Technical Note 42

Young Lives

Young Lives School Surveys 2016–17

The Design and Development of Transferable Skills Tests in India and Vietnam

Padmini Iyer and Obiageri Bridget Azubuike

www.younglives.org.uk

NOVEMBER 2017

Young Lives School Surveys 2016–17: The Design and Development of Transferable Skills Tests in India and Vietnam

Padmini Iyer and Obiageri Bridget Azubuike

First published by Young Lives in November 2017

Printed on FSC-certified paper from traceable and sustainable sources.

About Young Lives

Young Lives is an international study of childhood poverty, following the lives of 12,000 children in four countries (Ethiopia, India, Peru and Vietnam) over 15 years. **www.younglives.org.uk**

Young Lives is core-funded by UK aid from the Department for International Development (DFID).

Core-funded by

UKaid
from the British people

Young Lives, Oxford Department of International Development (ODID), University of Oxford, Queen Elizabeth House, 3 Mansfield Road, Oxford OX1 3TB, UK

Tel: +44 (0)1865 281751 • Email: younglives@younglives.org.uk

Contents

The authors

Padmini Iyer is an education research officer at Young Lives, and she led on the design and implementation of the 2016-17 Young Lives secondary school survey in Vietnam. Her research interests include educational access and equity, twenty-first century skills, gender, adolescence, and sexual and reproductive health.

Bridget Azubuike is a quantitative research assistant in the Young Lives education team. She provides support to the research and education teams in the development of the school surveys, data analysis and research dissemination. She has an MSc in Economics from the University of Sussex and a BSc in Economics and Statistics from University of Benin, Nigeria.

Acknowledgements

The work discussed in this technical note was conducted by the Young Lives education team at the University of Oxford. We would like to thank Caine Rolleston, Rhiannon Moore and Jack Rossiter for their comments and contributions to this note, and Rebecca Schendel for her support and advice on conceptualising and measuring problem solving and critical thinking in low and middle-income contexts. We would also like to thank the research department at the Council for Aid to Education (CAE) for granting Young Lives permission to adapt sample selected-response items from the College Work and Readiness Assessment Plus (CWRA+) to assess critical thinking in the Young Lives transferable skills test. Items used to assess problem solving in the Young Lives transferable skills test were adapted from OECD (2004a) *PISA Problem Solving Items and Scoring Guides*, licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 3.0 IGO (CC-BY-NC-SA 3.0 IGO) license.

1. Introduction

Young Lives is an international study of childhood poverty in Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam. Since 2002, Young Lives household surveys have followed the lives of two cohorts of 12,000 children in these four countries in two age cohorts: an 'Older Cohort' born in 1994-95, and a 'Younger Cohort' born in 2001-02. In 2010, Young Lives introduced a series of school surveys in all four countries, which included a sub-sample of children in the Younger Cohort. Between 2010 and 2013, the school surveys examined issues of school quality and effectiveness in primary schools in Young Lives sites in Ethiopia, India (Andhra Pradesh and Telangana), Peru and Vietnam.

Building on the design of the primary school surveys, the 2016-17 Young Lives school surveys examined school effectiveness at upper primary level in Ethiopia, and at secondary level in India and Vietnam (Rossiter et al. 2017; Moore et al. 2017; Iyer et al. 2017). The surveys examined school effectiveness through multiple outcome measures, including students' learning progress in maths and English. This involved the administration of two linked maths tests and two linked English tests, administered at the beginning and the end of the school year (Wave 1 and Wave 2 of data collection respectively) (Azubuike, Moore and Iyer 2017).

In addition to assessing learning progress in maths and English, the 2016-7 school surveys included a transferable skills test, administered at Wave 2 in India and Vietnam.[1] This note outlines the design and development of this cognitive test, which assessed students' problem-solving and critical-thinking skills. After a brief introduction to the literature around the development of transferable skills, the note outlines the ways in which problem solving and critical thinking were conceptualised and measured in the 2016-17 school survey. The process of developing, piloting and selecting items for the test in India and Vietnam is then described, followed by results of scale validation using transferable skills test data from the main survey.

2. Skills for the twenty-first century

Transferable skills provide young people with 'critically needed tools to be able to succeed in terms of employment, health and personal well-being' (Rankin et al. 2015). Such skills are seen as essential to prepare young people for 'complex life and work environments' in the twenty-first century (P21 2015). Transferable skills are also referred to as twenty-first century skills (P21 2015) and 'key soft skills for youth workforce success' (Lippman et al. 2015). Shared across these definitions is an emphasis on social and emotional skills such as communication and teamwork, and 'higher order cognitive skills' such as problem solving and critical thinking (P21 2015; Lippman et al. 2015; Rankin et al. 2015; World Bank 2014).

If the development of foundational skills such as numeracy and literacy is central to the 'first phase' of quality education, then the development of transferable skills such as problem solving and critical thinking can be seen as the 'next phase'. This next phase of quality education is of increasing policy interest at secondary level in both India and Vietnam. In India,

1 The transferable skills test was not administered in the 2016-17 Ethiopia school survey. While transferable skills are relevant to education policy in Ethiopia, since the 2016-17 Ethiopia school survey took place at upper primary level, it was decided that the assessment of 'higher order' cognitive skills is less relevant at this level of schooling (see Iyer and Moore 2017).

the New Education Policy (drafted in 2016) seeks to ensure that schools equip students with 'life skills' (including creativity, critical thinking and problem solving) suited to the requirements of a developing 'knowledge-based' economy (MHRD 2016). Meanwhile, in Vietnam, a new competency-based curriculum under the General Renovation of Education seeks to encourage the development of skills such as communication, teamwork and problem solving through learner-centred approaches at primary and secondary level (World Bank 2015).

While the importance of transferable skills at secondary level is emphasised globally and within Young Lives study countries, there is limited evidence on the extent to which young people develop these skills at school in low and middle-income countries (as highlighted by Rankin et al. 2015). The transferable skills test in the 2016-17 school survey in India and Vietnam therefore aimed to provide evidence on some of the student, class, teacher and school factors associated with the development of problem solving and critical thinking, and the relationship between these skills and learning outcomes in school subjects.

3. Conceptualising problem solving and critical thinking

The conceptualisation of problem solving and critical thinking as 'higher order' cognitive skills originates in Bloom's influential Taxonomy of Educational Objectives (1956, revised in 1990), in which more simple cognitive processes (remembering, understanding) must be developed in order to support more complex cognitive processes (analysing, evaluating, creating) (Mayer 2002; Funke 2010). Within an educational context, problem solving and critical thinking can be considered both as domain-specific skills (as required within specific subjects such as maths and the sciences), or as cross-curricular skills (those which can be developed and applied across domains and within 'real-life' situations) (Kuhn 1999; Scherer and Beckman 2014; Greiff et al. 2013). There is also debate as to the relationship between problem solving and critical thinking; for example, are they linked constructs within a hierarchical relationship (with problem solving as the lesser of the two skills), or completely separate constructs?

In the 2016-17 school survey, we examined problem solving and critical thinking as **cross-curricular skills** rather than domain-specific skills. In light of our interest in problem solving and critical thinking as transferable skills, it was more relevant to consider the potential application of such skills in 'real life' contexts rather than narrowly within school subjects. We also understood problem solving and critical thinking as **separate constructs**, following the definitions below:

Problem solving: An individual's capacity to use cognitive processes to resolve real, cross-disciplinary problems where the solution path is not immediately obvious (Greiff et al. 2013: 74).

Critical thinking: Skills such as inference and evaluation which are applied to ill-structured problems, and for which 'there are no definitive solutions' (Kuhn 1999; Thomas and Lok 2015).

The following section describes existing measures to assess problem solving and critical thinking among secondary school students, and considers their suitability to our study contexts.

4. Measuring problem solving and critical thinking

4.1. Problem solving

Although there has been extensive research on problem solving from a cognitive perspective, work on transferring this research into formats suitable for psychometric testing has only begun relatively recently (Greiff et al. 2013; Reeff et al. 2006). PISA (Programme for International Student Assessment) is the key example of assessing problem solving within educational contexts on an international scale. This began in 2003, when analytical problem solving was assessed using pen and paper tests; the definition of problem solving given above (Greiff et al. 2013: 74) is the definition of analytical problem solving adopted in PISA 2003. Subsequent assessments of problem solving in PISA 2012 and PISA 2015 have assessed interactive[2] and collaborative[3] problem solving respectively, using computer-administered adaptive testing. It has been argued that these more recent tests are more effective at assessing problem solving as it occurs in the 'real-world' (Greiff et al. 2013). However, the assessment of more complex problem-solving skills requires more sophisticated forms of test administration. The paper-and-pen format of the PISA 2003 problem-solving assessment was more suitable for Young Lives study contexts; moreover, the assessment of analytical problem solving still offered a valuable insight into students' problem-solving abilities.

Publicly-available items from the PISA 2003 problem-solving assessments include ten units which assess three different types of problem solving: decision-making, trouble-shooting, and system design and analysis. The items include multiple-choice, closed-constructed and open-constructed formats. When selecting items for adaptation and piloting, a key consideration was the extent to which the item format and/or problem context would need to be adapted for our test format (limited to multiple-choice questions), and how this might affect the cognitive processes assessed.

Based on these considerations, four decision-making units ('Energy Needs', 'Cinema Outing', 'Holiday', 'Transit System'), one system analysis and design unit ('Course Design') and one trouble-shooting unit ('Irrigation') were selected for adaptation and translation (for original units and scoring guides, see OECD 2004a and OECD 2004b). These units were chosen because they were made up of multiple-choice items and were set in contexts which would be familiar to young people in our study countries, or because they were made up of items which could adapted to multiple-choice format without diminishing the problem-solving skills assessed. Examples of contextual adaptation included transforming a metro map described in 'Transit System' into a bus map in a renamed unit ('Journey'), and adapting the problem situation described in 'Cinema Outing' into one in which a group of students must make decisions about which extra classes they can attend at school (rather than which films they can see at the weekend – the unit was renamed 'Extra Classes'). These items were then

2 'Characterised by the dynamic interaction between a problem solver and the problem to generate and integrate information about the problem' (Greiff et al. 2013: 76).

3 'The capacity of an individual to effectively engage in a process whereby two or more agents attempt to solve a problem by sharing the understanding and effort required to come to a solution' (OECD 2012, in Greiff et al. 2013: 81).

translated into Telugu and Urdu for administration in Young Lives sites in Andhra Pradesh and Telangana, India and into Vietnamese for administration in Young Lives sites in Vietnam. Further contextual adaptation included replacing proper nouns (names of people, locations) with those suitable to each country.

4.2. Critical thinking

Critical thinking has long been considered an important component of education in Europe, but the concept has only gained prominence within international education in recent years – largely in reference to the global 'knowledge economy' (Schendel and Tolmie 2017). However, in spite of a recent international consensus around its importance, critical thinking and its measurement remains one of the most contested constructs in education (Schendel and Tolmie 2017). A key debate relates to whether critical thinking should be conceptualised as a set of cognitive skills (i.e. Halpern 1996) or whether it also requires a dispositional element (i.e. Facione 1990). As Schendel and Tolmie (2017) note, Kuhn's (1999) theory of critical thinking development incorporates both sides of this definition, with critical thinking conceptualised as requiring both cognitive skills (e.g. abstraction, the ability to differentiate theory from evidence) and metacognitive strategies that allow for 'a sophisticated epistemological world view that recognises the uncertainty of knowledge' (Schendel and Tolmie 2017). Six 'functional attributes' of critical thinking recently outlined by Thomas and Lok (2015) also incorporate cognitive skills (interpretation, explanation, analysis, inference and evaluation) as well as metacognitive skills (self-regulation).

The Council of Aid to Education (CAE) has developed 'College Work and Readiness Plus' (CWRA+) tests which assess middle and high school students' (aged 14-18) critical-thinking skills in the USA. CWRA+ includes selected-response items which assess critical reading and evaluation, and students' ability to critique an argument. These elements of the CWRA+ assessments were selected for the Young Lives transferable skills test as they offered a way to assess the cognitive elements of critical thinking highlighted by Thomas and Lok (2015) – interpretation, explanation, analysis, inference and evaluation – within a multiple-choice format, and at a level suitable for students in the Young Lives study contexts.[4]

As with the adaptation of PISA problem-solving items described above, existing CWRA+ items were adapted to ensure that the content was relevant for the study contexts in India and Vietnam in which they would be administered. One 'Critical Reading and Evaluation' exercise (two letters and five linked questions) and one 'Critique an Argument' exercise (one passage and three linked questions) were adapted for use in our study countries. The original 'Critical Reading and Evaluation' exercise included a proposal to install security cameras in school to address incidents of theft and bullying from a school principal, and an opposing view from a school parent. This was adapted so that the proposal to install security cameras was in response to concerns around girls' safety in India, while the original focus on theft and bullying was retained for Vietnam. The original 'Critique an Argument' exercise argued that computer programming should be taught in schools instead of foreign languages; for both India and Vietnam, the references to 'foreign languages' were replaced with 'English language' as this was more relevant in both contexts. In addition to the adaptation of these CWRA+ exercises, we developed two new exercises in a similar format: a 'Critical Reading

4 CAE have also developed College Learning Assessments (CLA), which assess both the cognitive and meta-cognitive elements of critical thinking using 'performance tasks' – see Schendel and Tolmie (2017). However, it was decided not to use the CLA performance tasks to assess critical thinking in the 2016-17 school surveys; these assessments require criterion-reference scoring and well-trained assessors, which were beyond the resources available to the Young Lives surveys.

and Evaluation' exercise with a proposal from a school principal to introduce yoga (in India) or *khí công* (in Vietnam) to address student fatigue, and a blog post from a parent criticising the proposal; and a 'Critique an Argument' exercise with questions linked to a passage arguing that low and middle-income countries should not be expected to reduce fossil fuel consumption until they are more economically developed.

4.3. Literacy or transferable skills?

The assessment of skills such as problem solving and critical thinking can be significantly affected by students' literacy levels (Marin and Halpern 2011). Poor performance on such tests may therefore be a reflection of students' reading comprehension ability rather than their limited problem-solving or critical-thinking skills. We therefore decided to include basic comprehension questions with each problem solving and critical thinking exercise, which required students to identify basic information from presented text or figures. The aim of these comprehension questions was to allow us to distinguish between students' problem-solving or critical-thinking skills and their literacy levels.

5. Piloting and item selection

Qualitative pre-piloting and larger scale piloting of problem-solving and critical-thinking items was conducted between March and December 2016 in India and Vietnam (see Table 1).

Table 1. *Transferable skills test development and administration timeline*

		India	Vietnam
2016	January – February		Adaptation and translation: problem-solving items
	March		Pre-pilot: problem-solving items (116 students, 6 schools)
	August	Adaptation and translation: problem-solving items	
	October – December	Adaptation and translation: critical-thinking items Pre-pilot and pilot: critical-thinking and problem-solving items (India: 219 students, 4 schools) (Vietnam: 355 students, 4 schools)	
2017	January – March	Data collection (Problem solving: 4,070 students) (Critical thinking: 4,058 students)	
	March – April		Data collection (Problem solving and critical thinking: 7,792 students)

Qualitative pre-piloting with students was carried out in order to check the suitability of item difficulty and content, and to identify any issues with translation; teachers were also consulted regarding the suitability of the tests. After revisions to the items based on student and teacher feedback (e.g. relating to language and contextual relevance), larger-scale pilots were then conducted in order to help identify 'floor' and 'ceiling' effects among different groups (e.g. items which might be too easy or too difficult for students in urban vs. rural schools, and/or government vs. private schools, and so on).

In total, 24 problem solving items (across six different units) and 22 critical-thinking items (across four different passages) were piloted in India and Vietnam. For piloting in India, the items were translated into Telugu and presented in dual-medium format (i.e. English and Telugu), and for piloting in Vietnam, the items were translated into Vietnamese. A smaller qualitative pilot was also undertaken in Urdu-medium schools at a later date, using a bilingual form with each item presented in both English and Urdu.

Feedback from students during pre-piloting provided useful insights into further revisions required to the items themselves. For example, in Vietnam, student feedback indicated that the Vietnamese translation of 'bias' had not adequately captured the intended meaning in English (which affected students' ability to respond to two critical-thinking items on identifying potential bias), and so the translation of these items was revised. Meanwhile, observations during piloting in India revealed that a combined transferable skills test which included both problem-solving and critical-thinking items was too long for students to complete in an hour – many students required 75 minutes or more to complete both parts of the test, particularly in state government schools and those in rural areas. It was therefore decided to administer the transferable skills test as separate test forms in India; in the main survey, half the students in each class were randomly assigned either the problem-solving test or the critical-thinking test to complete. In both countries, student feedback generally indicated that the content of the problem-solving and critical-thinking exercises was appropriate to their contexts. There were some concerns in urban areas that students in rural areas would be unfamiliar, for example, with the bus map presented in the problem-solving exercise 'Journey', but students in rural areas did not raise these concerns themselves.

Due to the relatively small pilot sample sizes in the two countries, item-level pilot data analysis focused on percentage correct to indicate the difficulty of each item, and the functioning of distractors on each item (i.e. what the selection of a response tells us about how students of different ability levels understand an item). These item-level statistics were used to inform the selection of items for the final test forms in India and Vietnam. The final problem-solving and critical-thinking items selected for the main survey in each country are outlined in Table 2.

Table 2. *Problem-solving and critical-thinking items selected for main survey*

Country	Skill	Exercise	No. items	No. comprehension items	Total
India	Problem solving	Extra Classes	4	2	11
		Energy Needs	3	1	
		Journey	4	2	
	Critical thinking	Critical Reading and Evaluation: Student Fatigue	7	2	11
		Critique an Argument: Languages in School	4	1	
Vietnam	Problem solving	Extra Classes	4	2	12
		Energy Needs	3	1	
		Journey	4	2	
		Irrigation	1	0	
	Critical thinking	Critical Reading and Evaluation: Student Fatigue	7	2	11
		Critique an Argument: Languages in School	4	1	

6. Test reliability and cross-country calibration

The transferable skills items were administered during data collection at the end of the academic year (Wave 2) in India and Vietnam. The problem-solving and critical-thinking items were completed as one transferable skills test by 7,792 students in Vietnam, and as separate problem-solving and critical-thinking tests by 4,070 students and 4,058 students respectively in India.

6.1. Test validity and reliability

6.1.1. Individual country scales

In light of our conceptualisation of problem solving and critical thinking as two distinct skills – and additionally, since the problem-solving and critical-thinking items were administered as separate tests in India – we investigated the validity and reliability of the tests by conducting exploratory factor analysis (EFA) and generating the Cronbach's alpha separately for problem-solving items and critical-thinking items. Exploratory factor analysis is a method of modelling the covariance structure of observed variables in terms of how they relate to latent factors or traits (Hayashi and Yuan 2010), while Cronbach's alpha is a means of quantifying the reliability of a measure to capture one underlying trait (Multon and Coleman 2005).

When using exploratory factor analysis, eigenvalues greater than 1 are commonly used to determine the number of underlying factors or traits which are being measured from a set of items administered to survey participants (Hinkin et al. 1997). Exploratory factor analysis results for problem solving and critical thinking in India and Vietnam show that the eigenvalues for Factors 1-3 are all greater than 1, indicating that items in both the problem-solving and critical-thinking tests are measuring more than one underlying trait (see Appendix 1). However, Factor 1 accounts for the largest proportion of variance in the both tests for both India and Vietnam.

Table 3 reports the results from the Cronbach's alpha analysis of the individual country analysis conducted for the transferable skills tests. An alpha value of at least 0.7 or above is generally accepted to indicate that a measure is reliably capturing one underlying trait (Tavakol and Dennick 2011). Our results indicate that the problem-solving items in both India and Vietnam had a stronger average correlation in explaining problem-solving abilities as one underlying trait than the critical-thinking items. However, since the alpha values for both tests in the two countries are below 0.7, this suggests that they did not necessarily capture a single underlying trait.

Table 3. *Problem solving and critical thinking, individual countries – Cronbach's alpha*

Test	India	Vietnam
Problem solving	0.6115	0.6096
Critical thinking	0.3552	0.4480

6.1.2. Cross-country scales

As indicated in Table 2, the final test forms in India and Vietnam were almost identical in terms of the items included, with one additional problem-solving unit (Irrigation) included on the Vietnam test. This meant that it was possible to consider cross-country analysis of the transferable skills data, based on a common cross-country scale. To determine whether such analysis would be valid – that is whether we measured a single problem-solving construct and a single critical-thinking construct across the two countries – we conducted exploratory factor analysis and generated the Cronbach's alpha for problem-solving and critical-thinking data from both countries.

Exploratory factor analysis for the cross-country scale indicated similar results to the individual country scales. For the cross-country problem-solving scale, the eigenvalues for Factors 1-3 were greater than 1, with Factor 1 accounting for the largest proportion of variance. For the cross-country critical-thinking scale, the eigenvalues for Factors 1-4 were greater than 1, with Factor 1 again accounting for the largest proportion of variance (see Appendix 1). Meanwhile, the alpha value for the cross-country problem-solving scale indicated a stronger average correlation in explaining problem-solving abilities compared to the critical-thinking items. However, as with the individual country scales, alpha values below 0.7 suggest that the tests did not necessarily capture one underlying trait (see Table 4).

Table 4. *Problem solving and critical thinking, cross-country scale – Cronbach's alpha*

Test	Cross-country scale
Problem solving	0.6488
Critical thinking	0.4097

6.2. Validation of cross-country scales

Following exploratory factor analysis and the generation of Cronbach's alpha for each test, we proceeded to explore the functioning of individual problem-solving and critical-thinking items on a cross-country scale, using Item Response Theory (IRT). The results of this analysis are presented after a brief introduction to IRT.

6.2.1. Item Response Theory (IRT)

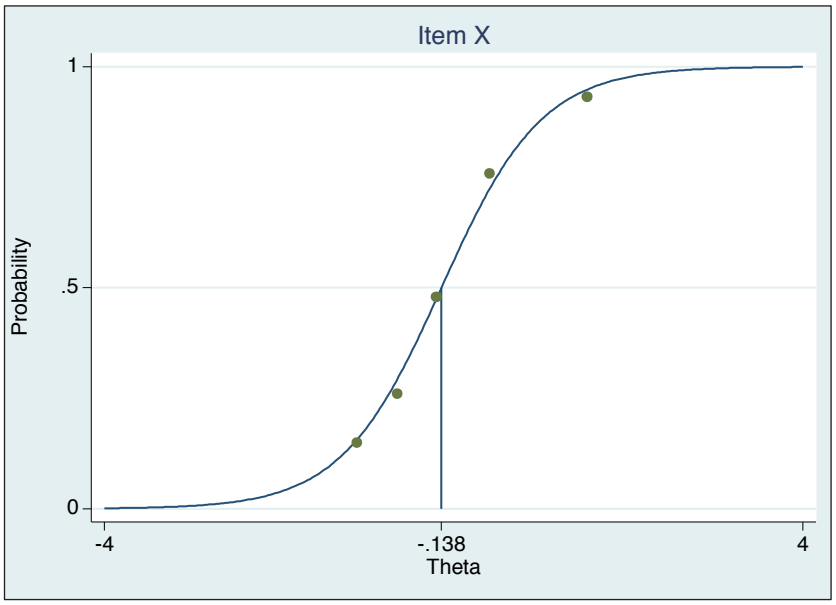
Item Response Theory provides the basis for modelling the relationship between latent traits (unobservable attributes) and their manifestations (observed outcomes or performance). IRT is used in the comparison of tests and similar instruments whose purpose is to measure unobservable characteristics of the respondents. It assumes that as the trait level increases, so does the probability of a correct response to the test item (see Edelen and Reeve 2007).

IRT is typically used to model a latent trait of individuals from a test or instrument in cases where the item score is dichotomous, being either correct or incorrect. For tests or instruments that follow polytomous scoring patterns, where responses may be scored 0 (incorrect), 1 (partially correct) or 2 (correct), we use the Partial Credit Models (PCM). The PCM is an IRT model that can be applied in a situation where responses on an item are recorded in two or more ordered categories, and there is an intention to combine results across items to obtain measures on some underlying variable (Masters and Wright 1997).

The item characteristic curve (ICC) is the basis of IRT. It yields a trace line (s-shaped) that is described by the difficulty and the slope (discrimination) of the item. Figure 1 provides an illustration of an ICC. The dots on the curve represent the mean quintiles of students' estimated ability/trait of interest; the higher up on the curve a student is, the higher the

probability of getting item X correctly and the higher the students' estimated ability (theta). The point at which a student has a 50 per cent chance of getting item X correctly is the estimated difficulty of item X (-0.138) and the median student's estimated ability (theta).

Figure 1. *Example item characteristic curve*



Item parameters in IRT are:

- Discrimination (a): The slope parameter is also known as the discrimination of the item, and it represents the slope of the ICC at the difficulty level. This slope also indicates the extent to which the item is related to the underlying construct and the ability of a test item to distinguish between individuals of differing ability/knowledge. A steeper slope indicates a closer relationship to the construct and therefore it is more discriminating.
- Difficulty (b): The difficulty level (also known as location parameter) is simply defined as the point on the ICC at which the probability of a positive response to the item is 50 per cent. The location parameter is usually between -2 and 2, with a mean of around zero. The higher the location parameter, the more difficult the item, and a respondent must have a higher ability (the measured construct) to answer that item correctly.
- Guessing or pseudo-guessing (c): an estimate of the probability with which a student with no knowledge of the item can obtain a correct response.

There are three types of IRT models:

- One parameter logistic (1-PL) model is the simplest form of IRT models, which predicts the probability of giving the correct response to an item as a function of the respondent's ability and the difficulty of an item. The discrimination parameter in a 1-PL model is fixed for all items.
- Two parameter logistic (2-PL) model predicts the probability of giving the correct response to an item as a function of the respondent's ability, item difficulty and the discrimination of the item. The discrimination parameter is allowed to vary between items.

- Three parameter logistic (3-PL) assumes that the probability of an individual getting an item right is dependent on the three factors above; item discrimination, item difficulty and guessing.

For the purpose of this analysis, the IRT two-parameter (2-PL) model was used to model students' responses on the critical-thinking test, while IRT PCM was used to model students' responses on the problem-solving test.[5]

6.2.2. Differential item functioning and cross-country calibration process

In order to determine whether items favoured students in either India or Vietnam, we examined differential item functioning (DIF) by generating item characteristic curves (ICCs) with quartiles of mean performance in each country, to provide a visual indication of whether students of the same ability in the two countries were equally likely to answer each question correctly. Figure 2 presents the ICCs for all problem-solving items, with India and Vietnam scaled together. Four items were identified as potential candidates for DIF (items 4, 6, 8 and 9), and so these items were split into individual country items (i.e. item 4 became items 4i and 4v, and so on), and the data were rescaled. For critical thinking (ICCs presented in Figure 3), items 4, 6, 7, 8 and 11 were identified as items that had either not functioned well on the cross-country scale and/or as potential candidates for DIF, so these items were split into individual country items, and the data were rescaled.

As evident from Figures 2 and 3, the ICC curves for several items (e.g. critical-thinking items 4, 6 and 7) had a very poor fit to the IRT model; in some cases the item difficulty is very high, and the slope of the curves are essentially flat, which indicates that the items discriminate poorly between students of differing abilities. Given that problem solving and critical thinking are potentially less easily defined as single underlying traits (for example, compared to maths ability), it is perhaps unsurprising that some items did not demonstrate a good fit to the IRT model. The fit of the curves to the IRT model were an additional consideration when deciding to split or drop specific items.

An iterative process was followed for both problem solving and critical thinking, in which ICCs were reviewed after rescaling, and further items were split and/or dropped for one or both of the countries. See Appendix 2 for the cross-country problem-solving and critical-thinking ICCs produced at the end of this calibration process.

5 The 3-PL IRT model was not used in this analysis as the design of the tests included distractor options in the multiple-choice options, which are intended to reduce the probability of guessing the correct response to an item. In light of this test design, the 2-PL model was felt to be more suitable.

Figure 2. *Cross-country problem-solving scale, all items – item characteristic curves*

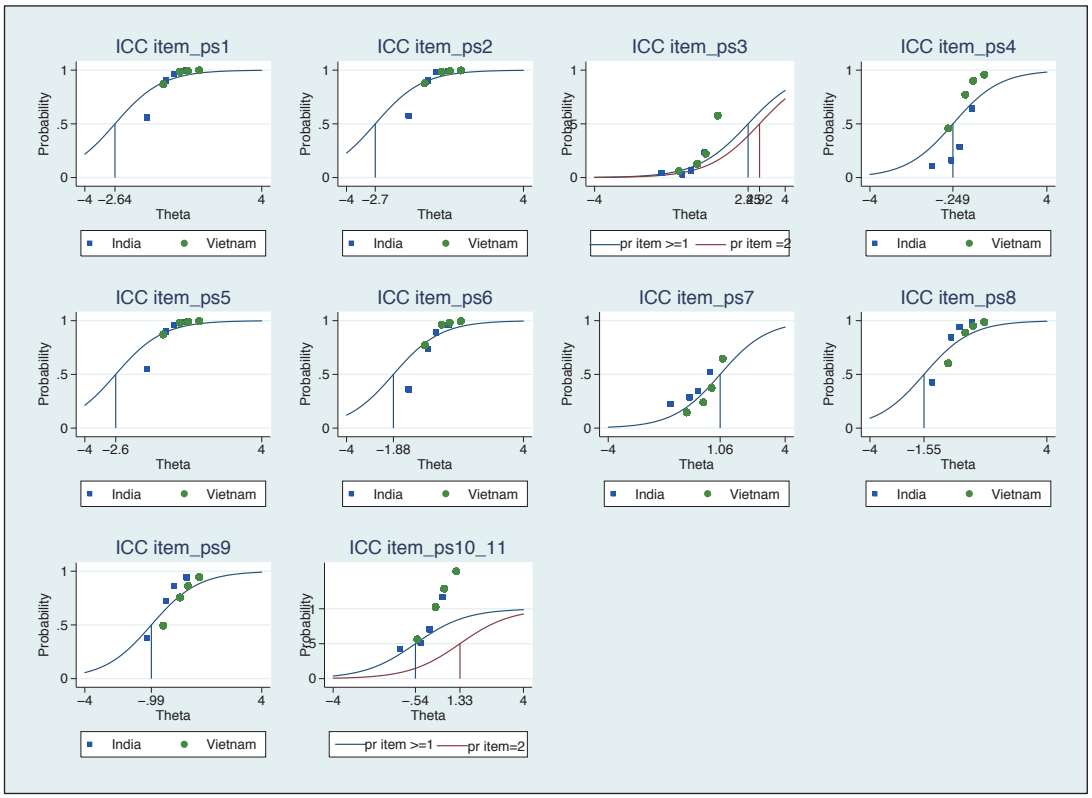
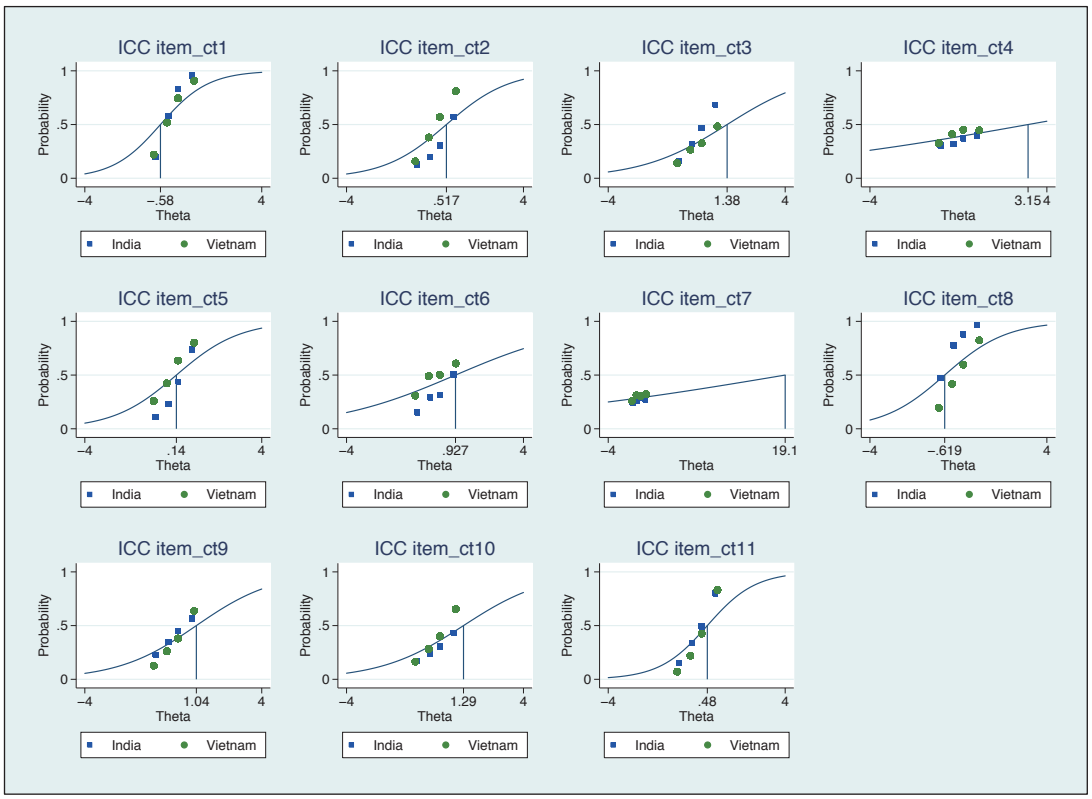


Figure 3. *Cross-country critical-thinking scale, all items – item characteristic curves*



7. Final test scales for problem solving and critical thinking

After the identification of differential item functioning and the cross-country calibration process described above, the cross-country scale for problem solving consisted of eight cross-country common items and three items split as individual country items, while the cross-country scale for critical thinking consisted of seven cross-country items and three items split as individual country items.[6]

While seven or eight common items are sufficient to support a common scale across the two countries, findings from the cross-country calibration process indicated that individual country scales for problem solving and critical thinking would be more valid than cross-country scales. For problem solving, five out of the eight cross-country items were comprehension items; that is, half the problem-solving items on the test functioned differently in India and Vietnam. For critical thinking, while four out of the seven cross-country common items assessed critical thinking, there were still indications of DIF on these common items, again indicating that the critical-thinking items had functioned differently in the two countries.

The validation of our problem-solving and critical-thinking scales indicated that the measures functioned better as individual country scales, rather than as cross-country scales. Exploratory factor analysis, Cronbach's alpha and IRT analysis suggested that revised individual country scales were more reliable in terms of measuring single constructs (see Table 5 for revised individual country scales, and Appendix 3 for results of exploratory factor analysis, Cronbach's alpha and IRT analysis for these scales). This may reflect differences in the problem-solving and critical-thinking constructs captured by our tests in India and Vietnam, perhaps linked to cultural factors or different pedagogical approaches. While it is beyond the scope of this technical note to consider this in more depth, these differences present an important area of further study. In the meantime, future analysis of the Young Lives transferable skills data will examine problem solving and critical thinking in India and Vietnam separately, using the scales detailed in Table 5.

In light of the increasing global emphasis on the importance of transferable or twenty-first century skills, the Young Lives transferable skills tests were developed in order to provide much-needed evidence on the extent to which young people develop these skills at school in low and middle-income countries. The validation process presented in this technical note indicates that our tests provide fairly reliable assessments of problem solving and critical thinking in the Young Lives sites in India and Vietnam; moreover, they can be used as a basis for developing future assessments of these skills in other low and middle-income contexts.

The Young Lives transferable skills tests will be available via the Young Lives website (www.younglives.org.uk) from December 2017, and Young Lives transferable skills data from India and Vietnam will be publicly available via the UK Data Archive (https://discover.ukdataservice.ac.uk/series/?sn=2000060) from February 2018.

6 Two poorly functioning critical thinking items were dropped from both countries, and one poorly functioning critical thinking item was dropped from Vietnam.

Table 5. *Final individual-country problem-solving and critical-thinking scales*

Test	Exercise	Skill assessed	Item no.	Answer key	Included in final scale?
India					
Problem solving	Extra Classes	Comprehension	1	D	Yes
		Comprehension	2	C	Yes
		Decision-making	3a	Yes	No*
			3b	No	
			3c	No	
			3d	No	
			3e	Yes	
			3f	Yes	
		Decision-making	4	C	Yes
	Energy Needs	Comprehension	5	A	Yes
		Decision-making	6	C	Yes
		Decision-making	7	B	Yes
	Journey	Decision-making	8	A	Yes
		Decision-making	9	D	Yes
		Decision-making	10	B	Yes**
			11	D	
Critical thinking	Critical Reading and Evaluation: Student Fatigue	Comprehension	1	C	Yes
		Comprehension	2	B	Yes
		Support a position	3	C	Yes
		Identify assumptions in arguments	4	D	No
		Identify connected and conflicted information	5	A	Yes
		Refute a position	6	B	Yes
		Evaluate the reliability of information provided	7	C	No
	Critique an Argument: Languages in School	Comprehension	8	C	Yes
		Address information that could strengthen or weaken the argument	9	D	Yes
		Evaluate alternate conclusions	10	A	No
		Evaluate the reliability of information provided	11	B	Yes
Vietnam					
Problem solving	Extra Classes	Comprehension	1	D	Yes
		Comprehension	2	C	Yes
		Decision-making	3a	Yes	Yes*
			3b	No	
			3c	No	
			3d	No	
			3e	Yes	
			3f	Yes	
		Decision-making	4	C	Yes
	Energy Needs	Comprehension	5	A	Yes
		Decision-making	6	C	Yes
		Decision-making	7	B	Yes
	Irrigation	Trouble-shooting	8a	No	Yes
			8b	Yes	
			8c	Yes	
	Journey	Decision-making	9	A	Yes
		Decision-making	10	D	Yes
		Problem solving	11	B	Yes**
			12	D	
Critical thinking	Critical Reading and Evaluation: Student Fatigue	Comprehension	13	C	Yes
		Comprehension	14	B	Yes
		Support a position	15	C	Yes
		Identify assumptions in arguments	16	D	No
		Identify connected and conflicted information	17	A	Yes
		Refute a position	18	B	No
		Evaluate the reliability of information provided	19	C	No
	Critique an Argument: Languages in School	Comprehension	20	C	Yes
		Address information that could strengthen or weaken the argument	21	D	Yes
		Evaluate alternate conclusions	22	A	Yes
		Evaluate the reliability of information provided	23	B	Yes

Notes: *Partial credit item: five out of six correct answers = partial credit; all correct answers = full credit.
**Partial credit item: one correct answer = partial credit; both correct answers = full credit.

References

Azubuike, O.B., Moore, R. and Iyer, P. (2017) *The design and development of cross-country Maths and English tests in Ethiopia, India and Vietnam*. Technical Note 39. Oxford: Young Lives.

Facione, P.A. (1990) *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction*. *Executive Summary*. Millbrae: California Academic Press.

Greiff, S., Holt, D.V. and Funke, J. (2013) 'Perspectives on problem solving in educational assessment: analytical, interactive, and collaborate problem solving', *The Journal of Problem Solving* 5 (2): 71-91.

Halpern, D.F. (1996) *Thought and Knowledge: An Introduction to Critical Thinking*. 3rd ed. Mahwah, NJ: L. Erlbaum Associates.

Hayashi, K. and Yuan, K. (2010) 'Exploratory Factor Analysis' in N.J. Salkind (ed) *Encyclopaedia of Research Design.* Sage. Available online: http://sk.sagepub.com/reference/researchdesign/n144.xml. [Accessed 24/04/2017].

Iyer, P., Azubuike, O.B. and Rolleston, C. (2017) *Young Lives School Survey 2016-17: Evidence from Ethiopia.* Oxford: Young Lives.

Iyer, P. and Moore, R. (2017) 'Measuring learning quality in Ethiopia, India and Vietnam: from primary to secondary school effectiveness', *Compare: A Journal of Comparative and International Education*. DOI: http://dx.doi.org/10.1080/03057925.2017.1322492.

Funke, J. (2010) 'Complex problem solving: a case for complex cognition?', *Cognitive Processes* 11: 133-142.

Hinkin, T.R., Tracey, J.B., and Enz, C.A. (1997) 'Scale construction: Developing reliable and valid measurement instruments', *Journal of Hospitality & Tourism Research*, 21(1): 100-120.

Lippman, L., Ryberg, R., Carney, R. and Moore, K.A. (2015) *Workforce connections: Key 'soft skills' that foster youth workforce success: towards a consensus across fields.* Bethesda, MD: Child Trends Publications.

Kuhn, D. (1999) 'A Developmental Model of Critical Thinking', *Educational Researcher* 28 (2): 16 – 25.

Marin, L.M. and Halpern, D.F. (2011) 'Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains', *Thinking Skills and Creativity* (6): 1-13.

Masters, G.N. and Wright B.D. (1997) 'The Partial Credit Model', in van der Linden, W.J. and Hambleton, R.K. (eds) *Handbook of Modern Item Response Theory*. New York: Springer.

Mayer, R.E. (2002) 'Rote versus meaningful learning', *Theory into Practice* 41 (2): 226 – 232.

Ministry of Human Resource Development (MHRD) (2016) *Some Inputs for Draft National Education Policy 2016*. New Delhi: MHRD, Government of India.

Moore, R., Azubuike, O.B., Reddy, P.P., Rolleston, C. and Singh, R. (2017) *Young Lives School Survey 2016-17: Evidence from India.* Oxford: Young Lives.

Multon, K.D. and Coleman, J.S.M. (2005) 'Coefficient Alpha' in N.J. Salkind (ed) *Encyclopaedia of Research Design.* Sage. Available online: http://sk.sagepub.com/reference/researchdesign/n53.xml. [Accessed 24/04/2017].

OECD (2004a) *PISA Problem Solving Items and Scoring Guides.* Paris: Organisation for Economic Co-operation.

OECD (2004b) *Problem solving for tomorrow's world: First measures of cross-curricular competencies from PISA 2003.* Paris: Organisation for Economic Co-operation and Development.

Partnership for 21st Century Learning (P21) (2015) *P21 framework definitions*. Washington, DC: P21.

Rankin, K., Cameron, D.B., Ingraham, K., Mishra, A., Burke, J., Picon, M., Miranda, J. and Brown, A.N. (2015) *Youth and transferable skills: an evidence gap map*. *3ie Evidence Gap Report 2.* New Delhi: International Initiative for Impact Evaluation (3ie).

Reef, J-P., Zabal, A. and Blech, C. (2006) 'The assessment of problem-solving competencies: a draft version of a general framework'. Available at: http://www.die-bonn.de/esprid/dokumente/doc-2006/reeff06_01.pdf.

Rossiter, J., Azubuike, O.B. and Rolleston, C. (2017) *Young Lives School Survey 2016-17: Evidence from Ethiopia.* Oxford: Young Lives.

Schendel, R. and Tolmie, A. (2017) 'Beyond translation: adapting a performance-task-based assessment of critical thinking ability for use in Rwanda', *Assessment and Evaluation in Higher Education* 42 (5): 673 – 689.

Scherer, R. and Beckmann J.F. (2014) 'The acquisition of problem solving competence: evidence from 41 countries that math and science education matters', *Large Scale Assessments in Education* 2 (10): 1 – 22.

Tavakol, M. and Dennick. R. (2011) 'Making sense of Cronbach's alpha', *International Journal of Medical Education* 2: 53 – 55.

Thomas, K. and Lok, B. (2015) 'Teaching Critical Thinking: An Operational Framework', in Davies, M. and Barnett, R. (eds.) *The Palgrave Handbook of Critical Thinking in Higher Education*, 99 – 106. New York: Palgrave MacMillan.

World Bank (2014) *Skilling up Vietnam: Preparing the workforce for a modern market economy.* Main report. Hanoi: Vietnam Development Information Centre.

World Bank (2015) *Project Appraisal Document for a Renovation of General Education Project, Vietnam.* Washington, DC: World Bank.

Appendix 1: Results of exploratory factor analysis for individual and cross-country scales

Individual country scales

Table 6. *Problem solving, India – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.59223	1.50720	0.2592	0.2592
Factor 2	1.08503	0.05084	0.1085	0.3677
Factor 3	1.03419	0.03885	0.1034	0.4711
Factor 4	0.99534	0.02473	0.0995	0.5707
Factor 5	0.97061	0.08664	0.0971	0.6677
Factor 6	0.88397	0.12249	0.0884	0.7561
Factor 7	0.76148	0.11657	0.0761	0.8323
Factor 8	0.64491	0.03271	0.0645	0.8968
Factor 9	0.61220	0.19215	0.0612	0.9580
Factor 10	0.42005	.	0.0420	1.0000

Table 7. *Problem solving, Vietnam – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.14321	0.99644	0.1948	0.1948
Factor 2	1.14677	0.09067	0.1043	0.2991
Factor 3	1.05610	0.09047	0.0960	0.3951
Factor 4	0.96563	0.02139	0.0878	0.4829
Factor 5	0.94424	0.01795	0.0858	0.5687
Factor 6	0.92628	0.01559	0.0842	0.6529
Factor 7	0.91070	0.08391	0.0828	0.7357
Factor 8	0.82679	0.04312	0.0752	0.8109
Factor 9	0.78367	0.04691	0.0712	0.8821
Factor 10	0.73676	0.17691	0.0670	0.9491
Factor 11	0.55985	.	0.0509	1.0000

Table 8. *Critical thinking, India – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.68741	0.58286	0.1534	0.1534
Factor 2	1.10455	0.07470	0.1004	0.2538
Factor 3	1.02985	0.00737	0.0936	0.3474
Factor 4	1.02248	0.05405	0.0930	0.4404
Factor 5	0.96843	0.02399	0.0880	0.5284
Factor 6	0.94445	0.00771	0.0859	0.6143
Factor 7	0.93673	0.04499	0.0852	0.6994
Factor 8	0.89175	0.03237	0.0811	0.7805
Factor 9	0.85937	0.04035	0.0781	0.8586
Factor 10	0.81903	0.08308	0.0745	0.9331
Factor 11	0.73594	.	0.0669	1.0000

Table 9. *Critical thinking, Vietnam – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.82342	0.73376	0.1658	0.1658
Factor 2	1.08966	0.06362	0.0991	0.2648
Factor 3	1.02604	0.04211	0.0933	0.3581
Factor 4	0.98393	0.01335	0.0894	0.4475
Factor 5	0.97058	0.06262	0.0882	0.5358
Factor 6	0.90796	0.01318	0.0825	0.6183
Factor 7	0.89478	0.03326	0.0813	0.6997
Factor 8	0.86152	0.00116	0.0783	0.7780
Factor 9	0.86036	0.04090	0.0782	0.8562
Factor 10	0.81946	0.05715	0.0745	0.9307
Factor 11	0.76231	.	0.0693	1.0000

Cross-country scales

Table 10. *Problem solving, India and Vietnam – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	2.14321	0.99644	0.1948	0.1948
Factor 2	1.14677	0.09067	0.1043	0.2991
Factor 3	1.05610	0.09047	0.0960	0.3951
Factor 4	0.96563	0.02139	0.0878	0.4829
Factor 5	0.94424	0.01795	0.0858	0.5687
Factor 6	0.92628	0.01559	0.0842	0.6529
Factor 7	0.91070	0.08391	0.0828	0.7357
Factor 8	0.82679	0.04312	0.0752	0.8109
Factor 9	0.78367	0.04691	0.0712	0.8821
Factor 10	0.73676	0.17691	0.0670	0.9491
Factor 11	0.55985	.	0.0509	1.0000

Table 11. *Critical thinking, India and Vietnam – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor 1	1.70054	0.58376	0.1546	0.1546
Factor 2	1.11678	0.08103	0.1015	0.2561
Factor 3	1.03576	0.03521	0.0942	0.3503
Factor 4	1.00054	0.03063	0.0910	0.4412
Factor 5	0.96991	0.01819	0.0882	0.5294
Factor 6	0.95173	0.07265	0.0865	0.6159
Factor 7	0.87908	0.00957	0.0799	0.6958
Factor 8	0.86951	0.02724	0.0790	0.7749
Factor 9	0.84227	0.01099	0.0766	0.8515
Factor 10	0.83129	0.02871	0.0756	0.9270
Factor 11	0.80258	.	0.0730	1.0000

Appendix 2: Item characteristic curves for cross-country problem-solving and critical-thinking scales

Problem solving

Figure 4. *Problem-solving items 1 - 4v*

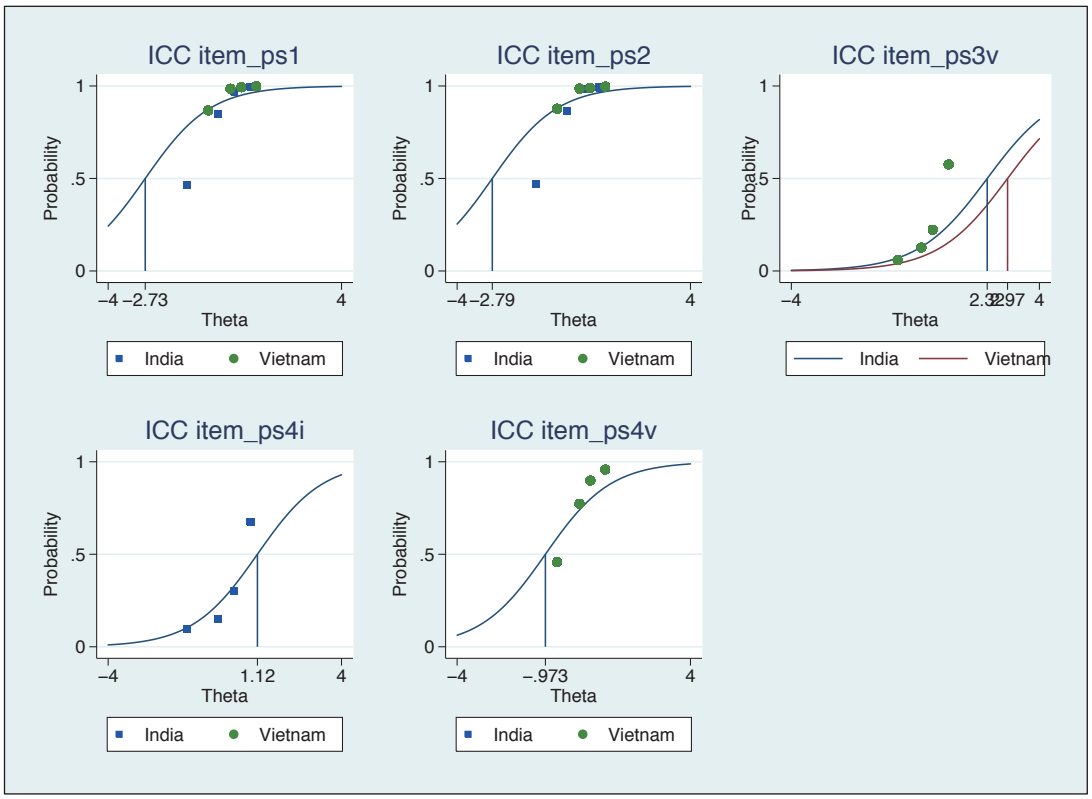


Figure 5. *Problem-solving items 5 - 8*

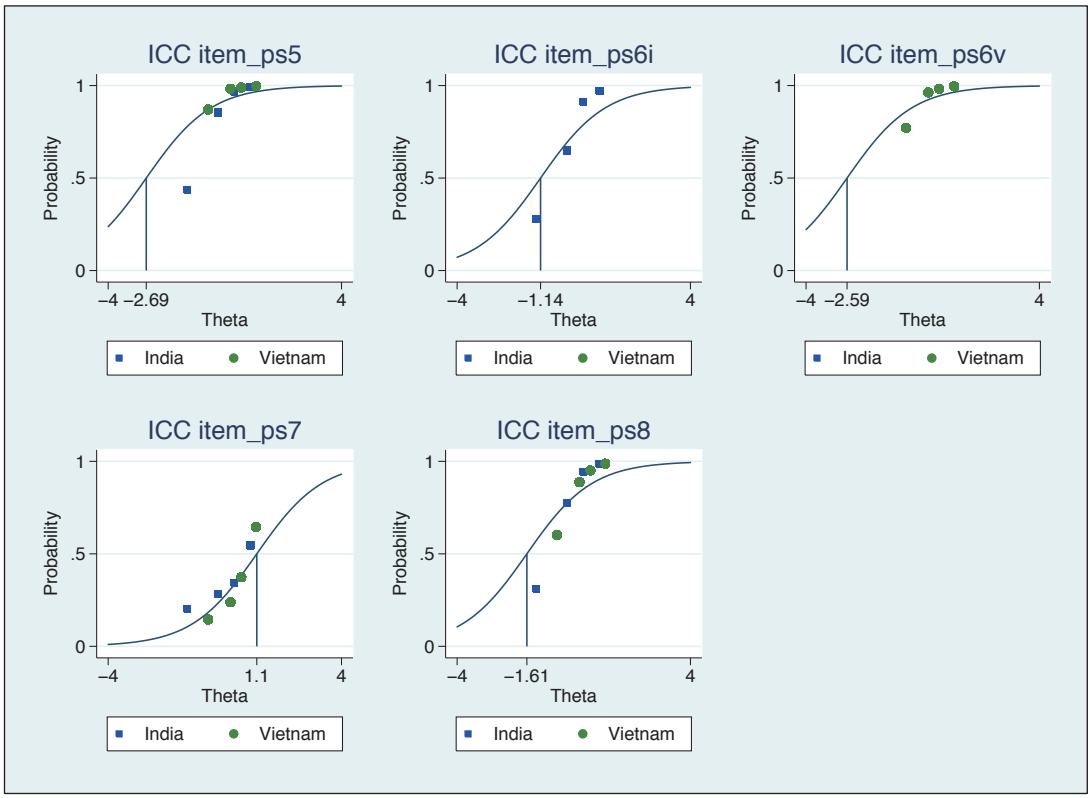
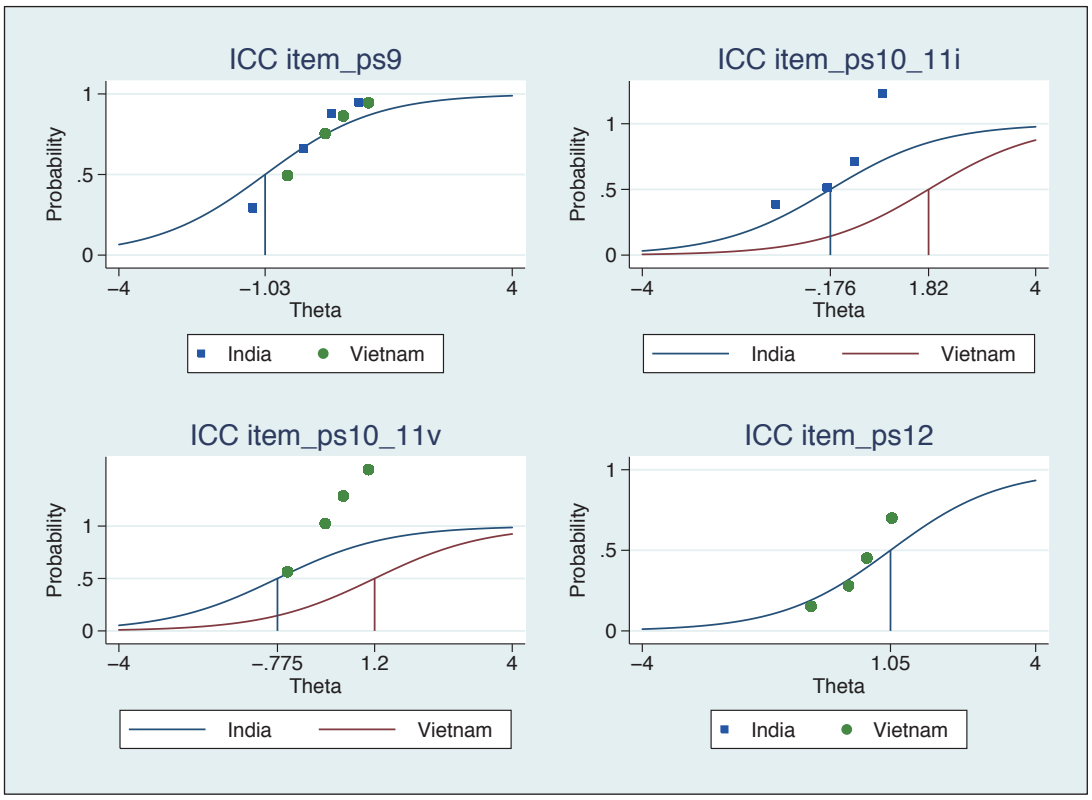


Figure 6. *Problem-solving items 9 - 12*



Critical thinking

Figure 7. *Critical-thinking items 1 - 6i*

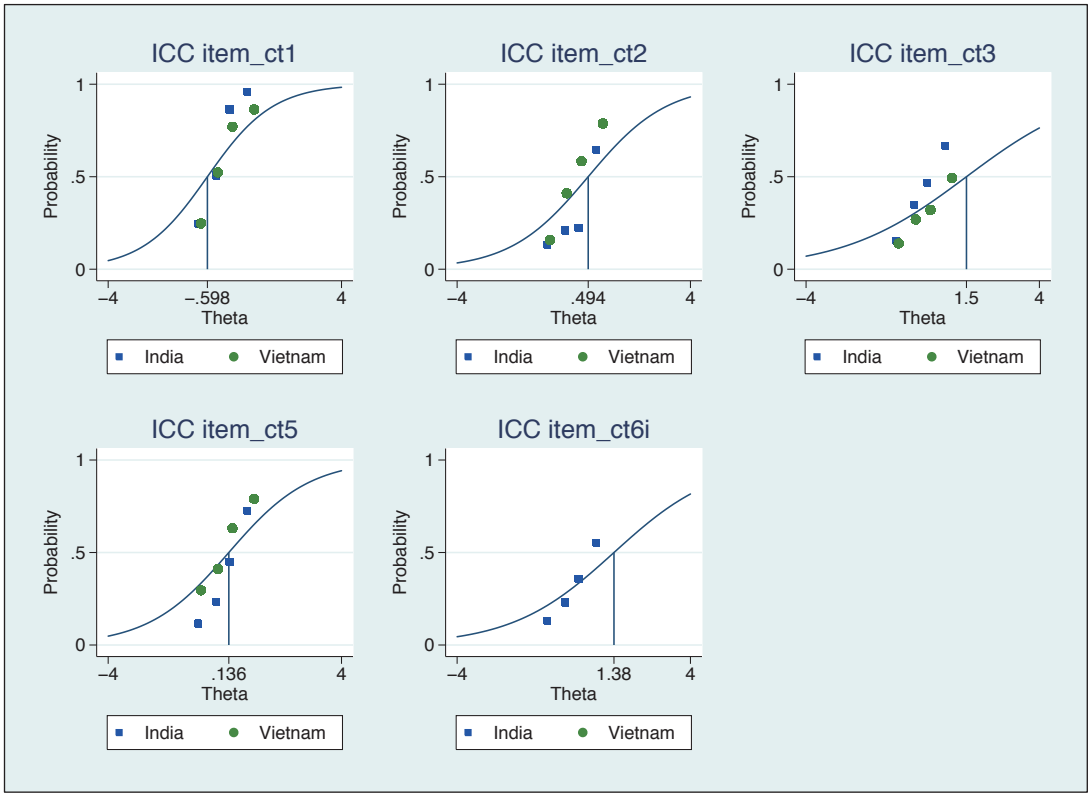
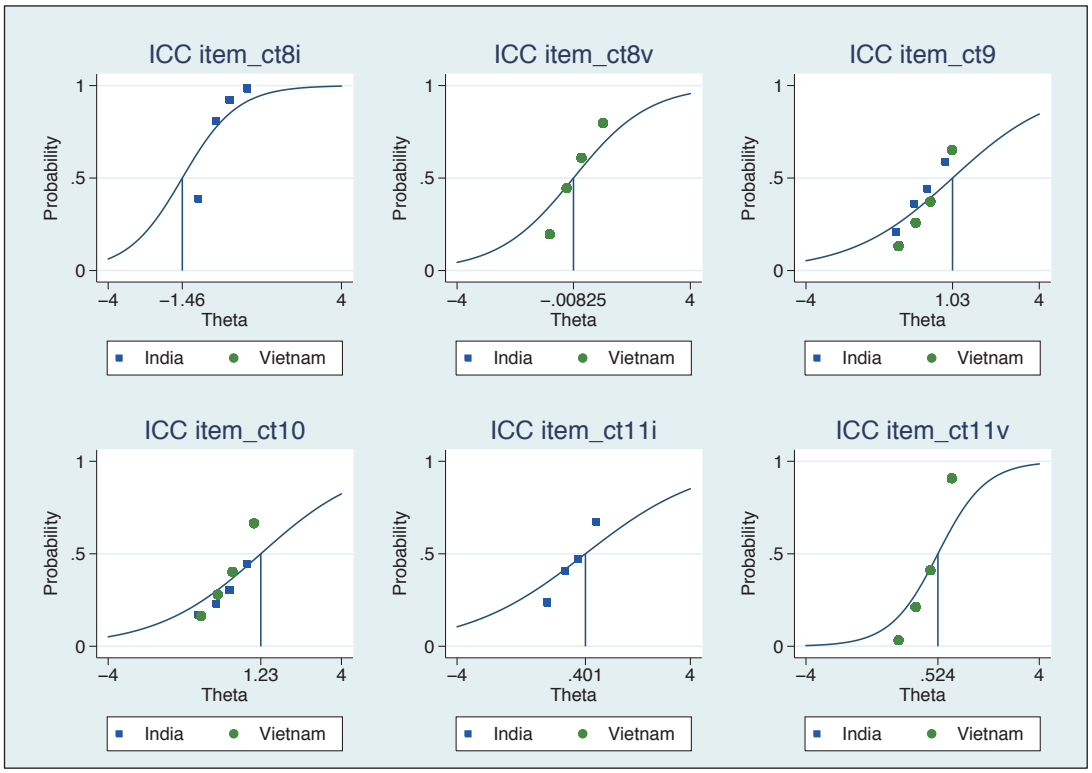


Figure 8. *Critical-thinking items 8i - 11v*



Appendix 3: Final individual country problem-solving and critical-thinking scales

India

Table 12. *Final problem-solving scale, India – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	2.58459	1.50671	0.2872	0.2872
Factor2	1.07787	0.04500	0.1198	0.4069
Factor3	1.03287	0.06149	0.1148	0.5217
Factor4	0.97138	0.08060	0.1079	0.6296
Factor5	0.89078	0.12858	0.0990	0.7286
Factor6	0.76220	0.11461	0.0847	0.8133
Factor7	0.64759	0.03493	0.0720	0.8853
Factor8	0.61266	0.19260	0.0681	0.9533
Factor9	0.42006	.	0.0467	1.0000

Table 13. *Final critical-thinking scale, India – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.67917	0.63245	0.2099	0.2099
Factor2	1.04672	0.06863	0.1308	0.3407
Factor3	0.97809	0.03095	0.1223	0.4630
Factor4	0.94714	0.02647	0.1184	0.5814
Factor5	0.92067	0.05537	0.1151	0.6965
Factor6	0.86530	0.04489	0.1082	0.8046
Factor7	0.82041	0.07792	0.1026	0.9072
Factor8	0.74249	.	0.0928	1.0000

Table 14. *Final problem-solving and critical-thinking scales, India – Cronbach's alpha*

Test	Cronbach's alpha
Problem solving	0.6340
Critical thinking	0.4225

Figure 9. *Final problem-solving scale, India – item characteristic curves*

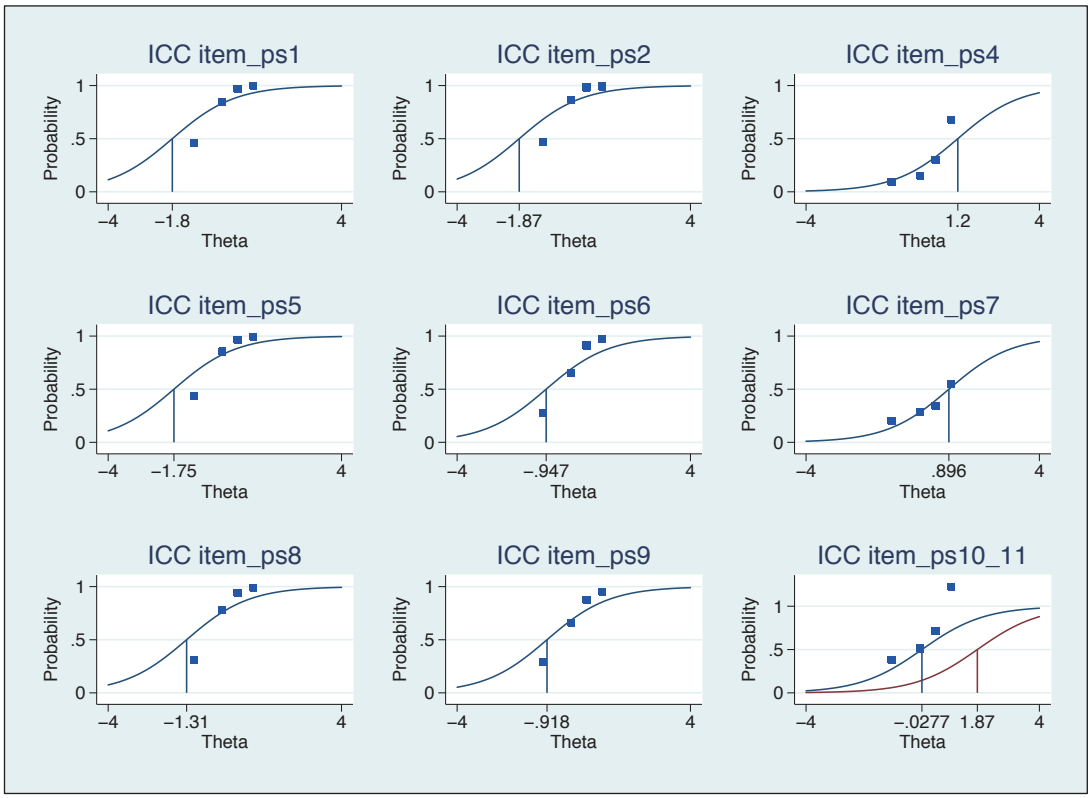
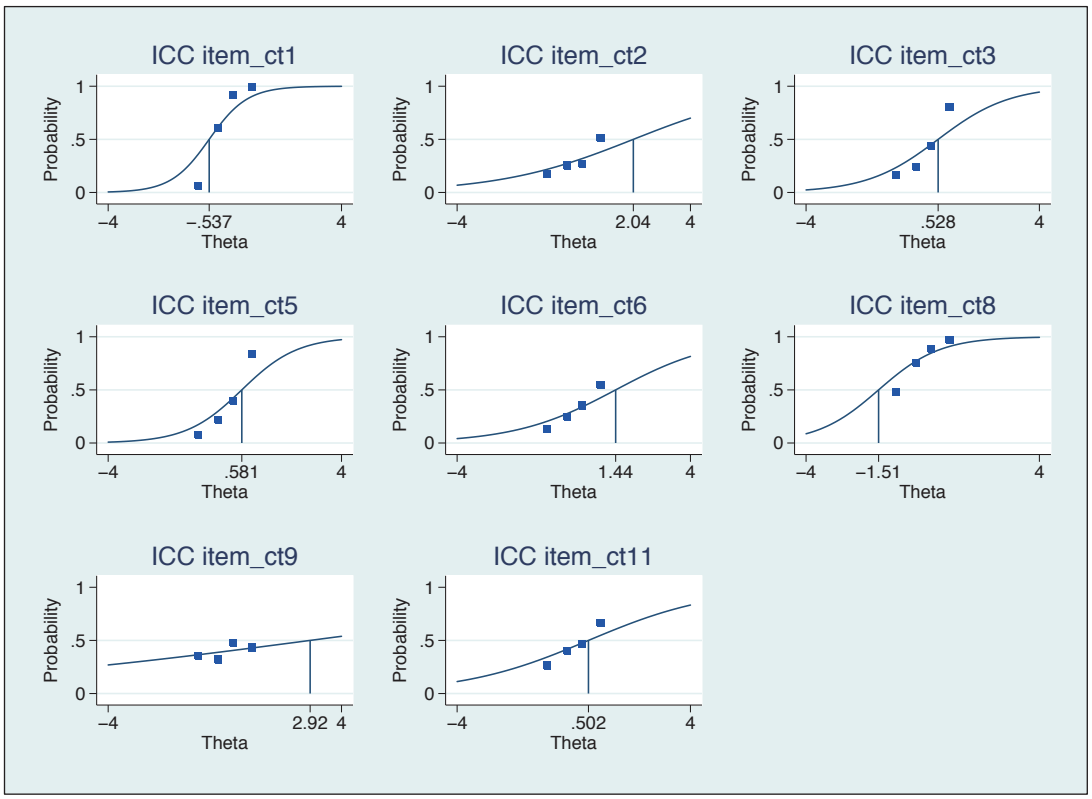


Figure 10. *Final critical-thinking scale, India – item characteristic curves*



Vietnam

Table 15. *Final problem-solving scale, Vietnam – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	2.14321	0.99644	0.1948	0.1948
Factor2	1.14677	0.09067	0.1043	0.2991
Factor3	1.05610	0.09047	0.0960	0.3951
Factor4	0.96563	0.02139	0.0878	0.4829
Factor5	0.94424	0.01795	0.0858	0.5687
Factor6	0.92628	0.01559	0.0842	0.6529
Factor7	0.91070	0.08391	0.0828	0.7357
Factor8	0.82679	0.04312	0.0752	0.8109
Factor9	0.78367	0.04691	0.0712	0.8821

Table 16. *Final critical-thinking scale, Vietnam – exploratory factor analysis*

Factor	Eigenvalue	Difference	Proportion	Cumulative
Factor1	1.79628	0.77848	0.2245	0.2245
Factor2	1.01780	0.06723	0.1272	0.3518
Factor3	0.95057	0.04496	0.1188	0.4706
Factor4	0.90561	0.02610	0.1132	0.5838
Factor5	0.87951	0.01619	0.1099	0.6937
Factor6	0.86331	0.03997	0.1079	0.8016
Factor7	0.82335	0.05977	0.1029	0.9046
Factor8	0.76358	.	0.0954	1.0000

Table 17. *Final problem-solving and critical-thinking scales, Vietnam – Cronbach's alpha*

Test	Cronbach's alpha
Problem solving	0.6096
Critical thinking	0.4955

Figure 11. *Final problem-solving scale, Vietnam – item characteristic curves*

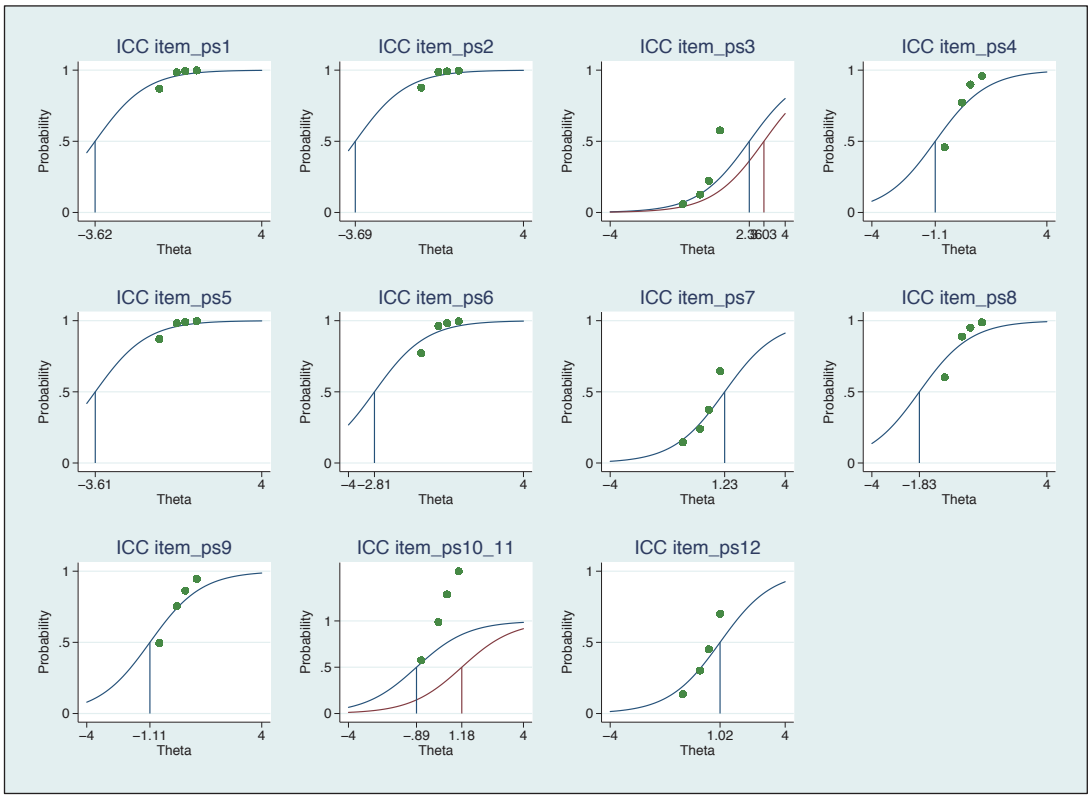
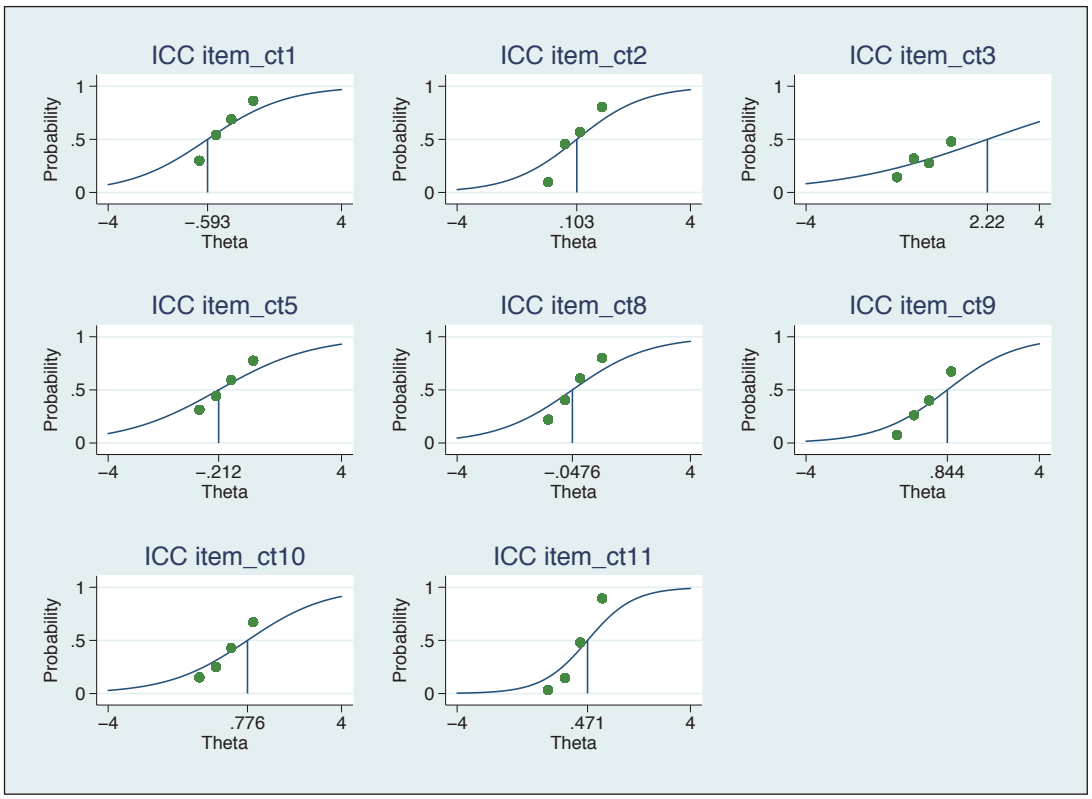


Figure 12. *Final critical-thinking scale, Vietnam – item characteristic curves*



Young Lives School Surveys 2016–17: The Design and Development of Transferable Skills Tests in India and Vietnam

In 2016-17, Young Lives school surveys examined school effectiveness at upper primary level in Ethiopia, and at secondary level in India and Vietnam. The surveys examined school effectiveness through multiple outcome measures, including students' learning progress in maths and English. This involved the administration of two linked maths tests and two linked English tests, administered at the beginning and the end of the school year (Wave 1 and Wave 2 of data collection respectively) In addition to assessing learning progress in maths and English, the 2016-7 school surveys included a transferable skills test, administered at Wave 2 in India and Vietnam.

This technical note outlines the design and development of this cognitive test, which assessed students' problem-solving and critical-thinking skills. After a brief introduction to the literature around the development of transferable skills, the note outlines the ways in which problem solving and critical thinking were conceptualised and measured in the 2016-17 school survey. The process of developing, piloting and selecting items for the test in India and Vietnam is then described, followed by results of scale validation using transferable skills test data from the main survey.

Young Lives

An International Study of Childhood Poverty

About Young Lives

Young Lives is an international study of childhood poverty, involving 12,000 children in four countries over 15 years. It is led by a team in the Department of International Development at the University of Oxford in association with research and policy partners in the four study countries: Ethiopia, India, Peru and Vietnam.

Through researching different aspects of children's lives, we seek to improve policies and programmes for children.

Young Lives Partners

Young Lives is coordinated by a small team based at the University of Oxford, led by Professor Jo Boyden.

- *Ethiopian Development Research Institute, Ethiopia*
- *Pankhurst Development Research and Consulting plc, Ethiopia*
- *Centre for Economic and Social Studies, Hyderabad, India*
- *Sri Padmavathi Mahila Visvavidyalayam (Women's University), Andhra Pradesh, India*
- *Grupo de Análisis para el Desarollo (GRADE), Peru*
- *Instituto de Investigación Nutricional (IIN), Peru*
- *Centre for Analysis and Forecasting, Vietnamese Academy of Social Sciences, Vietnam*
- *General Statistics Office, Vietnam*
- *Oxford Department of International Development, University of Oxford, UK*

Contact:
Young Lives
Oxford Department of International Development,
University of Oxford,
Mansfield Road,
Oxford OX1 3TB, UK
Tel: +44 (0)1865 281751
Email: younglives@younglives.org.uk
Website: www.younglives.org.uk

Young Lives