR through a single training programme given the need to acquire some of these competencies through practice.[11] Hence, trainees should be encouraged to undergo self-study and systematically identify opportunities for learning and/or applying IR competencies through different on-the-job experiences.[11] Additionally, lifelong exchanges and collaboration between trainers and well-situated trainees in different professions should be encouraged.

The self-assessment tools described in this study were adapted from ones described by Moore *et al* to assess training in a knowledge translation programme.[10] However, the tools in this study differ in that in addition to mapping to core competencies for IR, they were also assessed for reliability and validity.[12–16]

This study is also the first time an objective-assessment has been used to accompany self-assessments of IR competencies or training programmes and draws on both classical test theory and IRT in generating a valid approach for assessing IR competencies in LMICs.[41 43] We assume if the objective questionnaire is structurally valid,

Table 5 IR activities that participants were involved with at the time of data collection

	All (N=158)* n (%)
Applying IR	
Using IR theories, models, and frameworks in new or existing project	80 (50.6)
Using IR methods and study designs in new or existing project	83 (52.5)
Using systematic approaches to work with different stakeholders	73 (46.2)
Using systematic approaches in IR to disseminate and communicate research findings with different stakeholders	70 (44.6)
Using IR tools in your implementation project	93 (59.6)
Training others in IR	
Sharing IR course materials within your networks	75 (47.8)
Training others on how to apply IR theories, models and frameworks	48 (30.8)
Training others on how to apply IR methods and study designs	49 (31.0)
Training others on using systematic approaches to work with different stakeholders	45 (28.7)
Training others on using systematic approaches in IR to disseminate and communicate research findings with different stakeholders	47 (30.0)
Satisfaction and perception of quality with index IR course	(N=140)† n (%)
I was extremely satisfied with the session readings and resources	97 (69.3)
Overall, I was satisfied with the presentations	100 (71.4)
Overall, I was satisfied with how the content applies to my work	99 (70.7)
I was satisfied with the session activities	100 (71.4)
I was satisfied with the format of the session (presentation, group activity, etc.)	101 (72.1)
I was satisfied with content of the sessions	103 (73.6)
Overall, I thought the course was implemented with high quality	102 (72.9)

*There were 166 in all. However, information was missing on the current IR activities for eight individuals (5%). Hence, the total sample size for this table n=158.
†This includes only those who participated in the index IR course.
IR, implementation research.

then the distribution of the item difficulty from a one-parameter IRT model will be the same for the same population before and after the index IR courses as observed by our results.[43] The consistency regarding the item difficulty characteristic identified by the tool before and after the course may suggest that the objective-assessment tool is valid and reliable in assessing IR competence, thus reflecting on the validity of the underlying IR competency framework.

It is important that assessment of IR competencies does not rely on only self-assessment given the low correlation (and sometimes negative correlation) observed between the self-assessment and objective-assessment. Trainees tend to rate their own IR knowledge and self-efficacy highly (even before completing an index IR course) than the objective-assessment suggests due to social desirability bias given their expressed interest in IR. While the correlations between the self-assessment and objective-assessment scores improved after the index IR course, these were still low overall and for specific competency themes. We expect that the correlation coefficient between objectively assessed and self-assessed IR knowledge will be positive and should increase in absolute value after the index IR course (compared with before) as observed, assuming that the index IR course is teaching relevant IR competencies. However, the low correlations after the index IR course may suggest that while the relevant IR competencies are being covered, which may give trainees a perception that they have high knowledge in these competencies, they are not covered in depth for the students to overcome the difficulty of items examining these competencies in the objective-assessment. Future studies may further unpack these relationships.

We were able to further examine the competency gaps of the IR training programmes in the various institutions (online supplemental tables 2 and 3)—these analyses, based on the framework and tools, provide guidance for strengthening the training programmes. Other uses of the framework and assessment tools include guiding curriculum development and interdisciplinary efforts for new training and degree programmes in IR in LMICs, implementing competency-based evaluations of IR training programmes in LMICs, and IR theory building and facilitating effective IR and practice in LMICs.

This study has some limitations. First, the COVID-19 pandemic disrupted data collection, which differentially impacted some sites and could contribute to selection bias with respect to the studied sample. Furthermore, drop-off of respondents between the before and after self-assessment and between the self-assessment and objective-assessment could have introduced additional selection bias. Second, the different IR training programmes that formed the basis of the assessment were at different stages of maturity and their index IR courses, while they covered similar contents, were not the same. Hence, no comparative country/institutional performance assessment is possible from this study. Third, despite our attempt to minimise social desirability bias (with the introduction of the objective-assessment), this is still possible as we observed from the correlation study, and the extent may not be none. Fourth, while the interviews were implemented by locally based personnel proficient in both English and any other official language (to facilitate

multilingual conversation where necessary), the decision to conduct the interviews in English may have created some language barriers for non-native English speakers. Last, the conclusion from the KII data may not be transportable across different contexts outside of the included training programmes and settings.

CONCLUSION

Self-assessment and objective-assessment tools were developed based on an IR competency framework for LMICs and were shown to be reliable and valid for assessing IR competencies in LMIC settings. The IR competency framework and tools provide guidance for designing and evaluating IR training programmes in LMICs and reducing the know-do gap in implementing evidence-supported interventions to address health problems globally.

Author affiliations
[1]Sparkman Center for Global Health, The University of Alabama at Birmingham, Birmingham, Alabama, USA
[2]International Health, Johns Hopkins University Bloomberg School of Public Health, Baltimore, Maryland, USA
[3]University of the Witwatersrand, School of Public Health, Johannesburg, Gauteng, South Africa
[4]Social and Behavioural Sciences, University of Ghana School of Public Health, Accra, Ghana
[5]Faculty of Medicine, Public Health and Nursing, Universitas Gadjah Mada, Yogyakarta, Daerah Istimewa Yogyakarta, Indonesia
[6]Facultad Nacional de Salud Publica, Universidad de Antioquia, Medellin, Colombia
[7]Epidemiology and biostatistics, University of Zambia, Lusaka, Zambia, Zambia
[8]University of Ghana, Legon, Greater Accra, Ghana
[9]Special Programme for Research & Training in Tropical Diseases (TDR), WHO, Geneva, Switzerland

Acknowledgements We thank Dr Anna Thorson from TDR and Dr Vanessa Brizuela from UNDP/UNFPA/UNICEF/WHO/World Bank Special Programme of Research, Development and Research Training in Human Reproduction (HRP), Department of Sexual and Reproductive Health and Research, WHO for their valuable feedback in the design of the study.

Contributors OA conceptualised the study, was the principal author, contributed to the data collection and analysis and managed all revisions. AR and AK contributed to the data collection, analysis, and critical revisions of the paper. LI, PD-G, YM, CAR, CJ, KT and ERGM contributed to critical revisions of the paper. MV and PL contributed to the data interpretation and critical revisions of the paper. OA acts as the guarantor for the overall content of this work.

Funding The Johns Hopkins University received funding from TDR, the Special Program for Research and Training in Tropical Diseases and HRP, the Special Programme of Research, Development, and Research Training in Human Reproduction, to identify gaps in IR at the universities supported by each of these two programmes. This manuscript describes the work mainly involving the universities supported by TDR, except the University of Ghana where principal investigators supported by both TDR and HRP are listed as co-authors. TDR acknowledges the commitment and support of its funders, including long-term core contributors from national governments and international institutions, as well as designated funding for specific projects within our current priorities. For the full list of TDR donors, please visit the website at https://www.who.int/tdr/about/funding/en. The funding source had no role in the design of this study or in the decision to submit the manuscript for publication

Competing interests None declared.

Patient and public involvement Patients and/or the public were not involved in the design, or conduct, or reporting, or dissemination plans of this research.

Patient consent for publication Not applicable.

Provenance and peer review Not commissioned; externally peer-reviewed.

Data availability statement Data are available upon reasonable request.

ORCID iDs
Olakunle Alonge http://orcid.org/0000-0001-7642-2806
Anna Kalbarczyk http://orcid.org/0000-0002-6143-8634
Phyllis Dako-Gyeke http://orcid.org/0000-0002-4632-1833

REFERENCES

1 Theobald S, Brandes N, Gyapong M, *et al*. Implementation research: new imperatives and opportunities in global health. *Lancet* 2018;392:2214–28.
2 Alonge O, Rodriguez DC, Brandes N, *et al*. How is implementation research applied to advance health in low-income and middle-income countries. *BMJ Glob Health* 2019;4:e001257.
3 Brownson RC, Kumanyika SK, Kreuter MW, *et al*. Implementation science should give higher priority to health equity. *Impl Sci* 2021;16:28.
4 Alonge O, Rao A, Kalbarczyk A, *et al*. Developing a framework of core competencies in implementation research for low/middle-income countries. *BMJ Glob Health* 2019;4:e001747.
5 Leung WC. Competency based medical training. *BMJ: Br Med J* 2002;325:7366.
6 Gruppen LD, Mangrulkar RS, Kolars JC. The promise of competency-based education in the health professions for improving global health. *Hum Resour Health* 2012;10:1–7.
7 Bhandari S, Wahl B, Bennett S, *et al*. Identifying core competencies for practicing public health professionals: results from a delphi exercise in Uttar Pradesh, India. *BMC Public Health* 2020;20:1737.
8 Provvidenza C, Townley A, Wincentak J, *et al*. Building knowledge translation competency in a community-based hospital: a practice-informed curriculum for healthcare providers, researchers, and leadership. *Impl Sci* 2020;15.
9 Sistermans IJ. Integrating competency-based education with a case-based or problem-based learning approach in online health sciences. *Asia Pac Educ Rev* 2020;21:683–96.
10 Moore JE, Rashid S, Park JS, *et al*. Longitudinal evaluation of a course to build core competencies in implementation practice. *Impl Sci* 2018;13.
11 Schultes M-T, Aijaz M, Klug J, *et al*. Competences for implementation science: what trainees need to learn and where they learn it. *Adv Health Sci Educ Theory Pract* 2021;26:19–35.
12 Carlfjord S, Roback K, Nilsen P. Five years' experience of an annual course on implementation science: an evaluation among course participants. *Impl Sci* 2017;12.
13 Ullrich C, Mahler C, Forstner J, *et al*. Teaching implementation science in a new master of science program in Germany: a survey of stakeholder expectations. *Impl Sci* 2017;12.
14 Proctor EK, Landsverk J, Baumann AA, *et al*. The implementation research Institute: training mental health implementation researchers in the United States. *Impl Sci* 2013;8.
15 Meissner HI, Glasgow RE, Vinson CA, *et al*. The US training Institute for dissemination and implementation research in health. *Impl Sci* 2013;8:12.
16 Norton WE. On academics: advancing the science and practice of dissemination and implementation in health: a novel course for public health students and academic researchers. *Pub Health Rep* 2014;129:536–42.
17 Peters DH, Adam T, Alonge O, *et al*. Implementation research: what it is and how to do it. *BMJ* 2013;347.

18 Adesope OO, Trevisan DA, Sundararajan N. Rethinking the use of tests: a meta-analysis of practice testing. *Rev Educ Res* 2017;87:659–701.
19 Roediger HL, Karpicke JD. Test-enhanced learning taking memory tests improves long-term retention. *Psychol Sci* 2006;17:249–55.
20 Vaughn KE, Rawson KA, Pyc MA. Repeated retrieval practice and item difficulty: does criterion learning eliminate item difficulty effects? Psychon. *Psychon Bull Rev* 2013;20:1239–45.
21 Carpenter SK, Pashler H, Wixted JT, *et al*. The effects of tests on learning and forgetting. *Mem Cognit* 2008;36:438–48.
22 McDaniel MA, Wildman KM, Anderson JL. Using quizzes to enhance summative-assessment performance in a web-based class: an experimental study. *J Appl Res Mem Cogn* 2012;1:18–26.
23 Roediger HL, Agarwal PK, McDaniel MA, *et al*. Test-enhanced learning in the classroom: long-term improvements from quizzing. *J Exp Psychol Appl* 2011;17:382–95.
24 Karpicke JD, Aue WR. The testing effect is alive and well with complex materials. *Educ Psychol Rev* 2015;27:317–26.
25 Carpenter SK, Toftness AR. The effect of pre-questions on learning from video presentations. *J Appl Res Mem Cogn* 2017;6:104–9.
26 Hartley J. The effect of pre-testing on post-test performance. *Instr Sci* 1973;2:193–214.
27 Little JL, Bjork EL. Multiple-choice pretesting potentiates learning of related information. *Mem Cogn* 2016;44:1085–101.
28 Peeck J. Effect of pre-questions on delayed retention of prose material. *J Educ Psychol* 1970;61:241–6.
29 Richland LE, Kornell N, Kao LS. The pretesting effect: do unsuccessful retrieval attempts enhance learning. *J Exp Psychol Appl* 2009;15:243–57.
30 Latimier A, Riegert A, Peyre H, *et al*. Does pre-testing promote better retention than post-testing. *NPJ Sci Learn* 2019;4:15.
31 Gagné RM, Briggs LJ, Wager WW. *Principles of instructional design*. 4th edn. Forth Worth, TX: Harcourt Brace Jovanovich College Publishers, 2002.
32 Reteguiz JA. Relationship between anxiety and standardized patient test performance in the medicine clerkship. *J Gen Intern Med* 2006;21:415–8.
33 Ross JA. The reliability, validity, and utility of self-assessment. *Pract Assess Res Eval* 2006;11:10.
34 LaDonna KA, Hatala R, Lingard L, *et al*. Staging a performance: learners' perceptions about direct observation during residency. *Med Educ* 2017;51:498–510.
35 McCambridge J, Witton J, Elbourne DR. Systematic review of the hawthorne effect: new concepts are needed to study research participation effects. *J Clin Epidemiol* 2014;67:267–77.
36 Lockyer J, Sargeant J. Multisource feedback: an overview of its use and application as a formative assessment. *Can Med Educ J* 2022;13:30–5.
37 DeVellis RF, Thorpe CT. *Scale development: theory and applications*. Sage publications, 2021.
38 Lenzner T, Kaczmirek L, Lenzner A. Cognitive burden of survey questions and response times: a psycholinguistic experiment. *Appl Cogn Psychol* 2010;24:1003–20.
39 Floyd FJ, Widaman KF. Factor analysis in the development and refinement of clinical assessment instruments. *Psychol Assess* 1995;7:286–99.
40 Clark LA, Watson D. Constructing validity: basic issues in objective scale development. In: Kazdin AE, ed. *Methodological issues and strategies in clinical research*. American Psychological Association, 2016: 187–203.
41 Boateng GO, Neilands TB, Frongillo EA, *et al*. Best practices for developing and validating scales for health, social, and behavioral research: a primer. *Front Public Health* 2018;6:149.
42 Tavakol M, Dennick R. Making sense of cronbach's alpha. *Int J Med Educ* 2011;2:53–5.
43 Hambleton RK, Jones RW. Comparison of classical test theory and item response theory and their applications to test development. *Educ meas: issues pract* 1993;12:38–47.
44 Taylor R. Interpretation of the correlation coefficient: a basic review. *J Diagn Med Sonogr* 1990;6:35–9.
45 Ramaswamy R, Mosnier J, Reed K, *et al*. Building capacity for public health 3.0: introducing implementation science into an MPH curriculum. *Impl Sci* 2019;14:18.