

BMJ Open Measuring the outcomes of volunteering for education: development and pilot of a tool to assess healthcare professionals' personal and professional development from international volunteering

Natasha Tyler,¹ Carlos Collares,² Ged Byrne,³ Lucie Byrne-Davis⁴

To cite: Tyler N, Collares C, Byrne G, *et al.* Measuring the outcomes of volunteering for education: development and pilot of a tool to assess healthcare professionals' personal and professional development from international volunteering. *BMJ Open* 2019;**9**:e028206. doi:10.1136/bmjopen-2018-028206

► Prepublication history and additional material for this paper are available online. To view these files, please visit the journal online (<http://dx.doi.org/10.1136/bmjopen-2018-028206>).

Received 27 November 2018

Revised 14 June 2019

Accepted 18 June 2019



© Author(s) (or their employer(s)) 2019. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

¹Greater Manchester Patient Safety Translational Research Centre, University of Manchester, Manchester, UK

²School of Health Professions Education, Maastricht University, Maastricht, The Netherlands

³Health Education England, Manchester, UK

⁴The University of Manchester, Manchester, UK

Correspondence to

Dr Natasha Tyler;
natasha.tyler@manchester.ac.uk

ABSTRACT

Objective The development and pilot of a self-report questionnaire, to assess personal and professional development of healthcare professionals gained through experiences in low-income and middle-income countries.

Design The instrument was developed from a core set of the outcomes of international placements for UK healthcare professionals. Principal component analysis and multidimensional item response theory were conducted using results of a cross-sectional pilot study to highlight items with the best psychometric properties.

Setting Questionnaires were completed both online and in multiple UK healthcare professional events face-to-face.

Participants 436 healthcare professional participants from the UK (with and without international experience) completed a 110-item questionnaire in which they assessed their knowledge, skills and attitudes.

Measures The 110-item questionnaire included self-report questions on a 7-point Likert scale of agreement, developed from the core outcome set, including items on satisfaction, clinical skills, communication and other important healthcare professional knowledge, skills, attitudes and behaviours. Item reduction led to development of the 40-item Measuring the Outcomes of Volunteering for Education-Tool. Internal consistency was evaluated by the Cronbach's α coefficient. Exploratory analysis investigated the structure of the data using principal component analysis and multivariate item response theory.

Results Exploratory analysis found 10 principal components that explained 71.80% of the variance. Components were labelled 'attitude to work, adaptability, adapting communication, cultural sensitivity, difficult communication, confidence, teaching, management, behaviour change and life satisfaction'. Internal consistency was acceptable for the identified components ($\alpha=0.72-0.86$).

Conclusions A 40-item self-report questionnaire developed from a core outcome set for personal and professional development from international placements was developed, with evidence of good reliability and validity. This questionnaire will increase understanding of impact of international placements, facilitating comparisons of different types of experience. This will aid decision making about whether UK healthcare

Strengths and limitations of this study

- The Measuring the Outcomes of Volunteering for Education-Tool (MOVE-iT) was developed based on evidence from peer-reviewed literature and expert opinion.
- The underlying structures of the instrument were explored using a large dataset of 436 multidisciplinary healthcare professionals.
- This tool provides a way of evidencing benefits; however, there is a body of critical evidence outlining the ethical concerns of medical practice abroad, particularly when individuals practice in ways that they might not in a high-income country.

professionals should be encouraged to volunteer internationally and in what capacity.

BACKGROUND

Globalisation of the health workforce has inevitably led to large numbers of qualified healthcare professionals choosing to temporarily (ranging from 1 day to 2 years) work overseas in some capacity, with many choosing low-income and middle-income countries (LMICs).¹ In this paper, we describe international placements in any LMIC (as defined by the Organisation for Economic Cooperation and Development) in which the healthcare professional receives little or no remuneration; this is often referred to as volunteering. Such placements can take numerous forms, for example, a dentist delivering a service on a hospital train in India,² British healthcare professionals of many cadres working together in health partnerships with a hospital in Tanzania³ or healthcare scientists working in labs in sub-Saharan Africa.⁴

International health volunteering has been reported as resulting in personal and

professional development (PPD), for example, a change in attitudes on a personal level, or developing new/broadening existing professional skills, see our previous work for a full list of all reported PPD.⁵ Benefits have been reported for both the individual's practice and also patient outcomes on return.⁶ Many professionals report PPD outcomes as a result of the new experience and particularly that working in an LMIC encourages healthcare professionals to learn new skills in an effort to adequately adapt, for example, using new clinical techniques specific to the LMIC, or dealing with a new cultural phenomenon.⁷⁻⁹ Professionals report that LMICs provide staff with an opportunity to practice skills that they would not develop in a domestic work setting, as such giving them increased confidence in their work.^{8 10} In some academic papers, professionals report perceived/expected exposure to higher numbers of clinical cases and often clinical cases that are more challenging than those seen in high-income countries (HICs) as well as opportunities to lead, make decisions and work within new cultural and social norms.^{6 9 11 12} Many staff report a change in core attitudes or beliefs: a greater appreciation of caring, an acceptance of cultural differences or a changed/new/broader perspective.^{8 9 13 14} As a result, in the UK, some organisations have proposed that enabling and encouraging staff to work in LMICs may have great benefits to the National Health Service (NHS)^{6 7 15} and have expressed a desire to assess PPD outcomes^{16 17} to provide quantitative evidence of benefit.

Despite these reported benefits, volunteering is sometimes perceived as a loss to the HIC, for example, our research found that within the UK NHS, some management perceived volunteering as a loss of staff within a service that is already under pressure.¹⁵ As such, some employers are reluctant to release staff for international placements.¹⁵

Qualitative research into the benefits of international working or volunteering (from now on referred to as 'international placements' for ease), has reported similar PPD outcomes regardless of the host country, type of projects or individual's profession. Communication, leadership, attitude to work, flexibility and cultural awareness are frequently reported outcomes.^{2 6 9 12 18} However, from an educational perspective, precise information about this learning (process, outcomes, variables) is seldom reported. In a recent meta-synthesis and Delphi study, we reported a list of 116 outcomes⁵ from a review of literature on international placements for healthcare professionals. The list included benefits and costs that were agreed by stakeholders to be frequently experienced by healthcare professionals (of any cadre) in an international placement. Costs (eg, health outcomes, financial loss, clinical de-skilling) are not reported in this paper, but can be found in the meta-synthesis.⁵ We also summarised the moderating (factors that affect the strength of a relationship) and mediating variables (factors that explain the relationship between two items) that were reported in the

literature to potentially affect PPD outcomes (eg, length of stay, host country, level of experience, supervision).

There have been some attempts to quantify these outcomes, for example, a small number of previous UK papers have used a questionnaire approach to understand outcomes,^{8 19 20} but these have not taken a psychometric approach to the measurement of underpinning domains of learning (ie, developed and tested an evidence-based questionnaire). A number of psychometric questionnaires have been developed outside of the UK, but are based on non-domain-specific outcomes for any professional, hence are not specific to healthcare professionals.²¹⁻²³ For example, the IVIS used latent trait analysis and found 11 'volunteer outcome' factors including open-minded and intercultural relationships.²⁴ It is not known whether there are unique elements of learning or outcomes that are specific to healthcare professionals (from within the NHS) that differ from the non-domain-specific learning measured in existing tools, particularly as some of the qualitative research suggests unique outcomes, for example, related to patient interaction.^{9 25}

This study aimed to create a measure of the PPD outcomes of international placements. We worked on the large set of outcomes that stakeholders agreed were core outcomes from international placements for healthcare professionals.² We aimed to reduce the items to a short questionnaire using item response theory to establish and test a set of latent traits and their associated questions.

METHODS

Design

We followed traditional tool development methods in order to develop a measurement tool.²⁶ In summary, we took the PPD outcomes found in the previous study,²⁷ made them into questions and then reduced their number through a process of piloting with healthcare professionals and using statistical methods to eliminate items which were not congruent with other items or were redundant because they were too congruent with other items. We used a cross-sectional design, so participants were measured only at one time point. The study used item response theory, whereby 'constructs' are theoretical terms that refer to unobserved, idealised entities.²⁸ Latent traits are one type of construct, which are qualities possessed by individuals that can change, but only over the long term.²⁸ Latent traits include attitudes, preferences and dispositions, and elements that are important for professional development such as ability, expertise and aptitude.²⁹ No measure of a latent trait is ever considered perfectly accurate, instead different measures are used to estimate latent traits,³⁰ with varying levels of effectiveness.²⁸

Participants

Previous psychometric research on the sample size requirements for precise estimates of reliability

coefficients suggested we needed 400 participants.³¹ We therefore aimed to recruit the 400 participants across 4 different groups: 100 healthcare professionals who had been on international placements in the past, 100 who were about to undertake an international placement or currently working overseas, 100 with an interest in international placements but no past experience and 100 with no interest in or past experience of international placements. We included healthcare professionals who had and who had not worked internationally. It is usual to do item reduction with a sample of the population who will be using the tool. Since the tool could be used to compare PPD in healthcare professionals with or without international experience or before and after international experience, we decided to include, in the sample, healthcare professionals without international experience. We further subdivided our sample into people who were interested in international experience and not to ensure that the tool items were reduced on the basis of answers from people with all ranges of experience and perceptions of international placements. Participants were not excluded based on the years since NHS employment, provided they had this experience at some point. Inclusion criteria were that the participant be or have been an NHS employee (current or past), working/worked in a patient facing role as a qualified healthcare professional.

Procedure

Creating the pilot questionnaire

We developed a questionnaire based on the core outcome set reported in our previous paper.⁵ Two members of the team looked for common inductive, themes across the outcomes (LB-D, NT). We found experience, confidence and attitudes, where outcomes were to do with experience, we categorised them as experience and asked about the experiences they had during a suitable time period. If statements were about how confident they felt or attitudes they held, we categorised them as such and asked questions in that way. Statements were self-reported in terms of strength of agreement using a 7-point Likert scale. Where the core outcome reported in the previous paper could be interpreted in multiple ways, we referred back to the original papers where the outcome was originally reported from the metasynthesis⁵ and used this to make decisions about how to express the statement. If a statement could indicate change in experience, confidence and/or attitude, we developed questions for each.

Two members of the team (LB-D, NT) assessed each core outcome and generated 103 statements with Likert scales of agreement for each statement (from strongly disagree to strongly agree). We excluded 40 items from the core outcome set which would not be measurable through self-report questionnaires. These were items about organisational outcomes for the NHS,⁸ outcomes that were too vague to be specifically defined⁸ or overlapped in meaning with another and were combined.²⁴ For example, ‘*exposure to ethical dilemmas*’ and ‘*increased awareness of/knowledge about ethics*’ were combined into

“*I have frequently experienced ethical dilemmas*” (see online supplementary material for a record of the decisions and their reasons). In addition seven items from the Satisfaction with Life Scale were added.³²

We included 56 statements about the frequency with which the individual had an experience or exhibited certain behaviour in the last month (regardless of where this last month’s work took place). For example, “In the last month I frequently dealt with difficult people”. We generated 19 confidence statements. For example, “I am confident in my ability to teach others”. Other statements, which were more about attitudes and feelings were labelled included, for example, “I have an excellent work ethic” (n=35). Online supplementary material shows the matches between the outcomes and statements.

Pre-pilot

The questionnaire was prepiloted on 16 participants, including 7 from the MOVE research group (a group of Salford/Manchester researchers involved in similar research), to establish that the questionnaire was readable and understandable. We administered the tool online using eForms.³³ The authors, plus the wider institutional team of researchers in international placements, met face-to-face to consider all of the written comments from the pilot. We conducted a cognitive interview with four participants, using both think aloud interviewing and verbal probing, whereby participants were questioned/asked to think aloud as they completed the questionnaire.^{34 35} Any comments, issues, questions or suggestions raised during the cognitive interviews were inputted into a table, one member of the team (NT) decided how best to act on each one and whether changes needed to be made. The table was then reviewed by another team member (LB-D) and disagreements were discussed and resolved. This resulted in numerous changes being made to the statements, including using an existing Satisfaction With Life Scale (SWLS), previous research suggests using an existing validated scale if one exists and the cognitive interviews and pre-pilot process highlighted the necessity to do this.²⁶ As a result of this process, a 110-item tool was created for the pilot phase.

Pilot

There were two methods of recruitment: online and face-to-face. Face-to-face participants were recruited using an opportunistic sample at healthcare professional events nationwide (conferences, training events, exhibitions), many of which had an international focus (the majority of the sample gained this way were nurses and nursing assistants). Online participants were recruited in numerous ways, including links to the questionnaire posted on international volunteering blogs and in healthcare professional newsletters and bulletins. The majority of the online sample was gathered using snowball sampling with key contacts within companies, projects and hospital health links that place professionals internationally agreed to

send the link via email to healthcare professionals, the majority of the doctors were responded online.

The tool was completed by participants either online or face-to-face, as was convenient and appropriate for the participants. Online participants received a link in an email, blog or online community and after giving consent. Face-to-face participants completed a paper version of the questionnaire. Of the 43 organisations that helped us recruit, 9 involved face-to-face recruitment (21%). Recruitment took place between April and July 2016.

Materials

Measure

The tool consisted of 110 statements measured on a 7-point Likert scale ranging from strongly agree to strongly disagree. The Likert scale contained the following descriptors: 1 strongly agree, 2, 3, 4 neither agree nor disagree, 5, 6, 7 strongly disagree (this was reverse coded for analysis as higher intensity ordinal constructs need to be higher values, strongly agree at 7, strongly disagree at 1). No statements were reversed.

An additional existing scale was used within the tool, the SWLS.³² This is a five-item scale that has been used frequently to measure satisfaction with life. This replaced a number of statements from the core outcome set about satisfaction with life, since the questions had already been refined and tested for validity and reliability and guidelines suggest using existing scales where possible.^{26 32}

In addition to the 110 statements, participants' demographic and placement data were also gathered. Each participant was asked basic demographic questions: age, gender, profession, employment status, nationality and years since registration. Past experience on international placements was also recorded (country, length of stay).

Analyses

Principal component analysis

We used successive iterations of principal component analysis (PCA) to reduce the pool of items, so that only the items with optimal psychometric properties would remain. The PCA is a dimension-reduction tool that can be used to reduce a large set of items to a small set that still contains most of the information in the large set.³⁶ Initially, a parallel analysis was performed to determine the number of factors. Items with low communalities (<0.500) or loadings below 0.3 were withdrawn in subsequent iterations. In the final iterations, exclusions were performed at an item-by-item basis. We decided that even if there were more items in one domain we would retain them if they had adequate psychometric properties. PCA was performed in IBM SPSS V.23.³⁷

Multidimensional item response theory

We created a multidimensional item response theory (MIRT) model, based on the results of the best iteration of the PCA, in order to test the structure of the factors we found and remove any items which did not improve the assessment of each factor. MIRT is analogous to

Table 1 Participants: anticipated and actual numbers

Group	Target	N included (%)	Percentage (%) of target
Currently overseas/due to depart	100	79 (18%) (26 currently overseas, 53 due to depart)	79
Past international experience	100	169 (39%)	169
No international experience—interested	100	78 (18%)	78
No international experience—not interested	100	104 (24%)	104
Total	400	436 (100%)	109

confirmatory factor analysis (CFA)³⁸ but, unlike CFA, MIRT considers all Likert scale variables as categorical, which is more appropriate for our data. MIRT parameters in this study were estimated using weighted least squares means-adjusted and variance-adjusted, given their appropriateness for categorical variables in comparison to Bayesian estimation, which would be an operationally attractive alternative, given the high dimensionality of the data.³⁹ MIRT analysis was performed in Mplus V.8.⁴⁰

Patient and public involvement

Patients were not involved in the design and conception of this study.

RESULTS

Pilot

Participants

Four hundred and thirty-six participants completed the questionnaire, 42% (182/436) of participants had no international experience (table 1).

All participants were NHS employees (past or present). Staff group representation was largely in line with the NHS North West employee data,⁴¹ whereby 30% of the workforce is nursing and midwifery (table 2). The other staff groups were also relatively proportionate, besides Medical and Dental which represents only 9% of the North West workforce and support to staff (28%). This suggests that any item reduction based on variability in responses from the sampled group were largely representative of the NHS workforce. Table 3 shows the participant demographics.

Principal component analysis

The PCA used the correlation matrix obtained from the application of the questionnaire to the 436 participants. The 436 responses included those with no international experience to account for the range of variability in response across the NHS workforce, regardless of experience. Twenty-one iterations of the PCA were performed. From the original set of items, only 40 items were chosen for the last iteration of the PCA. The

Table 2 Professions of participants

Staff group	n	Pilot sample (%)	NHSNW (%) ⁴¹
Medical and dental	146	34	9
Nursing and midwifery	135	31	30
Allied healthcare professionals	64	15	6
Healthcare scientists	13	3	3
Ambulance	13	3	2
Support to clinical staff	30	7	28
NHS infrastructure support	5	1	18
Other scientific, therapeutic and technical	3	1	4
Other	25	6	<1

NHS, National Health Service; NHSNW, National Health Service North West.

Kaiser-Meyer-Olkin (KMO) measure showed the level of sampling adequacy to be acceptable (KMO=0.896). The lowest measure of sample adequacy for an individual item was 0.810 (“*I demonstrated I’m a good teacher*”). The Bartlett’s sphericity test indicated that the interitem correlations were sufficient for proceeding with the analysis. The lowest value for the items’ communalities was 0.590 (“*If I could live my life over, I would change almost nothing*”), which is above the aimed threshold of 0.500. After *varimax* rotation, 10 factors were extracted taking into account the findings of the scree plot and of a Monte Carlo parallel analysis. The 10 factors explained 71.80% of the variance. On the scree plot (figure 1), it is possible to observe that the first five factors had the highest eigenvalues.

Multidimensional item response theory

The diagram with the resulting model, which contains the items selected for each one of the latent variables, the

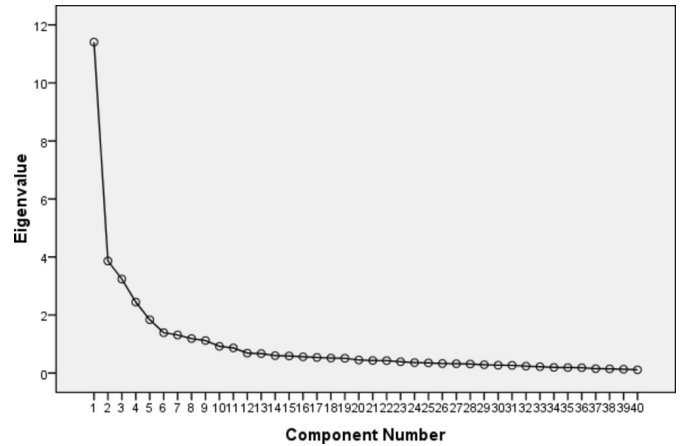


Figure 1 Scree plot.

loadings for each item and the correlation coefficients between the constructs, can be seen in figure 2. This model was chosen as it was the best possible solution to reconcile the need of creating a comprehensive, content-rich questionnaire while obtaining satisfactory evidence of validity based on its internal structure. In terms of goodness-of-fit, the model had significantly better-fit solution than a unidimensional solution in the X² test for difference testing ($\chi^2=2889.749$, $df=45$, $p<0.001$). The comparison of goodness-of-fit indices between the unidimensional solution and the proposed model can be observed in table 4. The X² is not the X² of any model but the X² of the difference of the X² of each model separately.

Reliability estimates were calculated using Cronbach’s α coefficients and using estimates of individual precision calculated based on the individual estimates of the SEs of measurement. Figure 1 shows the precision curves for each latent variable. While ‘confidence’, ‘life satisfaction’ and ‘attitude to work’ had the highest means for the individual precision estimates, ‘adaptability’ was the construct that achieved the highest precision estimates

Table 3 Participant demographic information: age, employment status, nationality, gender and career stage (years since registration was used as a proxy measure of experience)

Age (years)	n	Employment status	n	Nationality	n	Years since registration	n	Gender	n
Under 25	35	Full time	325	British	350	<5	98	Male	113
26–30	76	Part-time	72	English	7	6–15	137	Female	323
31–40	127	Retired	20	Irish	11	16–25	60	Total	436
41–50	84	Student	16	Scottish	4	26+	94		
51–60	81	Unemployed	3	Welsh	1	Total	389		
61–70	32	Total	436	Northern Irish	2	Missing data	47		
Total	435			EU	12				
Missing data	1			Non-EU	28				
				Dual British	7				
				Total	422				
				Missing data	14				

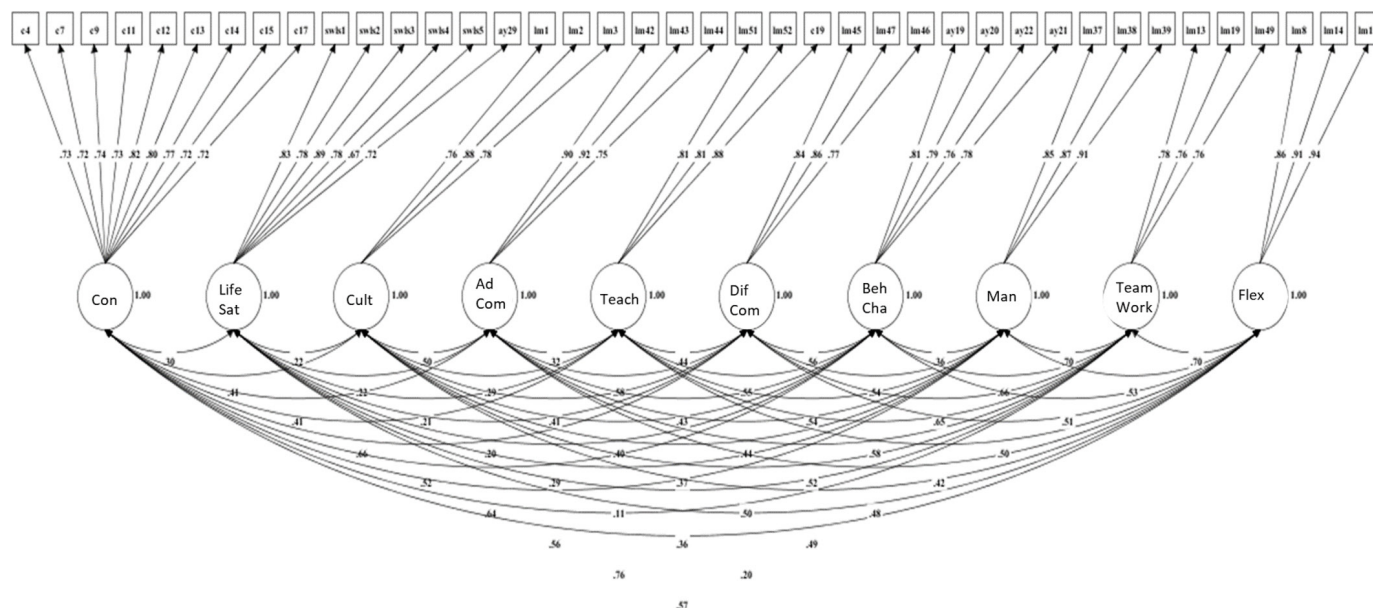


Figure 2 Latent variables and loadings.

for most of the theta spectrum. ‘Attitude to work’ had the lowest estimates for individual precision. Using the information functions as indicators of precision, ‘flexibility’ achieved the highest values and ‘attitude to work’, the lowest ones. As expected, an inverse situation is observable on the curves for the SEs of measurement, with ‘flexibility’ showing the lowest measurement errors and ‘attitude to work’ the highest ones. The precision, information and SE curves for the retrieved constructs under the MIRT analysis can be observed in figures 3, 4 and 5. The precision, information and SE curves demonstrate that the quality of the measures for each one of the proposed constructs varies across the latent spectrum, with lower levels of reliability and information and higher levels of SE of measurement in the extremes of the latent spectrum. The extreme right side of the spectrum has the worst reliability and highest error. The information curve, therefore, is indirect evidence of reliability with the advantage of being sample-independent.

Table 5 shows the Cronbach’s α coefficients for each one of the retrieved constructs. Taking the Cronbach’s α coefficients into account, the reliability estimates are somewhat divergent from the MIRT-based precision estimates. Using Cronbach’s α , the most reliable factor was ‘adapting communication’ and the least reliable was ‘cultural awareness’.

The analysis resulted in 40 items grouped into 10 constructs, the final list of constructs and the items that belong on each can be seen in table 6. Table 6 also shows the loading estimates, the SEs of the loading estimates, the ratios between the estimate and the SE and the two-tailed p values for the estimates. Table 6 shows the final selection of items with the dimension each one of them belongs.

DISCUSSION

In this study, we converted stakeholder-agreed PPD outcomes of healthcare professional international placements²⁷ into outcome statements, to assess which have the best psychometric properties for self-assessment. By piloting these statements with a large set of healthcare professionals and using item response theory to establish and test a set of latent traits and their associated questions, we were able to determine the 40 items with the best psychometric properties to create the MOVE-iT tool. Reliability evidence is favourable to the latent trait structure, both when using a single coefficient for the entire sample, and under the multidimensional item response theory approach. The validity evidence based on the internal structure of the questionnaire detailed in this study, combined with the content validity evidence based

Table 4 Comparison of selected goodness-of-fit indices between the unidimensional model and the proposed model

Models	χ^2	df	χ^2/df	RMSEA	CFI	TLI	WRMR
Unidimensional	8206.204	740	11.089	0.152	0.641	0.622	3.511
Proposed model	1736.922	695	2.499	0.059	0.950	0.944	1.271

CFI, Confirmatory Fit Index; RMSEA, Root Mean Square Error of Approximation; TLI, Tucker Lewis Index; WRMR, Weighted Root Mean Square Residual.

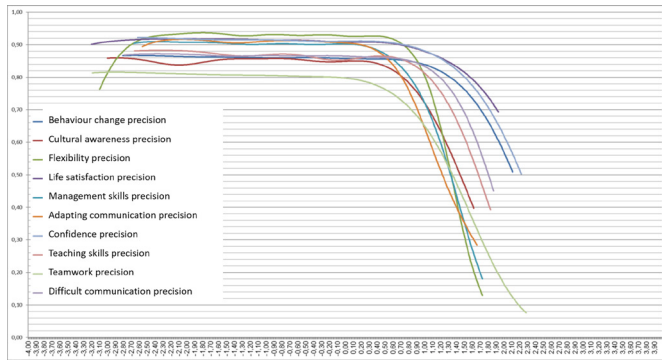


Figure 3 Estimates for mean individual precision of the latent variable scores.

on the selection of the initial pool of items⁵ helps build a strong validity argument in favour of the use of this questionnaire for the measurement of PDD-related dimensions of international placements. There were many more outcomes retained within the confidence domain as there were more items in the original data that we about confidence, and these items demonstrated more variability in responses regarding what people were confident about. We kept this as a large domain as we did not want to lose the richness of that data.

This paper aimed to consider whether a unique tool is needed to assess outcomes of UK healthcare professionals as a unique professional group, due to the qualitative reports of healthcare specific (ie, patient interaction outcomes) in the literature.^{9 25} We found that six of the outcome statements included in the MOVE-iT tool were specific to healthcare professionals (ie, “I am confident in my ability to manage myself in a clinical environment”). However, if one were to reduce the health specificity of the wording (eg, change the word clinical to work, or patient to customer), the tool has similarities to other psychometric measures introduced earlier in this paper.^{21 22} These similarities provide support for the application of all measures and suggest that MOVE-iT could be applicable outside of healthcare.

The 40 outcome statements that we found to have the best psychometric properties fell within the main outcome categories reported in past literature. For example, communication, leadership, attitude to work,

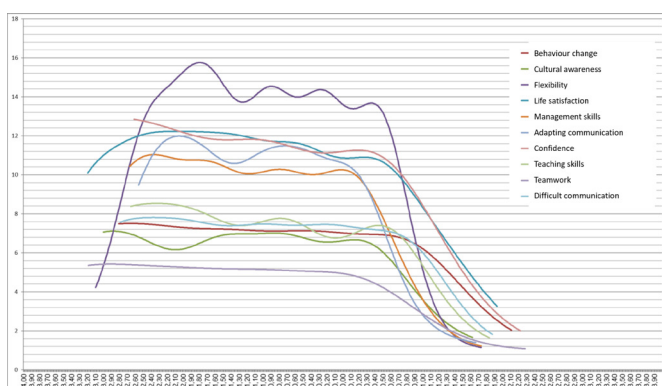


Figure 4 Information functions for the latent variables.

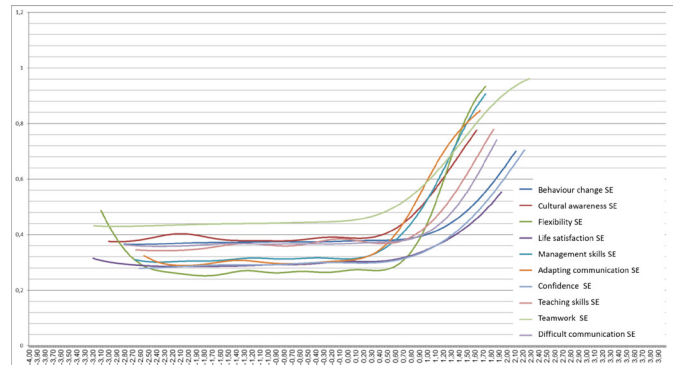


Figure 5 Estimates for individual SEs of measurement of the latent variable scores.

cultural awareness are frequently reported outcomes in the literature and domains within this tool.^{2 6 9 12 18} In our previous work, we criticise the current evidence base for being too vague in outcome reporting, as many papers report communication, leadership and cultural awareness as broad outcomes, rather than specify the relevant components within each that develop (specific skills, knowledge or attitudes).^{2 5 12} By using psychometric tests to assess latent traits, we further highlight the necessity for specific outcome reporting, as we found outcome statements associated with adapting communication and difficult communication to be two unique latent traits, rather than a single entity.

We hope that any healthcare professionals as individuals, project managers or NHS trusts may choose to use the tool in both a within-participant or between-participant manner (comparing outcomes preinternational and postinternational placements and comparing staff with and without international experience). By collecting data using the MOVE-iT tool and the variable statements developed in our previous work (to assess moderating or mediating variables that may affect outcomes), future researchers could begin to gather precise information about this learning (process, outcomes, variables).⁵ This should also be considered against measures of the list of costs reported in our previous work,⁵ as there is

Table 5 Cronbach's α coefficient for each construct	
Construct	Cronbach's α
Adapting communication	0.88
Confidence	0.86
Life satisfaction	0.86
Difficult communication	0.86
Management skills	0.86
Attitude to work	0.82
Flexibility	0.83
Teaching skills	0.78
Behaviour change	0.77
Cultural awareness	0.72

Table 6 Estimated discrimination parameters from the proposed MIRT model

Constructs/Items	Estimate	SE	P value (two-tailed)
Confidence			
I am confident in my ability to manage myself in a clinical environment.	0.727	0.030	0.000
I am confident in my abilities to work independently when necessary.	0.719	0.032	0.000
I am confident in my ability to deal with the unexpected.	0.743	0.025	0.000
I am confident in my ability to be adaptable and innovative as a leader.	0.733	0.024	0.000
I am confident in my ability to adapt and be flexible clinically.	0.823	0.021	0.000
I am confident in my ability to adapt and be flexible in general.	0.798	0.021	0.000
I am confident in my ability to find solutions despite limited resources.	0.770	0.022	0.000
I am confident in my ability to apply clinical skills to another context.	0.721	0.026	0.000
I am confident in my work.	0.724	0.025	0.000
Life satisfaction			
In most ways my life is close to my ideal.	0.834	0.02	0.000
The conditions of my life are excellent.	0.783	0.02	0.000
I am satisfied with my life.	0.893	0.017	0.000
So far I have gotten the important things I want in life.	0.776	0.024	0.000
If I could live my life over, I would change almost nothing.	0.667	0.029	0.000
Taking everything into consideration, I am satisfied with my job.	0.717	0.038	0.000
Cultural			
I demonstrated a good awareness about how culture influences health*.	0.761	0.036	0.000
I frequently demonstrated cultural sensitivity*.	0.881	0.031	0.000
I was constantly conscious of culture when working with patients*.	0.779	0.033	0.000
Adapting communication			
I changed the way I speak so that somebody can understand me (eg, purposely spoke slower and clearer)*.	0.899	0.024	0.000
I changed the way I communicate to make it more contextually appropriate (eg, to make it more culturally appropriate)*.	0.916	0.025	0.000
I frequently relied on my non-verbal communication (eg, hand gestures)*.	0.751	0.032	0.000
Teaching			
I demonstrated I'm a good teacher*.	0.813	0.024	0.000
I adapted the way I teach to make it better for the learner*.	0.807	0.023	0.000
I am confident in my ability to teach others.	0.883	0.031	0.000
Difficult communication			
I demonstrated that I am skilled in challenging conversations, even in high pressure situations*.	0.842	0.025	0.000
I demonstrated that I am able to manage difficult people effectively*.	0.862	0.021	0.000
I frequently dealt with difficult people*.	0.774	0.027	0.000
Behaviour change			
I am able to empower patients to help themselves.	0.807	0.026	0.000
I am able to empower colleagues to help themselves.	0.794	0.025	0.000
In my work I have demonstrated skills in changing colleagues' behaviour.	0.761	0.027	0.000
In my work I have demonstrated skills in encouraging and supporting patients to change behaviour.	0.778	0.027	0.000
Management			
I allocated tasks*.	0.848	0.021	0.000
I coordinated colleagues*.	0.868	0.02	0.000

Continued

Table 6 Continued

Constructs/Items	Estimate	SE	P value (two-tailed)
I demonstrated I am able to plan and organise*.	0.907	0.024	0.000
Attitude to work			
I was frequently proactive at work (eg, used my initiative, got on with things, thought on my feet)*.	0.778	0.027	0.000
I demonstrated that I am able to cope in work (eg, able to deal with stress)*.	0.763	0.028	0.000
I demonstrated that I am particularly good at working as part of team*.	0.765	0.026	0.000
Flexibility			
I demonstrated I'm good at dealing with the unexpected*.	0.857	0.037	0.000
I frequently had to find solutions despite limited resources*.	0.912	0.017	0.000
I demonstrated I am able to find solutions despite limited resources*.	0.937	0.017	0.000

*Items indicate that 'in the last month' is presented ahead of that statement, providing a time reference to consider the experience.

considerable literature regarding the ethical concerns of medical practice in LMICs, particularly when staff practice skills that they could not in an HIC.^{42 43} If mutual benefits could be evidenced using metrics, and costs minimised/mitigated by assessing the elements that increase mutual benefits, employers may be less reluctant to release staff to undertake such work.^{1 6} Evidence suggests that such work may be beneficial for the LMIC, the NHS and the individual professional.

Going forward we hope to develop a larger set of data, which will a) help us understand in more detail the processes associated with the outcomes, b) assess more thoroughly the reliability and validity of the tool, c) adapt or reduce the tool further based on future data and d) assess sensitivity of the tool to change.

Limitations

The tool only includes items which are either psychometrically related, or show variability of response. This means that many items that stakeholders considered important for inclusion in the core outcome set were not represented within the tool.⁵ This tool, therefore, compliments rather than replaces other tools which professionals reflect on all components of their PPD.¹⁹ This tool provides a way of evidencing benefits, however there is a body of critical evidence outlining the ethical concerns of medical practice abroad, particularly when individuals practice in ways that they might not in an HIC.^{43 44} A full cost-benefit analysis of this phenomena can be found in the authors' other work,¹⁵ the authors only advocate benefits in mutually beneficial, sustainable, ethical placements.

Conclusion

This evidence-based 40-item psychometric tool for self-assessment of outcomes from international placements (MOVE-iT) could be used in research and practice. Future work will reveal if the tool has the sensitivity to detect change in the domains.

Acknowledgements The authors would like to thank Anya Ahmed, Louise Ackers, Natalie Tate, Eileen Cunningham, John Chatwin, Hassan Osman and James Ackers-Johnson for being involved in the wider MOVE project. The authors would like to thank all of the stakeholders and participant groups involved for their significant contribution of time and effort.

Contributors NT participated in the design of the study, conducted the pilot and drafted the majority of the manuscript. LB-D conceived the design of the study, analysed data and contributed significantly to drafting the manuscript. CC provided oversight to the study design, conducted the PCA and statistical analysis and drafted the manuscript, GB provided oversight of study design, helped recruit participants and drafted the manuscript. All authors participated in the coordination of the research and read and approved the final manuscript.

Funding This work was supported by a project development grant from Health Education England (HEE), through the Global Health Exchange (GHE). Grant ref. NURA54. The full title of the study from which this analysis was derived was: Measuring the outcomes of volunteering for education (MOVE). The study was funded by Health Education England (Global Health Exchange). The research team were independent from the funding agency. The views expressed in this publication are those of the authors and not necessarily those of Health Education England or the Department of Health

Competing interests GB Ged Byrne is the Director of Global Engagement for Health Education England.

Patient consent for publication Not required.

Ethics approval Approval for the study was obtained from the Ethical Research Committee, University of Salford, and the University of Manchester Research Ethics Committee. Participants gave informed consent.

Provenance and peer review Not commissioned; externally peer reviewed.

Data sharing statement The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

REFERENCES

1. Crisp N. Mutual learning and reverse innovation--where next? *Global Health* 2014;10:14.
2. Sikka N. All aboard with Impact India. *Br Dent J* 2009;207:401-3.
3. Wood JB, Hills E, Keto FJ. Hands across the equator: the Hereford-Muheza link eight years on. *BMJ* 1994;308:1029-32.

4. Kosack CS. Experience of Médecins Sans Frontières in laboratory medicine in resource-limited settings. *Clin Chem Lab Med* 2012;50:1221–7.
5. Tyler N, Chatwin J, Byrne G, et al. *The benefits of International volunteering in a low-resource setting: Development of a core outcome set*.
6. Jones FA, Knights DP, Sinclair VF, et al. Do health partnerships with organisations in lower income countries benefit the UK partner? A review of the literature. *Global Health* 2013;9:38.
7. Abell C, Taylor S. The NHS benefits from doctors working abroad. *BMJ* 1995;311:133–4.
8. Kiernan P, O'Dempsey T, Kwalombota K, et al. Evaluation of effect on skills of GP trainees taking time out of programme (OOP) in developing countries. *Educ Prim Care* 2014;25:78–83.
9. Norton D, Marks-Maran D. Developing cultural sensitivity and awareness in nursing overseas. *Nurs Stand* 2014;28:39–43.
10. Standage R, Randall D. The benefits for children's nurses of overseas placements: where is the evidence? *Issues Compr Pediatr Nurs* 2014;37:87–102.
11. Hockey P, Tobin A, Kemp J, et al. Global health partnerships: leadership development for a purpose. *Leadersh Health Serv* 2009;22:306–16.
12. Dean E. 'Tanzania changed me'. *Nurs Stand* 2013;27:16–17.
13. Lumb A, Murdoch-Eaton D. Electives in undergraduate medical education: AMEE Guide No. 88. *Med Teach* 2014;36:557–72.
14. Button L, Green B, Tengnah C, et al. The impact of international placements on nurses' personal and professional lives: literature review. *J Adv Nurs* 2005;50:315–24.
15. Ackers HL, Ackers-Johnson J, Chatwin J, et al. *Healthcare, Frugal Innovation, and Professional Voluntarism [Internet]*. Cham: Springer International Publishing, 2017.
16. Tropical Health Education Trust. In our mutual interest [Internet]. 2017 <https://www.thet.org/wp-content/uploads/2017/08/In-Our-Mutual-Interest.pdf>.
17. Crisp N. *Turning the world upside down : the search for global health in the twenty-first century*: Royal Society of Medicine Press, 2010. 228.
18. Wright J, Walley J, Philip A, et al. Research into practice: 10 years of international public health partnership between the UK and Swaziland. *J Public Health* 2010;32:277–82.
19. Longstaff B. *Toolkit for the collection of evidence of knowledge and skills gained through participation in an international health project*.
20. Smith C, Pettigrew LM, Seo HN, et al. Combining general practice with international work: online survey of experiences of UK GPs. *JRSM Short Rep* 2012;3:1–8.
21. Lough BJ, McBride AM, Sherraden MS. Measuring Volunteer Outcomes Development of the International Volunteer Impacts Survey Measuring Volunteer Outcomes: Development of the International Volunteer Impacts Survey. *Online* 2009.
22. Hammer MR, Bennett MJ, Wiseman R. Measuring intercultural sensitivity: The intercultural development inventory. *International Journal of Intercultural Relations* 2003;27:421–43.
23. Braskamp L, Braskamp D, January KM-R. undefined. Global Perspective Inventory (GPI): Its purpose, construction, potential uses, and psychometric characteristics. academia.edu [Internet]. 2008 <http://www.academia.edu/download/44637248/manual.pdf> (cited 2019 Apr 18).
24. Lough BJ, McBride AM, Sherraden MS. *Perceived effects of international volunteering: Reports from alumni*. Center for Social Development, Washington University: St Louis, MO, 2009.
25. Briscoe L. Becoming culturally sensitive: a painful process? *Midwifery* 2013;29:559–65.
26. Streiner DL, Norman GR, Cairney J. *Health Measurement Scales [Internet]*. Vol. 1: Oxford University Press, 2015. Available from: <http://www.oxfordmedicine.com/view/>. (cited 2019 Apr 18).
27. Tyler N, Chatwin J, Byrne G, et al. The benefits of international volunteering in a low-resource setting: development of a core outcome set. *Hum Resour Health* 2018;16:69.
28. Wu M, Ping H, Jen T. Educational Measurement for Applied Researchers [Internet]. 2017 https://books.google.co.uk/books?id=EOHTDQAAQBAJ&pg=PA22&lpg=PA22&dq=tap+into+latent+traits&source=bl&ots=KNzd0XfZxn&sig=754d3_DYxsS8Hh2DBTny9h8Uf0&hl=en&sa=X&ved=0ahUKEwj-8rbsmPLTAhVIJ8AKHcFEAkoQ6AEILDAB#v=onepage&q=tap.
29. Hambleton RK, Cook LL. Latent trait models and their use in the analysis of educational test data1,2,3. *J Educ Meas* 1977;14:75–96.
30. Evans DB, Murray CJL. Health systems performance assessment : debates, methods and empiricism [Internet]. World Health Organization. 2003 https://books.google.co.uk/books?id=Ut_cxcWHobYC&pg=PA739&lpg=PA739&dq=tap+into+latent+traits&source=bl&ots=jh4GceWRQt&sig=MwkVD0ZpHWATuLYNwJU9u3snMQ&hl=en&sa=X&ved=0ahUKEwj-8rbsmPLTAhVIJ8AKHcFEAkoQ6AEIMTAC#v=onepage&q=tap.
31. Charter RA. Sample size requirements for precise estimates of reliability, generalizability, and validity coefficients. *J Clin Exp Neuropsychol* 1999;21:559–66.
32. Diener E, Emmons RA, Larsen RJ, et al. The Satisfaction With Life Scale. *J Pers Assess* 1985;49:71–5.
33. My eForms - eForms [Internet]. <https://eforms.mhs.man.ac.uk/> (cited 2017 May 19).
34. Ericsson KA, Simon HA. How to Study Thinking in Everyday Life: Contrasting Think-Aloud Protocols With Descriptions and Explanations of Thinking. *Mind Cult Act* 1998;5:178–86.
35. Caspar RA, Lessler JT, Willis GB. Reducing survey error through research on the cognitive and decision processes in surveys. *Short course Present 1999 Meet Am Stat Assoc* 1999.
36. Leung W. How to design a questionnaire. *Student BMJ* 2001;9:187–9.
37. IBM Corp. *IBM SPSS Statistics for Macintosh, Version 24.0*. 2016, 2016.
38. Osteen P. An Introduction to Using Multidimensional Item Response Theory to Assess Latent Factor Structures. *J Soc Social Work Res* 2010.
39. Liang X, Yang Y. An evaluation of WLSMV and Bayesian methods for confirmatory factor analysis with categorical indicators. *Int J Quant Res Educ* 2014.
40. Muthén L, Muthén B. *Mplus user's guide*. 8th ed. Los Angeles: Author, 2017.
41. Electronic Staff Record Programme Data Warehouse. *North West Staff in post* 2013.
42. Dowell J, Blacklock C, Liao C, et al. Boost or burden? Issues posed by short placements in resource-poor settings. *Br J Gen Pract* 2014;64:272–3.
43. Banatvala N, Doyal L. Knowing when to say "no" on the student elective. Students going on electives abroad need clinical guidelines. *BMJ* 1998;316:1404–5.
44. Dowell J, Merrylees N. Electives: isn't it time for a change? *Med Educ* 2009;43:121–6.