TASO Transforming Access and Student Outcomes in Higher Education

Validating a new questionnaire scale

Brief user guidance

Validating a new questionnaire scale

BACKGROUND

Evaluations of student access and student success programmes look to understand how such programmes affect a range of student outcomes. These outcomes may fall within one of two categories.

First, the category of attainment and progression: outcomes such as exam scores, higher education access, progression, degree completion, employment, and others, may be the ones programmes ultimately look to improve; but sometimes can require long waiting times until they can be measured.

And second, the category of intermediate outcomes: outcomes such as self-efficacy, sense of belonging, cognitive skills, metacognitive skills, attitudes and expectations, and many others, are important in and of themselves; can provide an indication of whether the programmes may in the future deliver improvements in terms of attainment and progression; and often rely on self-report measures administered via questionnaires.

When measuring intermediate outcomes, it is important to use the best possible questionnaire scales. Using well-designed and validated scales contributes to good-quality data, which is important for the success of robust evaluation.

While many questionnaire scales are available, information about their quality is not always forthcoming or complete. This makes it difficult to ascertain whether they are appropriate for administration to the populations of interest and how they may generate good-quality data.

The solution to this potential issue is to use a validated scale capturing the outcome of interest.

A validated scale is a measurement scale that has undergone a multi-step process to ensure it captures the outcome it intends to, does so consistently and reliably, is appropriate for the intended populations, and is associated with other relevant outcomes in an expected manner.

A range of validated scales are available for use:

- TASO's Access and Success Questionnaire (ASQ).
- TASO's rapid evidence review of intermediate outcomes for higher education access and success.
- The Education Endowment Foundation's SPECTRUM database.
- The Toolkit for Access and Participation Evaluation (TAPE).

Other scales are in the process of being validated across the sector.

Scale validation is complex and therefore, wherever possible, using an existing scale is a much better option than validating one from scratch.

Sometimes, though, a programme looks to affect a different outcome, for which a scale has not yet been validated. This document provides a brief non-technical guide through the validation process.

THE VALIDATION PROCESS

Step 1. Identify and define the outcome

The first step is to identify the outcome the programme, activity or intervention looks to affect, and establish whether a validated measurement scale already exists for it.

To support the above, review the literature around outcomes that programmes, activities, or interventions such as yours look to affect and how they do so. Also, consider the literature on their respective effectiveness. Engage critically with the way in which these outcomes may support your wider student access or success goals.

Generate a definition for the outcome. It is very likely that the outcome will have been measured at some point in the past, but perhaps with different populations, at different ages or educational stages, in different countries, and therefore cultural, contexts. So, while the definition of the outcome may be the same, the scale required for measuring it in the population and context of interest may need to be different.

Search existing resources to identify what measures do exist, even if they are not appropriate for the population of interest:

- TASO's Access and Success Questionnaire (ASQ)
- TASO's rapid evidence review of intermediate outcomes for higher education access and success
- The Education Endowment Foundation's SPECTRUM database
- The Toolkit for Access and Participation Evaluation (TAPE)

Having established that no validated scale exists, and a new one needs to be generated, start the validation process.

Step 2. Assemble an initial long list of items, with prompts and response options

Drawing on existing measures of the same outcome (even if with different populations or in different contexts), or on measures of outcomes deemed to be related, assemble a long list of items.

Assemble ideally between six and twelve items, guided by any suggestions from other relevant scales. Combine items from across scales if they are consistent in their framing. Sometimes a higher number of items will be needed if the outcome is very complex: those outcomes are best developed in collaboration with psychometricians or measurement experts.

Items

Adapting items from existing scales is strongly encouraged, while ensuring their framing and wording are relevant to the population and context of interest.

Framing: items may be framed around the frequency of behaviours, agreement with statements, reflections of one's views, the similarity of statements with one's actions, etc. The key is that all items of a scale are framed the same way.

Wording: items should only ever ask about one single behaviour, attitude, perspective, or action at a time; items should be worded as simply and clearly as possible, using simple words, avoiding jargon, and remaining as brief as possible

The prompt to be provided to potential respondents should also be considered. This should be as simple as possible while ensuring that it supports consistency of understanding across respondents. Provide definitions of key terms if these support that understanding.

Finally, response options should follow the framing of the items. Response options include frequency scales ('never' to 'always'), agreement scales (often called Likert scales: 'strongly disagree' to 'strongly agree'), and many others. In determining the response options, the number of options should also be considered: an odd number of options (usually five) offers respondents a neutral option, while an even number of options forces a decision in relation to the response. Adapting response options from any existing scales is strongly encouraged.

Response options

The most common response scales are five-point Likert scales: "strongly disagree" – "disagree" – "neither agree nor disagree" – "agree" – "strongly agree". Likert scales assume that the steps between each of these options are equal, which allows this data to be used in a range of statistical analyses.

Once assembled, the totality of the items, prompts, and response options should be considered together. Identify any inconsistencies, ensure clarity of expression, and prepare for testing.

Step 3. Test the scale and item long list with the population of interest

Once assembled, test the long list of items with a small group of individuals from the population of interest. This process is often referred to as cognitive testing, see here for further details.

This testing takes the form of short conversations about the meaning, clarity, and approach to responding to the scale items, including understanding the prompt and use of the response options.

These conversations can take place individually with relevant learners or as part of a group. The key aim is for individuals to offer their honest and full perspectives on every aspect of the scale.

Conversations, individually or in a group with at least four learners should ensure a range of feedback is provided. Ensure the group includes a diverse range of learners, as relevant to your population of interest.

A common learner testing format

Introduce participants to the aim of the conversation: for them to provide feedback as to the clarity, simplicity, and meaning of the statements/questions they're about to see.

Ask participants to read each statement/question and then answer it, and then tell you about its: clarity, any issues with wording, clarity of response options and what they think about when answering it.

Assemble all the feedback and review it. Discard any items that are shown to have inconsistent meaning, change any words that emerged as problematic, and make all adaptations to respond to all received feedback. This will result in a short list of items.

Step 4. Test the scale and item short-list with the population of interest

Once the scale with its revised short-list of items and revised prompt and response options has been assembled following feedback, it requires testing with another group of the population of interest, this time in the format it will be used during full-scale implementation: a questionnaire.

Generate the brief questionnaire to include the scale, with its prompt and response options.

In this questionnaire also include any other information that may be relevant:

- Background characteristics of learners who may want to understand if the scale works.
- One or more other validated scales measuring outcomes that you expect, or know from the literature, are related to the outcome you are looking to measure.
- Ideally one measure of educational attainment, achievement, or progression.

Administer the questionnaire either on paper or online, in the format most likely to be used during full-scale implementation.

Testing questionnaire set-up

This is a formal research questionnaire and therefore a data protection notice, consent form, and all relevant information must be provided to the individuals responding to the questionnaire. Collect only as much data as strictly required for your analysis, and ideally in an anonymous and de-identified manner.

The higher the number of items in the scale being validated, other background characteristics, and other relevant scales are included in the questionnaire, the higher the number of respondents to the questionnaire needs to be. At a minimum, 30-40 respondents per characteristic of interest are required. If only whole-sample analyses are planned, then a minimum of 30-40 total respondents are required.

Maximising responses

Non-response is normal when administering questionnaires and should be expected. Under circumstances where the questionnaire is administered online and there is no direct contact with the respondents, assume at least 50% of the group initially invited to respond will not actually do so.

Maximise responses by keeping the questionnaire short and simple. If able, work with a group with whom you can interact directly. This can often result in a better response rate.

Follow ethical guidelines and never force a response, either to the whole questionnaire or to specific items.

Once the data from the questionnaire's administration has been assembled, undertake as much of the analysis below to get a basic understanding of the scale's quality. All of these can be carried out in any spreadsheet software (Excel, Sheets, etc.) or in specialist statistical software (SPSS, R, Jasp, Stata, etc.)

The full statistical testing of the scale is outside the scope of this guide. Approach the analysis below carefully, as background knowledge is required for appropriate interpretation. Wherever possible, work with others, either in specific departments of a higher education provider, or with other colleagues who have undertaken this work before. Also explore the training options available via a higher education provider, usually around 'statistics for the social sciences' or similar.

With care, carrying out even part of the analysis below will provide you with information about how the scale and its items work.

Analyses

Face validity. Explore your items in terms of the feedback from learners. If any of the items on the shortlist are still problematic, then consider removing them from the scale. This may require further testing with one or two learners to obtain feedback.

Item behaviour. Explore your items descriptively: explore how responses are spread by response option, for each item; look at whether any response option is never used, or if most responses are at one (or both) extremes of the response scale.

Expected relationship between items. Explore your items in (all possible pairs): plot responses to each pair of items, checking if they behave as expected. For example: if you assume that stronger agreement for one item would normally also see stronger agreement on another, is this the case?

Internal consistency. Explore further how the items of the scale relate to each other. This is usually measured using Cronbach's alpha. This is a measure of how a scale of a given number of items performs, in terms of the overall variance across the scale items and the average co-variance between all the pairs of items in the scale. Values of 0.7 are usually considered good, although this varies, and few-item scales usually display lower values of the coefficient.

Factor analysis. Different types of factor analysis exist. Exploratory factor analysis (EFA) allows for a smaller set of items to be chosen from the shortlist in a way that sees them capture the same underlying outcome (as defined by yourself in Step 1). Confirmatory factor analysis (CFA) allows for the testing of your assumption about the fact that you are measuring one (defined by yourself in Step 1) outcome. Factor analysis can also help generate a single factor score, that is an overall score for the outcome which assumes that each item may contribute differently to this overall score. Such analyses should not be undertaken without prior knowledge as results can be misinterpreted.

Concurrent validity. Explore how the scale measures an outcome in comparison to another, validated, or established measure (the ones you may have added to the questionnaire). If you have included a different scale measuring a related (or similar) outcome in your questionnaire, explore whether these two measures are related to each other. The simplest way is to sum the scores of all items together (although this breaks the assumption of the earlier factor analysis, so use a factor score if you have generated one) and then plot their relationship to explore whether the two scores change together as expected.

Predictive validity. Explore how the scale is related to a different, usually external measure. If you have collected any measure of educational attainment, achievement, or progression, explore how the scale (either as a sum score, or via its factor score) is associated with this measure, and if this relationship is as expected. You may not always be looking for a positive relationship here, sometimes you may expect that as participants score more highly on your new scale, they score less highly in terms of the external measure you have collected. This is entirely down to what outcome you are measuring.

Step 5. Interpret the analysis results and plan the next steps

Document the results of any of the above analysis you have undertaken. Interpret the results holistically, looking across every element you have undertaken. If results point consistently towards items and the scale behaves as expected, then you have (basic) evidence of the quality of the scale and can proceed to deploy it in your evaluation.

Please remember that the process may result in scales without sufficient support from the above analysis. If so, return to Step 1 above and look to work with others also trying to measure your outcome of interest.

Scale validation is complex and therefore, wherever possible, using an existing scale is a much better option than validating one from scratch.

Helpful resources

The resources below provide a starting point for further insight into the topics and points raised above.

Artino Jr, A. R., La Rochelle, J. S., Dezee, K. J., & Gehlbach, H. (2014). Developing questionnaires for educational research: AMEE Guide No. 87. *Medical teacher*, *36*(6), 463-474. https://www.tandfonline.com/doi/full/10.3109/0142159X.2014.889814

Baines, L., Gooch, D., & Ng-Knight, T. (2022). Do widening participation interventions change university attitudes in UK school children? A systematic review of the efficacy of UK programmes, and quality of evaluation evidence. *Educational Review*, 1-20. https://www.tandfonline.com/doi/pdf/10.1080/00131911.2022.2077703

Boynton, P. M., & Greenhalgh, T. (2004). Selecting, designing, and developing your questionnaire. *British Medical Journal*, *328*(7451), 1312-1315. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC420179/

Designing Likert scales - TASO. (n.d.). TASO. https://taso.org.uk/evidence/evaluation-guidance-resources/survey-design-and-validation/survey-design-resources/evaluation-guidance-designing-likert-scales/

Overview of Cognitive Testing and Questionnaire Evaluation. (n.d.). Overview of Cognitive Testing and Questionnaire Evaluation | Harvard University Program on Survey Research. https://psr.iq.harvard.edu/book/overview-cognitive-testing-and-questionnaire-evaluation

Survey design and validation resources - TASO Webinar. (n.d.). TASO. https://taso.org.uk/evidence/evaluation-guidance-resources/survey-design-and-validation/survey-design-resources/

Trobia, A. (2011, January 1). *Sage Research Methods - Encyclopedia of Survey Research Methods*. Sage Research Methods - Encyclopedia of Survey Research Methods. https://doi.org/10.4135/9781412963947

Younger, K., Gascoine, L., Menzies, V., & Torgerson, C. (2019). A systematic review of evidence on the effectiveness of interventions and strategies for widening participation in higher education. *Journal of Further and Higher Education*, *43*(6), 742-773. https://www.tandfonline.com/doi/pdf/10.1080/0309877X.2017.1404558