Technical Note 51

Young Lives

Equating Cognitive Scores across Rounds and Cohorts for Young Lives in Ethiopia, India, Peru and Vietnam

Juan Leon

www.younglives.org.uk

JUNE 2020

Equating Cognitive Scores across Rounds and Cohorts for Young Lives in Ethiopia, India, Peru and Vietnam

Juan Leon

First published by Young Lives in June 2020

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 18 years. **www.younglives.org.uk**

Young Lives is funded by UK aid from UK Government.

The views expressed are those of the author(s). They are not necessarily those of, or endorsed by, Young Lives, the University of Oxford, DFID or other funders.

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@qeh.ox.ac.uk

Contents

The authors

Juan León has a PhD in Educational Theory and Policy and Comparative and International Education from Pennsylvania State University, United States. He has a Bachelor's degree in Economics from the Pontifical Catholic University of Peru. Juan is a Senior Researcher at GRADE in Peru. He is also a Lecturer in the Department of Psychology at the Universidad Antonio Ruiz de Montoya in Lima, and a Lecturer in the Economics Department at the Pontifical Catholic University of Peru.

1. Introduction

For longitudinal studies such as Young Lives, getting comparable measures of children´s cognitive abilities over time is essential for identifying individual, family, school or contextual variables that affect children´s development. Few longitudinal studies that follow birth/age cohorts have comparable cognitive measures over time; of those that are available, most are from developed countries and there are almost none from developing countries. For example, studies such as The National Education Longitudinal Study (NELS), the Early Childhood Longitudinal Study – Kindergarten (ECLS-K), Education Longitudinal Study (ELS), and the Rochester Longitudinal Study in the United States have achievement measures (maths and reading comprehension) that are comparable across waves. To help fill this gap, this technical note outlines the statistical procedures that Young Lives has followed to achieve comparable measures across rounds and age cohorts in Ethiopia, India, Peru and Vietnam.

The note has five sections. After this brief introduction, we present a description of each cognitive or achievement test. Subsequent sections present the methodology of analysis and the key results. The final section provides some concluding thoughts on the main findings of the analysis.

2. The Young Lives assessments

Young Lives is a longitudinal study into childhood poverty that has tracked the development of 12,000 children over 15 years in Ethiopia, India (in the states of Andhra Pradesh and Telangana), Peru and Vietnam. Young Lives has been following two cohorts (an Older Cohort born in 1994 and a Younger Cohort in 2001) since the beginning of the study. In Peru, the original sample was chosen randomly from 20 sites across the country, with the richest 5 per cent of districts excluded from the sampling framework. To date, the study has carried out five rounds of data collection (in 2002, 2006, 2009, 2013 and 2016). Young Lives gathers information on children's and their families' access to services, economic indicators, and work patterns, and assessments of children's nutritional and educational measures.

In order to identify which variables such as children, family and school characteristics can affect children's development, it is necessary to have comparable measures of children´s cognitive abilities and achievement over time. Table 1 shows the cognitive and achievement measures administered in Young Lives by round and age cohort.

The only common test administered across rounds was the Peabody Picture of Vocabulary Test (PPVT). For maths and reading, different tests were used across rounds, but a set of common items were kept across rounds, items that were used as anchor items in order to get comparable measures for both areas across rounds. Since Round 4 the PPVT or its Spanish version (TVIP) has been administered only to the Younger Cohort. The main reason for this change was the ceiling effects observed for the Older Cohort in Round 3 (Cueto and Leon 2012). It was possible to have comparable measures across age cohorts and rounds for the PPVT (from Rounds 2 to 5), maths (Rounds 2 to 5) and reading comprehension (Rounds 4 and 5). Finally, it is important to point out that while we carefully estimated comparable scores for maths achievement across rounds and age cohorts, the tests and administration procedures used across rounds were different, introducing a source of noise in those scaled scores.

Table 1. *Measures of abilities and achievement administered in Young Lives*

Round	Cohort	Cognitive	Reading	Mathematics
Round 1	Younger Cohort	NA	NA	NA
	Older Cohort	Raven's Progressive Matrices for children	One item on reading One item on writing	One item on multiplication
Round 2	Younger Cohort	PPVT	NA	CDA
	Older Cohort	PPVT	One item on reading and one on writing	One multiplication item and maths test
Round 3	Younger Cohort	PPVT	One item on reading One item on writing Early Grade Reading Assessment (EGRA).	One multiplication item and maths test
	Older Cohort	PPVT	Cloze test of reading comprehension	Maths test
Round 4*	Younger Cohort	PPVT	Reading comprehension	Maths test
	Older Cohort	NA	Reading comprehension	Maths test
Round 5*	Younger Cohort	PPVT	Reading comprehension	Maths Test
	Older Cohort	NA	NA	NA

Notes: NA = Not administered. * Rounds 4 and 5 consider PVVT for only 125 items in Peru, unlike other countries where a sub-sample of the original 204 items was administered.

The cognitive and achievement measures administered in Young Lives include:

- **The Peabody Picture of Vocabulary Test:** This is a widely used test of receptive vocabulary, originally developed in English in 1959 and updated several times since. For Ethiopia, India and Vietnam, we used PPVT version III (Dunn and Dunn 1997); while for Peru, we used a Spanish adaptation of the PPVT called 'Test de Vocabulario en Imágenes Peabody' (TVIP) developed by Dunn et al. (1986). The English and Spanish versions of this cognitive test have been used by several research studies that have found a positive strong correlation between the PPVT and some commonly used intelligence measures, such as the Wechsler and McCarthy Scales (e.g. Campbell, Bell and Keith 2001; Gray et al. 1999; Campbell 1998).

  The test is administered individually, orally, untimed, and norm-referenced. The task of the test taker is to select the picture that best represents the meaning of a stimulus word presented orally by the examiner. Not all items in the test are expected to be administered. Instead, the examiner administers enough items to establish a ceiling and a baseline. The basal set rule is one or no errors in a set of 12 items, and the ceiling set rule is eight or more errors in a set of 12 items. Non-administered items below the baseline are automatically given a score of 1, given that they are expected to be easier, while items above the ceiling are given a score of 0, as they are more difficult. The raw score is formed by all the items given a score of 1 (i.e. answered correctly or below the basal item).

  In Ethiopia, India and Vietnam, since Round 4, some noise was added in the PPVT scores measures by the difficulty levels of individual items having changed in translation and often leading to a change in the order of the sets. Given that the PPVT is one of the few longitudinal measures available in the Young Lives data, it was worthwhile to try and keep it, albeit in a modified form. With this end in mind, a subset of items that seemed to have performed well in Rounds 2 and 3 were kept and administered again in Rounds 4 and 5 to the Younger Cohort. Also, after Round 4, the PPVT test stopping rule was

abandoned and all children answered all items. The criteria followed to select the subset of items within each country was: (i) Item Response Theory scores (3PL) were calculated for PPVT in Ethiopia (Amharic, Oromifa, and Tigrinya), India (Telugu) and Vietnam (Vietnamese), restricting the sample to the main languages in each country; (ii) ability cut-offs were identified to split the full sample (combined Older Cohort and Younger Cohort, Rounds 2 and 3) into four equal bands of ability; (iii) items were sorted within these four bands based on their difficulty parameters; (iv) the items were further sorted within these four bands based on their discrimination; (v) finally, a roughly equal number of items were selected across bins, inspecting the Item Characteristic Curves: specifically, items with bad fit, high guessing parameter or zero variation were excluded. This resulted in a specific subset of items for each country (55 in Ethiopia, 57 in India, and 76 in Vietnam) that were administered in Round 4. In Peru, no changes were introduced in the PPVT test administration.

- **Maths achievement:** From Round 2, Young Lives started to administer maths achievement tests to the Older Cohort, while the Younger Cohort has been tested since Round 3. As the maths achievement tests where different in each round, the Young Lives education team decided to keep a set of common items across rounds and age cohorts that allows us to equate maths scores across rounds and cohorts. Since Round 3, the competency measured in the maths achievement tests is number and number sense, since numeracy has a positive impact in individual decision-making processes (Bateman et al. 2007).
- **Reading comprehension:** Since Round 4, the education team has administered reading comprehension tests to both age cohorts in the four countries. The items included in the assessments measure reading literacy, from basic decoding to identifying underlying ideas from a narrative. As with the maths tests, a set of common items were kept across age cohorts and rounds in order to equate the reading comprehension scores. Like with maths, Young Lives decided to test reading comprehension since not only it is fundamental at school but also outside of school. Several studies have found that good reading comprehension skills are good predictors of social (e.g. poverty reduction) and individual benefits (e.g. higher income) in the long run (McMahon 1997; Wolfe and Zuvekas 1997).

3. Methodology

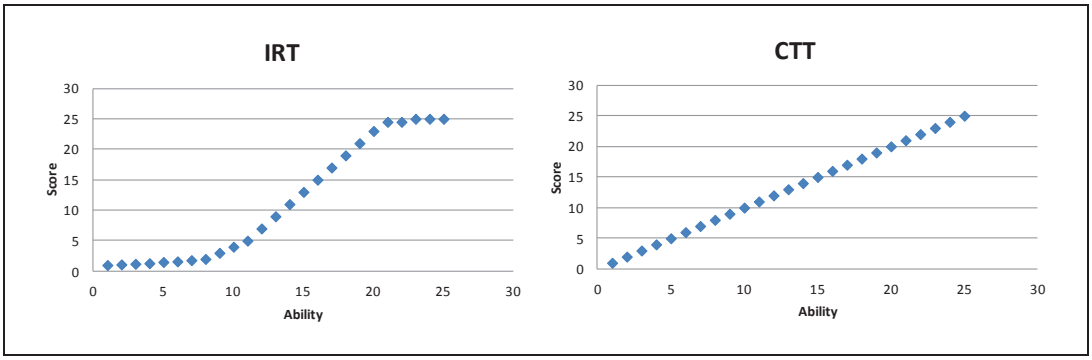
For the psychometric analysis of the tests, we used Modern Test Theory or Item Response Theory (IRT). This statistical technique is part of the latent traits theory since it attempts to explain a latent construct (e.g. maths knowledge) using observed outcomes (e.g. items).

3.1. Why IRT scores instead of CTT scores?

Unlike Classical Test Theory (CTT), IRT is more focused on the item rather than the test. Also, the standard error of measurement in IRT is a function of the ability of individuals, thus it varies at each level of ability, and nonetheless, the interpretation is the same. IRT estimates the probability of answering the item correctly through a logistic function based on the difference between the item difficulty and the individual´s ability. The idea is that individuals with higher ability will have a greater probability of answering easier items correctly than difficult ones.

Figure 1 shows the relationship between ability and score in CTT and IRT. In the case of CTT, the raw score increases in the same proportion as ability, thus it follows a linear and monotonic trend. In contrast, in IRT as ability increases the score does not increase in the same proportion, in other words the growth is nonlinear. This implies that, under CTT, the growth in the score is the same if ability changes from 10 to 15, or 20 to 25. However, under IRT it is not the same since it follows a different functional form that relies on the characteristics of the items.

Figure 1. *Functional form between scores and ability, by theory*



Therefore, we used IRT to build children's composite scores for PPVT in Ethiopia (Amharic, Oromifa and Tigrigna), India and Vietnam. The main advantages of using IRT instead of CTT are: (i) the principle of invariance – the item parameters do not depend on an individual's ability, being invariant over different samples of examinees, and an individual's ability does not depend on the items presented, being invariant over different samples of items; (ii) allowing for comparing individuals' ability from different populations if tested with instruments that have common items; and (iii) allocation of individuals' ability and item difficulty on the same scale or metric, creating an interval scale in logits for both scores. Thus, using this statistical technique, we were able to build comparable scores by cohort and round.

3.2. The IRT three-parameter model

IRT models rely on two main assumptions. First is the local independence assumption, which asserts that the probability of answering an item correctly depends on an individual's ability only and not his/her answer to other items. Second, the models assume unidimensionality, in other words, that only one latent trait is measurable across all items or at least one dominant factor is observed behind the set of items tested. Of these two assumptions, the latter is the more difficult to accomplish since different factors could be affecting individual performance (i.e. test anxiety) (see Cueto et al. 2009).

The model assumes that an individual's ability depends on three item parameters – item difficulty, item discrimination, and item guessing. Item difficulty refers to the proportion of individuals who get each item right. Item discrimination indicates how well an item can discriminate between high and low achievers. The guessing parameter refers to the chances that an individual has to get an item right. This parameter is mainly considered for multiple-choice tests since these allow examinees to guess.[1] These parameters and the individual´s ability level are part of the Item Characteristic Curve (ICC) that defines the probability that

1 As PPVT is a multiple-choice test it is necessary to consider a guessing parameter.

each individual has to get an item right given the item characteristics (difficulty, discrimination and guessing) and individual ability. The following equation represents the general ICC model:

$$P_i(\theta) = c_i + (1 - c_i)\frac{e^{a_i(\theta - b_i)}}{1 + e^{a_i(\theta - b_i)}} \qquad i = 1,2, \dots, n$$

$P_i(\theta)$: the probability that an individual with ability θ get right the item i

a_i : item discrimination

b_i : item difficulty

c_i : guessing parameter

n: number of items in the test

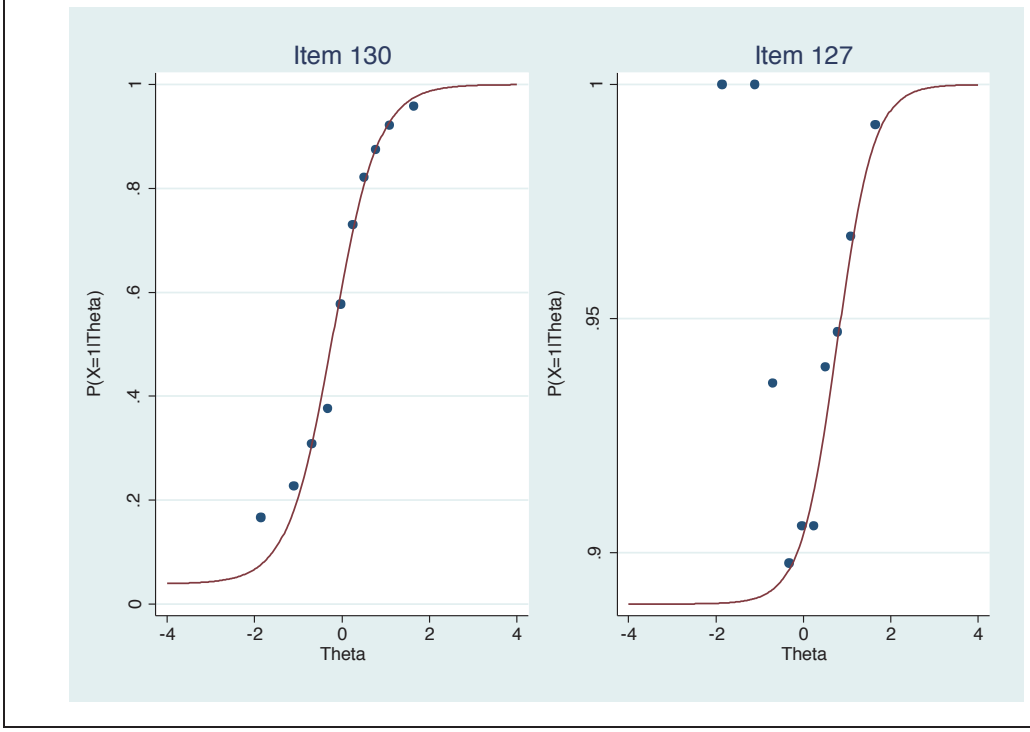
θ: individual´s ability parameter

The two-parameter model uses the same equation but assumes that the guessing parameter (c_i) is equal to zero, while the one-parameter model not only assumes a guessing parameter (c_i) of zero but also that the item discrimination (a_i) is constant across items.

3.3. Item fit

An item has *good fit* if the ICC shows that the proportion of children who answer an item correctly varies monotonically as a function of a child's ability. Figure 2 shows an example of an item with good fit (item 130) and one with poor fit (item 127). For the item with good fit the proportion of children who answer correctly varies monotonically with the average child´s ability, while the item with poor fit shows no correspondence between the proportion of children who answer correctly and the average child´s ability.

Figure 2. *Item Characteristic Curves of items with good and bad fit*

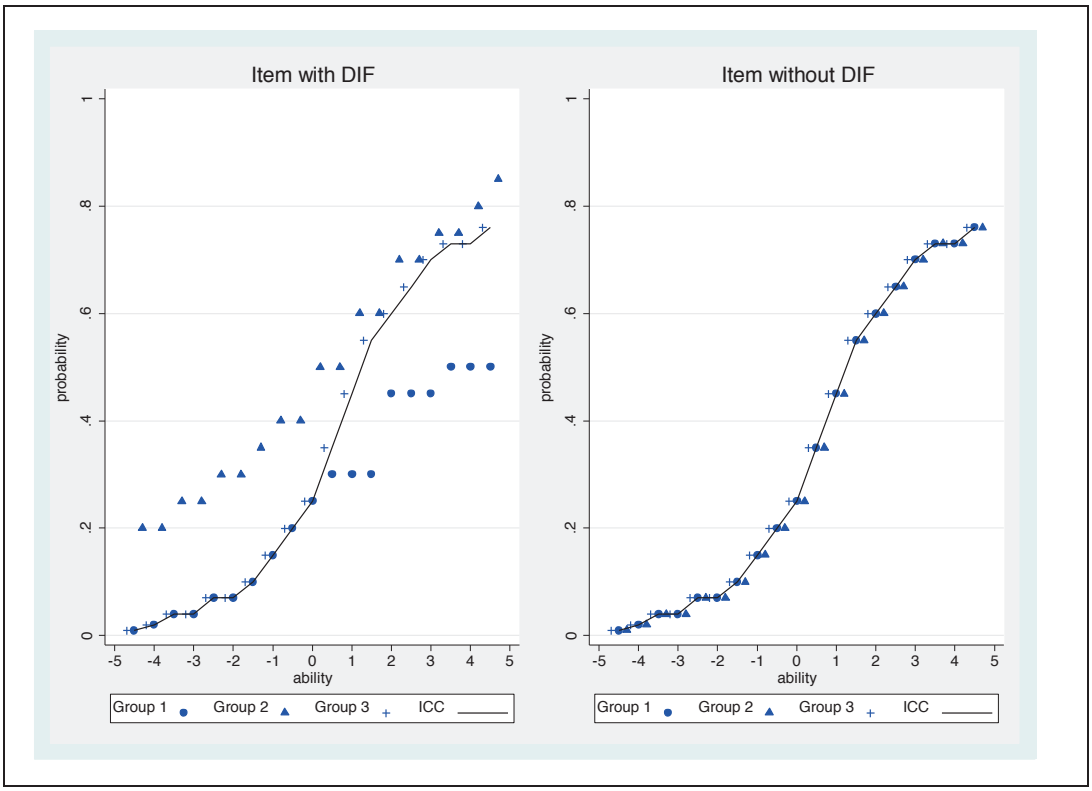


3.4. Differential Item Functioning (DIF)

An item is considered to have Differential Item Functioning (DIF) if the probability of answering an item correctly differs across groups or memberships (e.g. gender), controlling for level of ability (Hambleton and Swaminathan 1985; Dorans and Holland 1993; Linacre 2008). DIF analysis, however, could be sensitive to sample size since the standard errors of the item difficulty depend on the size of the groups that are being compared. Thus, large sample sizes could lead to accepting even small differences between item difficulties as DIF. Therefore, it is necessary to use normalised standard errors in order to have better estimates of DIF between groups. Both the Educational Testing Services in the United States and scholars (Wright and Douglas 1976) suggest that for large sample sizes logit differences in item difficulty above 0.50 signal DIF between groups.

For this note, we used two approaches to check item DIF. The first was graphically, in that we estimated the ICC for each item from the full sample and each group (cohort and round) that we want to check for DIF; then, an item was considered as DIF by each group if the ICC has a different shape than the ICC for the full sample, as the example in Figure 3 shows.

Figure 3. *Item Characteristic Curves of items with DIF and without DIF*



Our second approach was to calculate the Welch test using the one-parameter model, with an item flagged with DIF if the difference between item difficulties across groups was statistically different at 5 per cent according this test. Finally, in both analyses, we considered an item as having DIF across groups if the number of children who took the item was equal or above 30.

3.5. Scores equating

One of the main advantages of using IRT modelling is that it helps to build comparable scores using common items. Hambleton (1989) indicates that if we have different tests (common items across them) and the items in those tests meet the IRT assumptions (good item fit indicators), then it is possible to estimate a score for each individual that is independent of the group of items that he/she answered. Thus, it is possible to use those PPVT items with an adequate fit index as anchors in order to have a score that could be comparable across rounds and cohorts.

The main types of test equating are (Linacre 2008):

- *Common item equating:* There are different examinees but common items across all test forms. Two different type of analysis could be performed. First, the common and non-common items are analysed simultaneously (e.g. vertical equating). Second, common items across all test forms are analysed and calibrated in order to use them to adjust the mean and standard deviation of each test form.
- *Common person equating:* There are different tests of the same subject (e.g. maths) but common examinees across tests. The average ability of the common examinees is used to adjust examinees' mean and standard deviations.
- *Virtual item equating:* There are different examinees and different tests but both tests cover the same subject (e.g. maths). This type of equating involves identifying test pairs of items that cover the same subject and using them as pseudo-anchor items for the equating analysis.

For our analysis, we used the common item equating approach since we have the same test or set of items across cohort and rounds. It was not possible to use common person equating since having the scores of the same examinee at two different time points is similar to having different examinees.

The procedures followed for the equating analysis were: (i) run the three-parameter model for the pool sample;[2] (ii) identifying those items with poor item fit and deleting them from our analysis; (iii) identifying those items with DIF for all the groups and deleting them from the analysis; (iv) identifying those items with the presence of DIF and separating them into different items; and (v) running the three-parameter model again using as anchor items those with the absence of DIF by round and cohort. In the case of the sibling analysis, the IRT model was run using as fixed parameters those estimated in the analysis for the index children.

The analysis was carried out using the ado file *openirt*. This file was developed by Tristan Zajonc, who not only provided the STATA files to run the analysis but also technical assistance to interpret and improve the IRT analysis performed in STATA.[3]

2 For maths test scores, we used a two-parameter model.

3 All the item parameters for the different IRT models estimated are available in Appendix A. Appendix B contains the Item Characteristics Curves for each of the final IRT models estimated for each test. Appendix C shows all the DIF analysis performed on the final IRT model estimated for each test.

3.6. Limitations of IRT scores

One main limitation of IRT scores is that they are not comparable across languages for vocabulary and reading comprehension. IRT scores are specific for each language or country and each scale is independent from each other. This caveat emerged as the PPVT test used for Ethiopia, India and Vietnam is the English version and it was difficult to ensure item cognitive equivalence across languages. Instead, Young Lives ensured the comparability of the items within each language, in order to have PPVT IRT scores comparable across rounds and age cohorts.

Therefore, we need to be careful when using IRT scores for analysis and we need to be clear *what we could do* and *what we could not do* with these scores. For example, we could not use the IRT scores to compare the vocabulary, reading comprehension or maths achievement between children who took the test in Amharic and Spanish. Instead, we could use the IRT scores to compare the vocabulary, reading comprehension or maths achievement between children from the Younger and Older Cohort who took the test in Amharic or another main language.

4. Results

4.1. PPVT scores

We estimated a three-parameter IRT analysis for the pool sample for each of the main languages in the four countries: Ethiopia (Amharic, Tigrigna and Oromifa), India (Telugu), Peru (Spanish) and Vietnam (Vietnamese). This first analysis performed allowed us to identify those items with poor fit that should be dropped from each of the composite scores. Table 1 shows the percentage of items that were dropped because of poor item fit or DIF for all the comparison groups (round and cohort). The percentage of items dropped, on average, was less than one third of the total items, the Oromifa language had the highest percentage (33 per cent) of items dropped, and Peru was the only country where no item was dropped.

Table 1. *Number of items dropped by language*

Language	Total items	Items dropped	%
Amharic	204	29	14
Tigrinya	204	65	32
Oromifa	204	67	33
Telugu	204	44	22
Spanish	125	0	0
Vietnam	204	33	16

Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

However, one main concern was the number of common items or anchor items dropped between Round 4 and the previous rounds. Table 2 shows that the percentage of anchor items dropped was less than 3 per cent. These results indicate that we have enough anchor items to ensure adequate equating across rounds and age cohorts. Those items that have a good item fit but have DIF were considered as a different item.

Table 2. *Number of anchor items dropped by language*

Language	Total anchor items	Anchor items dropped	%
Amharic	55	0	0
Tigrinya	55	1	2
Oromifa	55	1	2
Telugu	57	0	0
Peru	125	0	0
Vietnam	76	1	1

Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Once items with poor fit and DIF for all groups were dropped, we ran the three-parameter model again to get corrected IRT scores for each child. Table 3 shows the average mean scores for all the languages by cohort and round. IRT scores, for both cohorts, have increased over time.

Table 3. *Mean scores by language for each round and age cohort (standard deviation)*

	Older Cohort			Younger Cohort		
	Round 2	Round 3	Round 2	Round 3	Round 4	Round 5
Amharic	2.3	2.6	0.0	1.5	2.3	2.9
	(0.95)	(1.17)	(1.00)	(1.28)	(1.36)	(1.40)
Tigrigna	2.7	2.9	0.0	1.5	2.5	3.0
	(1.04)	(1.25)	(1.00)	(1.10)	(1.24)	(1.25)
Oromifa	2.1	2.3	0.0	0.9	2.7	3.0
	(1.00)	(1.00)	(1.00)	(1.04)	(1.18)	(1.22)
Telugu	2.2	2.6	0.0	0.8	1.9	2.6
	(1.00)	(1.06)	(1.00)	(0.97)	(1.05)	(1.23)
Spanish	2.8	3.2	0.0	1.5	2.8	3.3
	(0.79)	(0.66)	(1.00)	(0.76)	(0.74)	(0.65)
Vietnamese	3.1	3.0	0.0	1.8	3.1	3.4
	(1.20)	(1.12)	(1.00)	(0.89)	(1.16)	(1.34)

Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Table 4 shows the increment in the IRT scores over time by cohort and language. We found that almost all the increments are statistically significant by age cohort and language; also, with the Younger Cohort, since we have four time points, we could estimate the increments between rounds. Our results show that Amharic, Tigrigna and Vietnamese children have the highest increment between Rounds 2 and 3, Oromifa, Vietnamese and Spanish children have the highest increment between Rounds 3 and 4, and Amharic and Telugu children have the highest increment between Rounds 4 and 5.

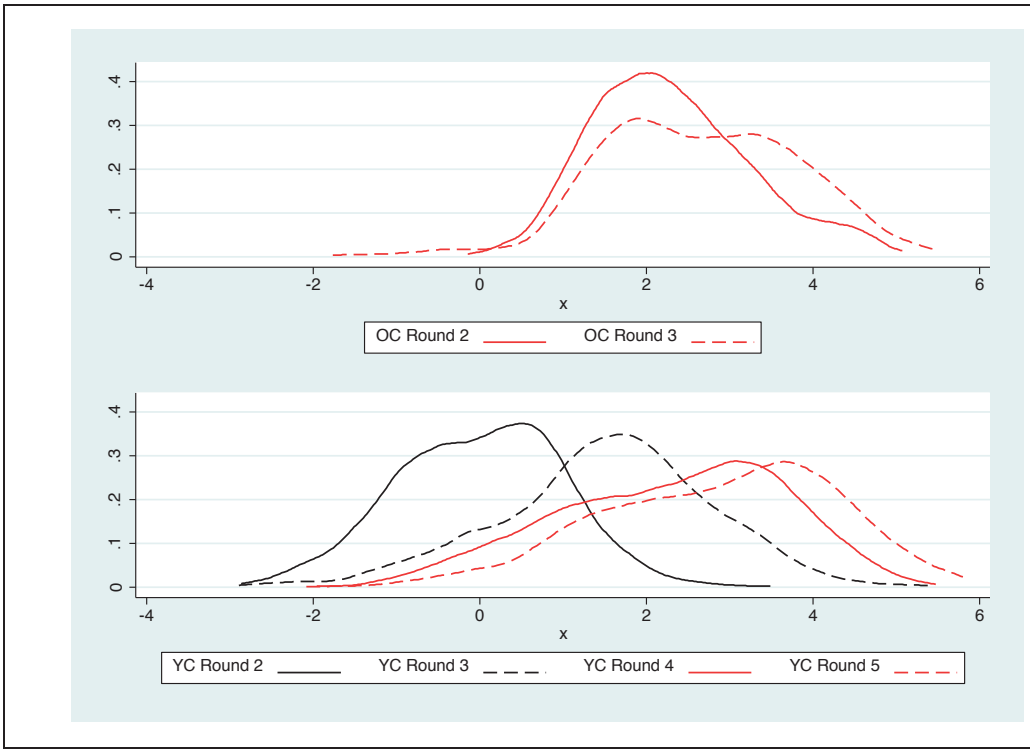
Table 4. *Gap analysis for each age cohort (standard error)*

	Older Cohort	Younger Cohort		
	R3-R2	R3-R2	R4-R3	R5-R4
Amharic	0.34	1.51	0.78	0.58
	(0.07)	(0.06)	(0.06)	(0.07)
Tigrigna	0.16	1.51	1.01	0.48
	(0.12)	(0.08)	(0.09)	(0.09)
Oromifa	0.24	0.89	1.79	0.30
	(0.11)	(0.08)	(0.08)	(0.09)
Telugu	0.45	0.81	1.12	0.65
	(0.05)	(0.03)	(0.03)	(0.04)
Spanish	0.41	0.78	1.28	0.44
	(0.04)	(0.02)	(0.03)	(0.02)
Vietnamese	0.11	1.76	1.38	0.24
	(0.05)	(0.03)	(0.03)	(0.04)

Source: Young Lives main survey, Rounds 2, 3, 4 and 5.
Note: Mean scores differences between rounds in bold indicate that is statistically significant at 5 per cent, according the ttest for dependent or correlated samples.

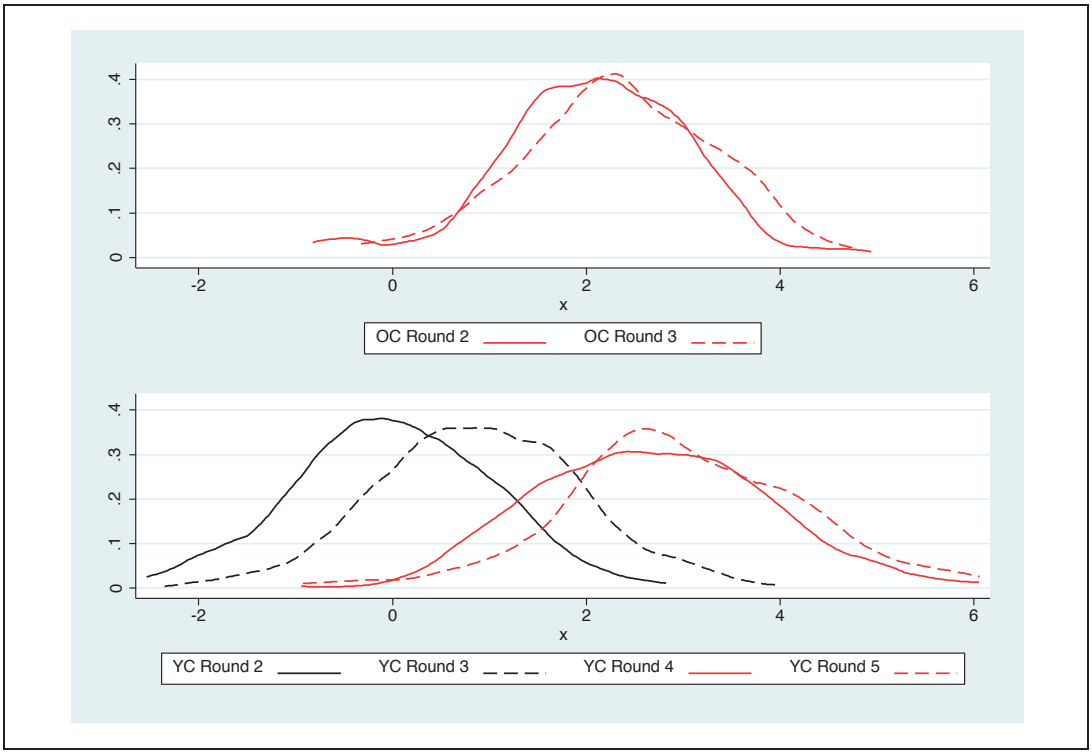
Figures 4 to 9 show that average scores for Younger Cohort children increased over time for all the main languages analysed, while the scores for the Older Cohort show some stagnation for Telugu, Spanish and Vietnamese children: therefore, the average scores are fairly similar across rounds, confirming possible ceiling effects.

Figure 4. *Distribution of IRT scores for Amharic by round and age cohort*



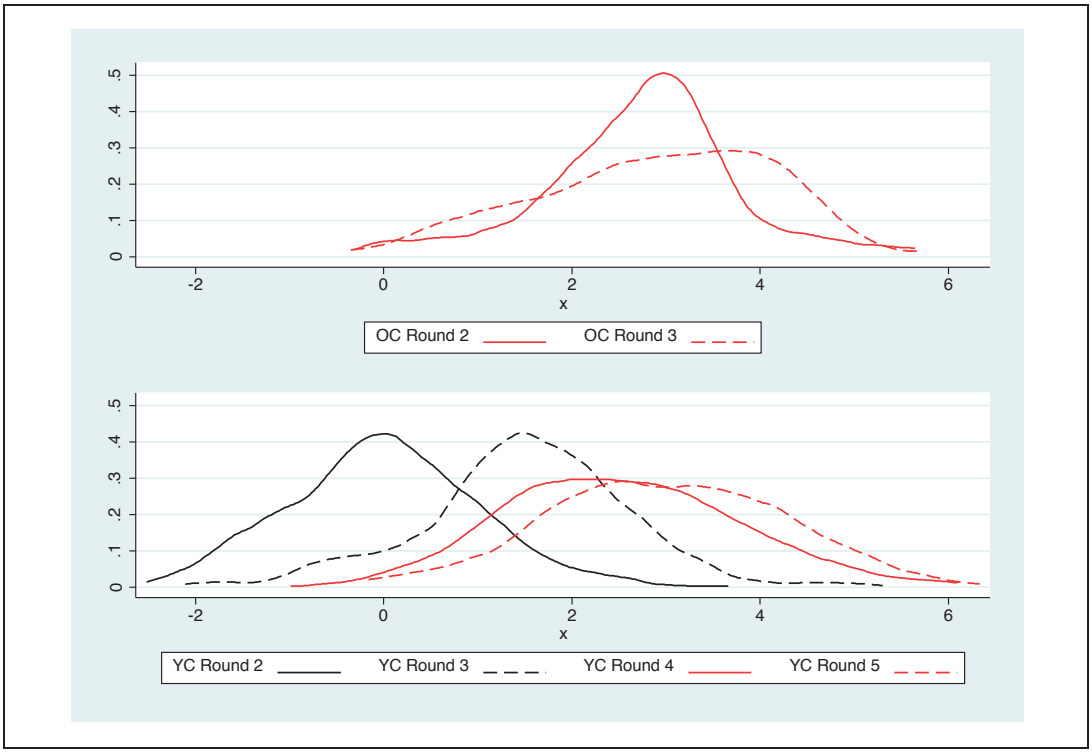
Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 5. *Distribution of IRT scores for Oromifa by round and age cohort*



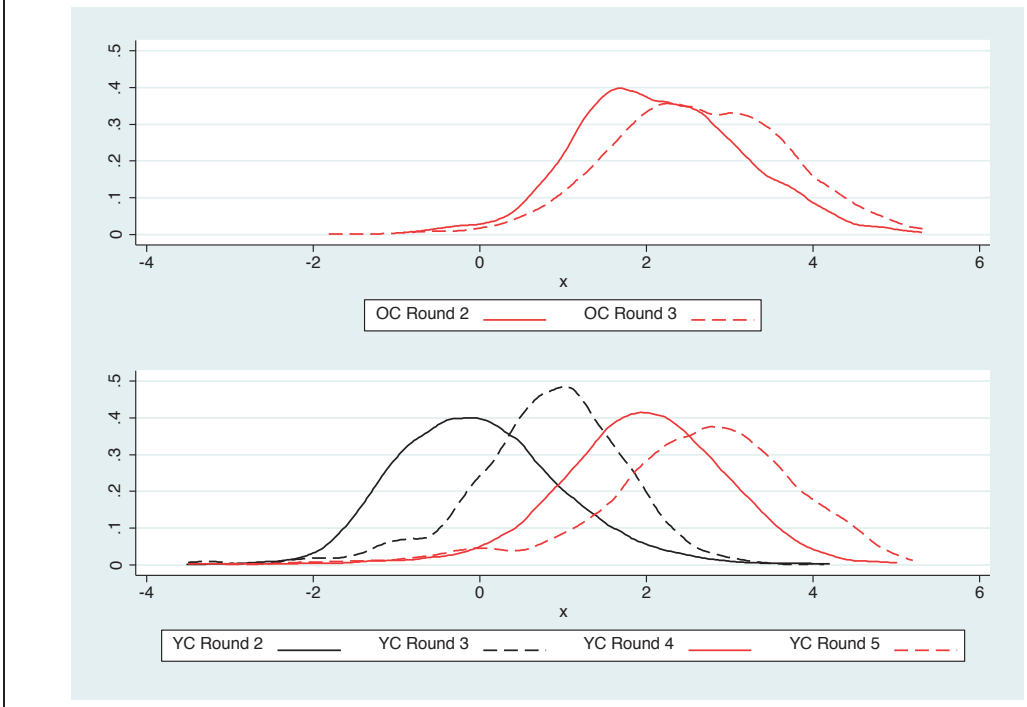
Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 6. *Distribution of IRT scores for Tigrinya by round and age cohort*



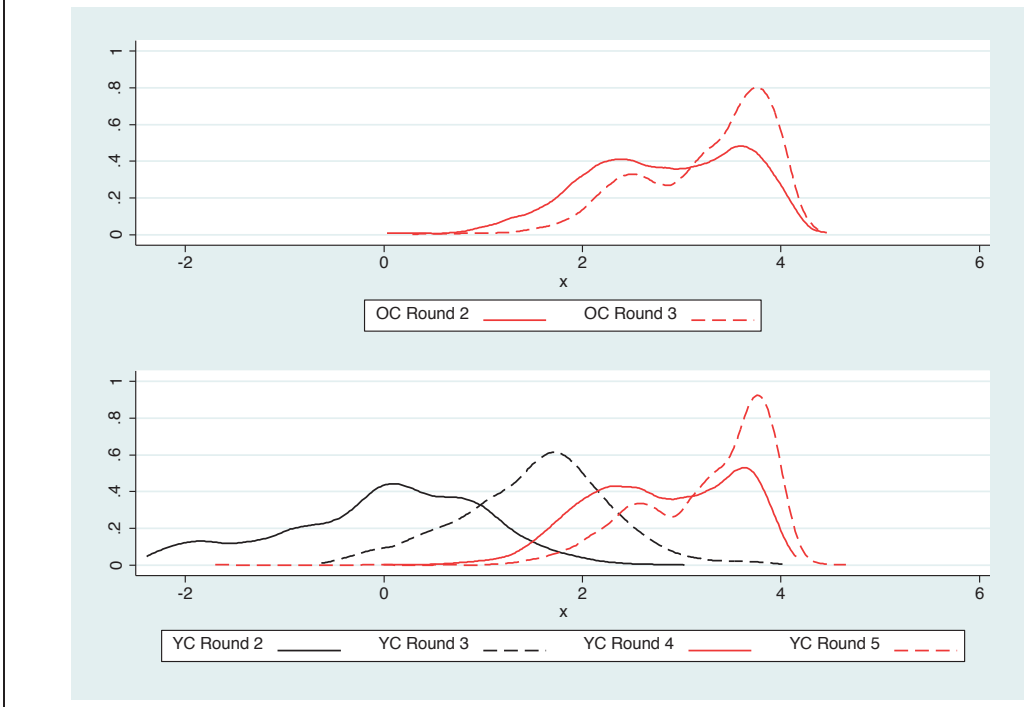
Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 7. *Distribution of IRT scores for Telugu by round and age cohort*



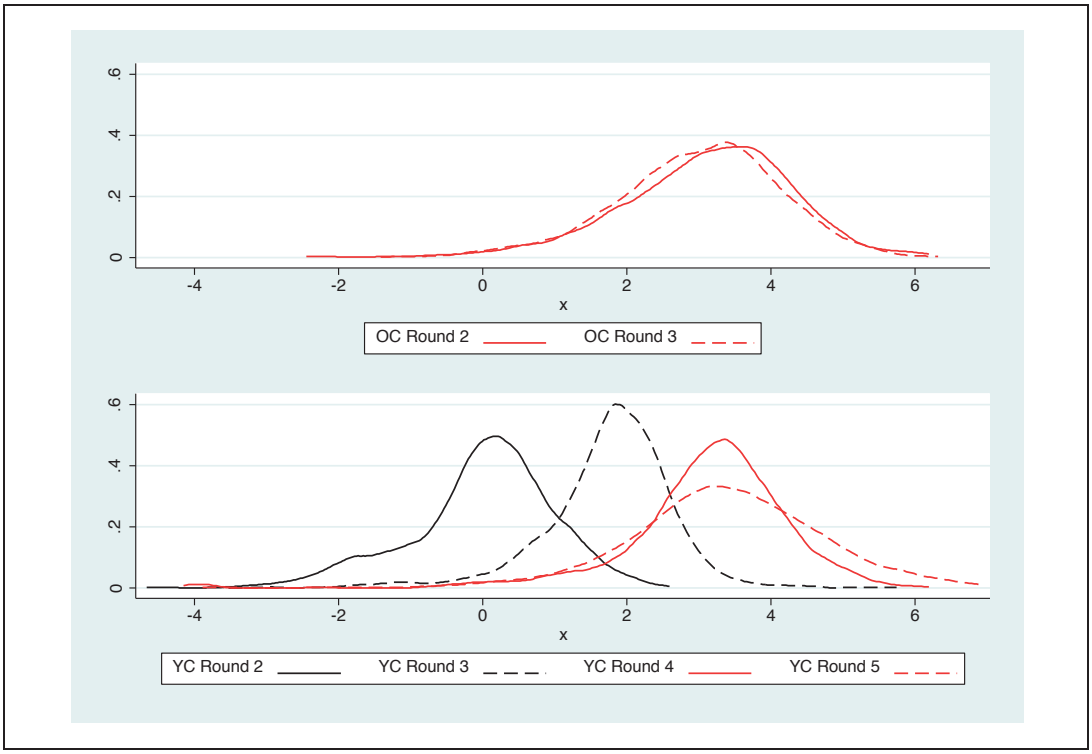
Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 8. *Distribution of IRT scores for Spanish by round and age cohort*



Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 9. *Distribution of IRT scores for Vietnamese by round and age cohort*



Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Finally, we equated the PPVT (Ethiopia and Vietnam) and TVIP scores (Peru) of the siblings with the IRT scores of the index children in order to have comparable measures. However, for Ethiopia, it was not possible to fit an IRT scale for Tigrigna and Oromifa children since most of the items have poor ICC fit (no correlation between the proportion of children who correctly answer an item and the average child´s ability), as a result of the reduced number of observations in the sibling data compared with the index children data. Appendix A provides details of the item parameters for each of the PPVT scales.

4.2. Maths achievement

We estimated a two-parameter IRT analysis for the pool sample for each country (Ethiopia, India, Peru and Vietnam). One main difference between maths and vocabulary is that at early grades most of the items involve solving a simple algorithm, such as addition, subtraction, multiplication or division. Therefore, we decided not to perform an analysis by main language, and instead analysed the whole sample.

The first analysis performed allowed us to identify those items with poor fit that should be dropped from each of the composite scores. Table 5 shows the percentage of items that were dropped because of poor item fit or DIF for all the comparison groups (round and cohort). The percentage of items dropped, on average, was less than 10 per cent of the total items; Ethiopia had the highest percentage (13 per cent) of items dropped and India (3 per cent) had the lowest percentage.

Table 5. *Items dropped by country*

	Total items1	Items dropped	%
Ethiopia	88	11	13
India	92	3	3
Peru	91	7	8
Vietnam	107	5	5

Note: [1] Number of items administered from Round 2 to Round 5.

Once items with poor fit and DIF for all groups were dropped, we ran the two-parameter IRT model again in order to get corrected scores for each child. Table 6 shows the average mean scores for all the countries by cohort and round. We can see a clear increase over time in maths scores for Younger Cohort children in all countries. The results from the Older Cohort in Ethiopia, India and Vietnam show an increase over time between Round 2 and Round 4, but a decrease between Rounds 2 and 3. Only Peru shows an increment in maths achievement scores over time for both age cohorts.

Table 6. *Mean scores by country for each round and age cohort (standard deviation)*

	Older Cohort			Younger Cohort		
	Round 2	Round 3	Round 4	Round 3	Round 4	Round 5
Ethiopia	0.0	-0.1	0.5	-1.5	-0.3	0.1
	(1.00)	(1.06)	(1.04)	(0.88)	(0.92)	(0.98)
India	0.0	-0.4	0.0	-1.7	-0.3	0.0
	(1.00)	(1.02)	(1.21)	(0.93)	(0.96)	(1.03)
Peru	0.0	0.2	0.5	-1.7	0.0	0.3
	(1.00)	(1.02)	(1.15)	(0.96)	(0.84)	(1.04)
Vietnam	0.0	-0.1	0.4	-0.2	0.0	0.2
	(1.00)	(1.11)	(1.09)	(1.19)	(0.97)	(1.06)

Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Table 7 shows the increment in the IRT scores over time by age cohort and country. For both age cohorts, the achievement gaps between rounds were statistically different from zero. However, for Ethiopia, India and Vietnam, we see a decline in the maths achievement between Rounds 2 and 3, while between Rounds 3 and 4, there is an increase in maths achievement. Younger Cohort children in all countries show an increment over time in their maths abilities, compared to their older peers.

Table 7. *Gap analysis by country and age cohort (standard error)*

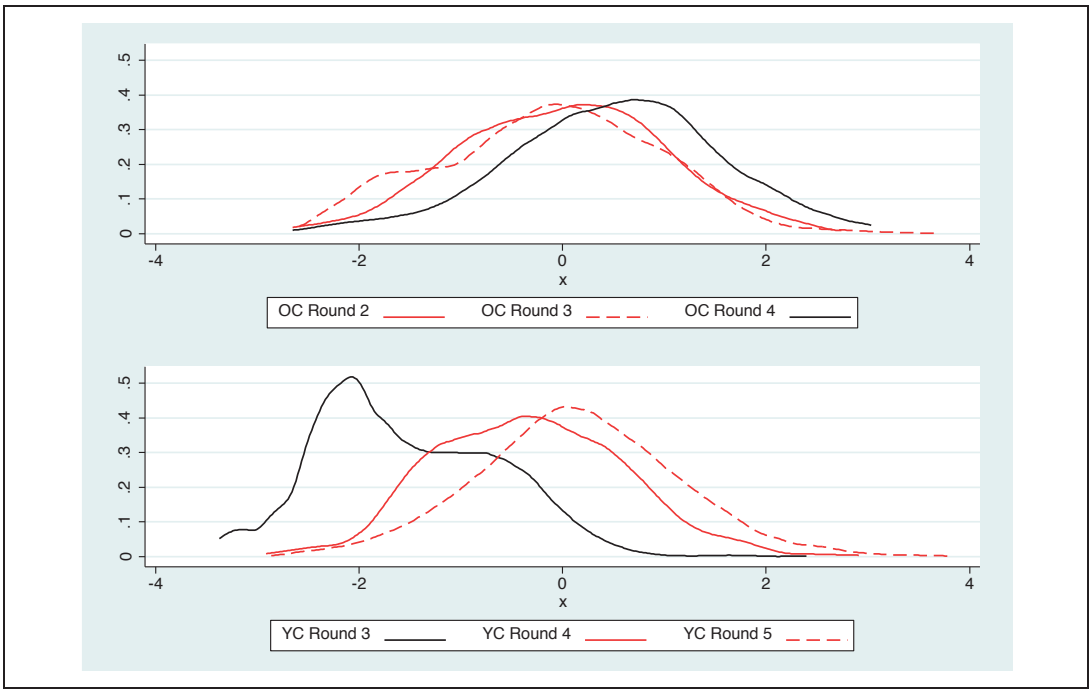
	Older Cohort		Younger Cohort	
	R3-R2	R4-R3	R4-R3	R5-R4
Ethiopia	-0.13	0.63	1.19	0.47
	(0.05)	(0.05)	(0.03)	(0.03)
India	-0.36	0.33	1.38	0.28
	(0.05)	(0.05)	(0.03)	(0.03)
Peru	0.16	0.33	1.77	0.27
	(0.05)	(0.06)	(0.03)	(0.03)
Vietnam	-0.11	0.53	0.14	0.27
	(0.05)	(0.05)	(0.03)	(0.03)

Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Note: Mean scores differences between rounds in bold indicate that is statistically significant at 5 per cent, according the ttest for dependent or correlated samples.

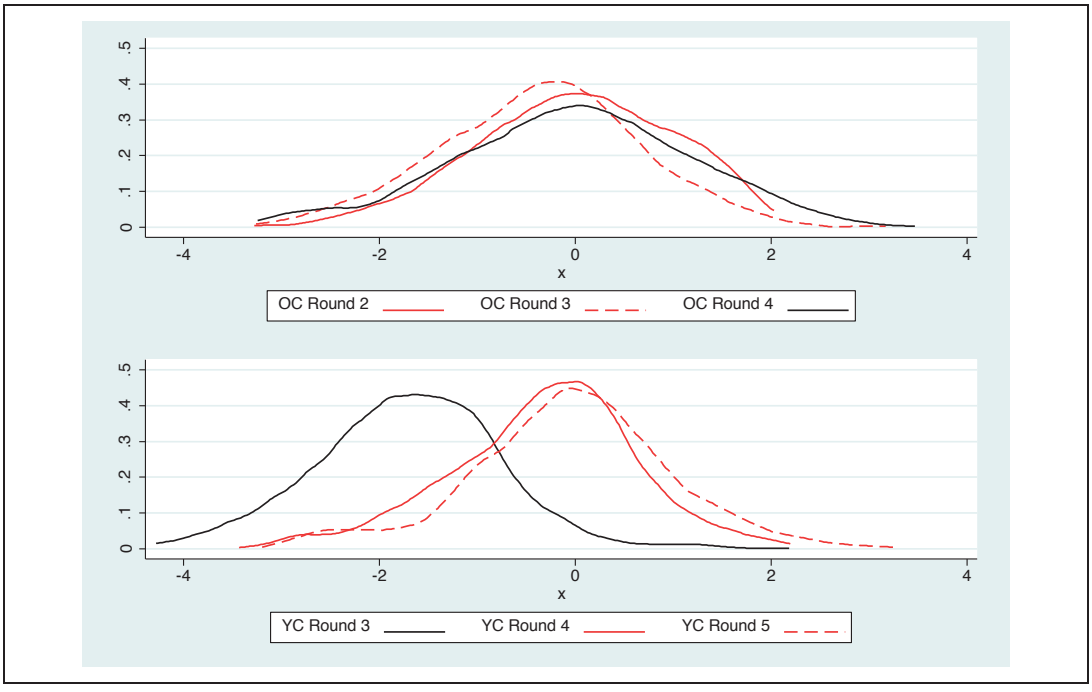
Finally, Figures 10 to 13 show that average scores for Younger Cohort children increase over time for all the countries, while the scores for the Older Cohort show some stagnation for India, Ethiopia and Vietnam between Rounds 2 and 3.

Figure 10. *Distribution of IRT scores for Ethiopia by age cohorts*



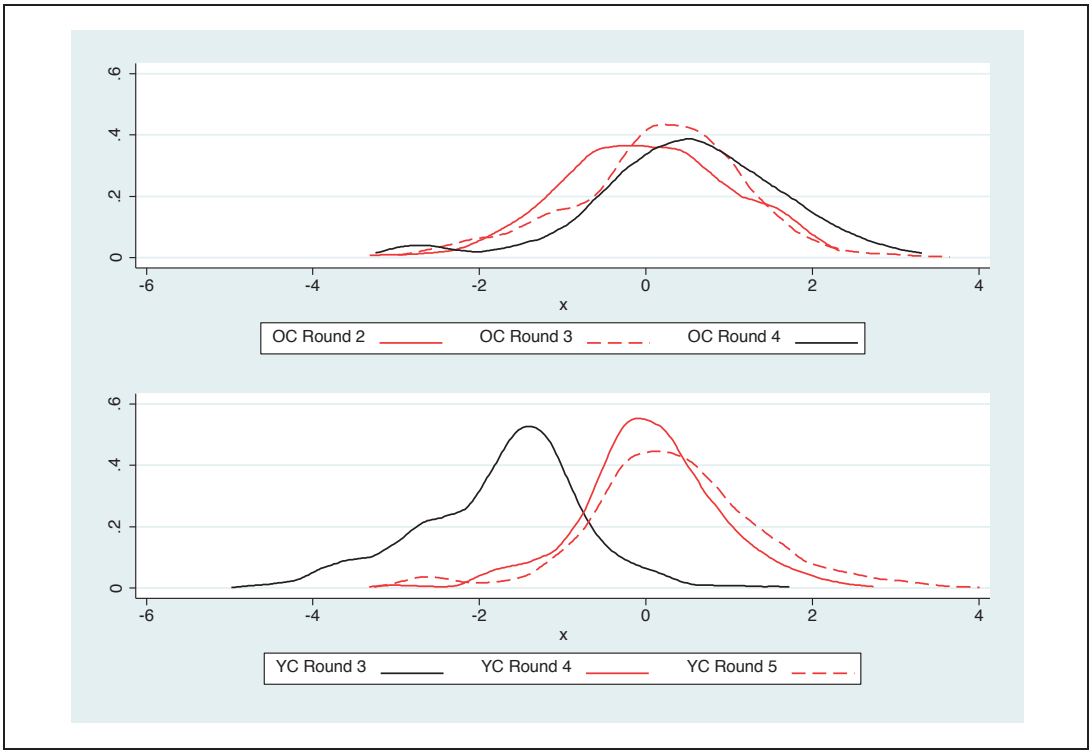
Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 11. *Distribution of IRT scores for India by age cohorts*



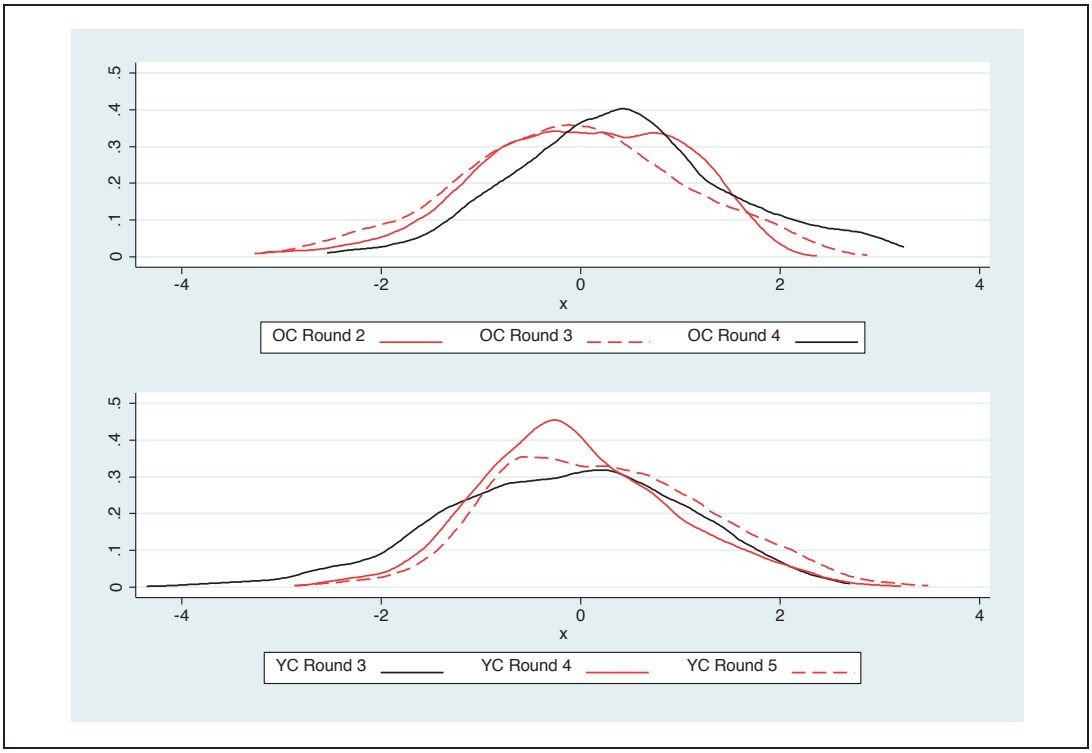
Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 12. *Distribution of IRT scores for Peru by age cohorts*



Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

Figure 13. *Distribution of IRT scores for Vietnam by age cohorts*



Source: Young Lives main survey, Rounds 2, 3, 4 and 5.

4.3. Reading comprehension

We estimated a three-parameter IRT analysis for the pool sample for each of the main languages in the four countries: Ethiopia (Amharic, Tigrigna and Oromifa), India (Telugu), Peru (Spanish) and Vietnam (Vietnamese). Like in the previous tests, the first analysis performed allowed us to identify those items with poor fit that should be dropped from each of the composite scores.

Table 8 shows the percentage of items that were dropped because of poor item fit or DIF for all the comparison groups (round and cohort). The percentage of items dropped, on average, was less than one-fifth of the total items; the Oromifa language had the highest percentage (17 per cent) of items dropped, and Vietnam was the country with the lowest percentage of items dropped (3 per cent).

Table 8. *Number of items dropped by language*

Language	Total items	Items dropped	%
Amharic	36	2	6
Tigrinya	36	4	11
Oromifa	36	6	17
Telugu	36	5	14
Spanish	36	3	8
Vietnam	39	1	3

Source: Young Lives main survey, Rounds 4 and 5.

Once items with poor fit and DIF for all groups were dropped, we ran the three-parameter model again in order to get corrected IRT scores for each child. Table 9 shows the average mean scores for all the languages by cohort and round. Our results show that IRT scores increased over time for both age cohorts.

Table 9. *Mean scores by language for each round and age cohort (standard deviation)*

	Older Cohort		Younger Cohort	
	Round 4	Round 5	Round 4	Round 5
Amharic	0.5	-	0.0	0.4
	(0.98)		(1.00)	(1.03)
Tigrigna	0.9	-	0.0	0.4
	(1.05)		(1.00)	(0.95)
Oromifa	0.7	-	0.0	0.5
	(1.10)		(1.00)	(1.09)
Telugu	0.6	-	0.0	0.4
	(1.07)		(1.00)	(0.98)
Spanish	0.7	-	0.0	0.5
	(1.34)		(1.00)	(1.09)
Vietnamese	0.6	-	0.0	0.4
	(1.20)		(1.00)	(1.16)

Source: Young Lives main survey, Rounds 4 and 5.

Table 10 shows the increment in the IRT scores over time for the Younger Cohort by language. We found that all the increments were statistically significant. The languages with the highest increment in reading comprehension scores were Spanish and Oromifa, while Amharic had the lowest increment.

Table 10. *Gap analysis for the Younger Cohort (standard error)*

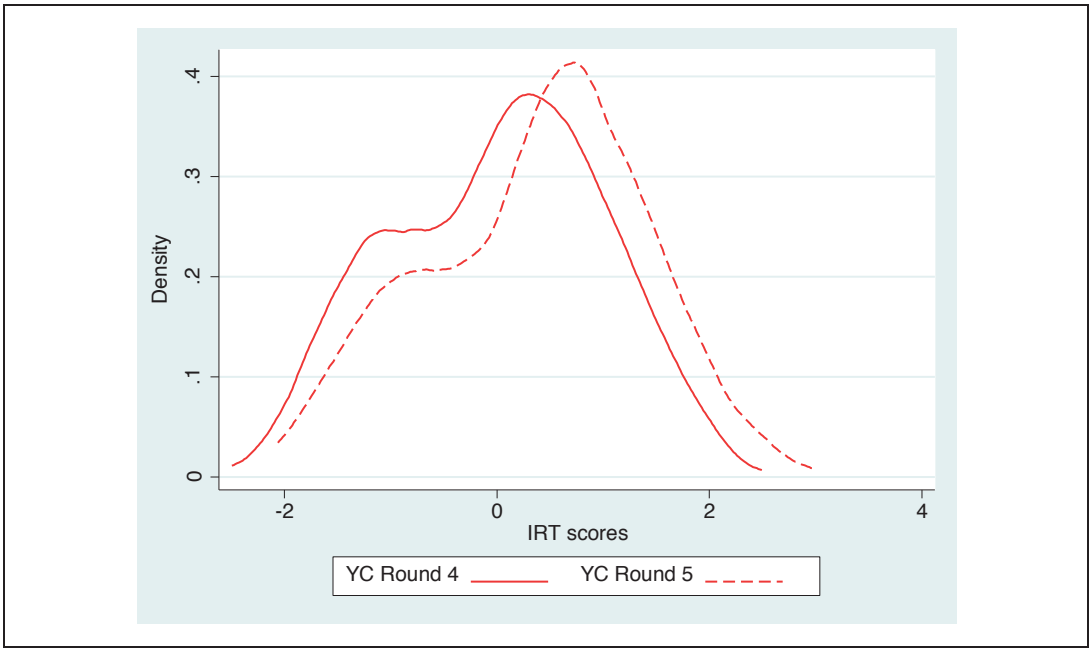
	R5 - R4
Amharic	0.36
	(0.05)
Tigrigna	0.44
	(0.07)
Oromifa	0.53
	(0.08)
Telugu	0.37
	(0.04)
Spanish	0.51
	(0.03)
Vietnamese	0.37
	(0.04)

Source: Young Lives main survey, Rounds 4 and 5.

Note: Mean scores differences between rounds in bold indicate that is statistically significant at 5 per cent, according the ttest for dependent or correlated samples.

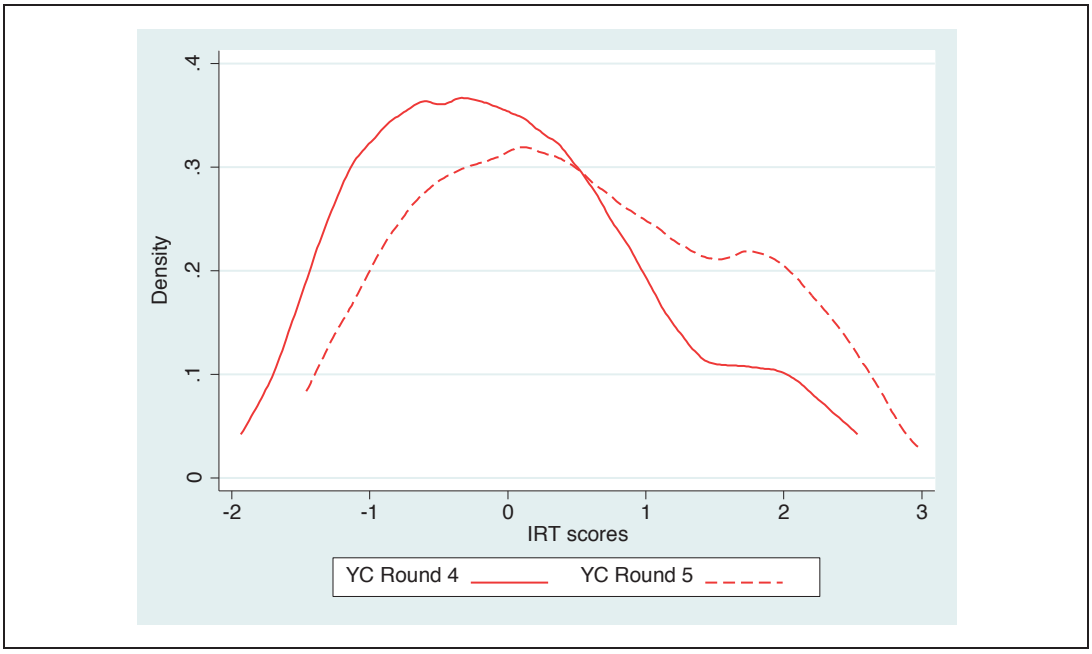
Figures 14 to 19 show that the average scores for the Younger Cohort increased over time for all the main languages analysed.

Figure 14. *Distribution of IRT scores for Amharic by round, Younger Cohort*



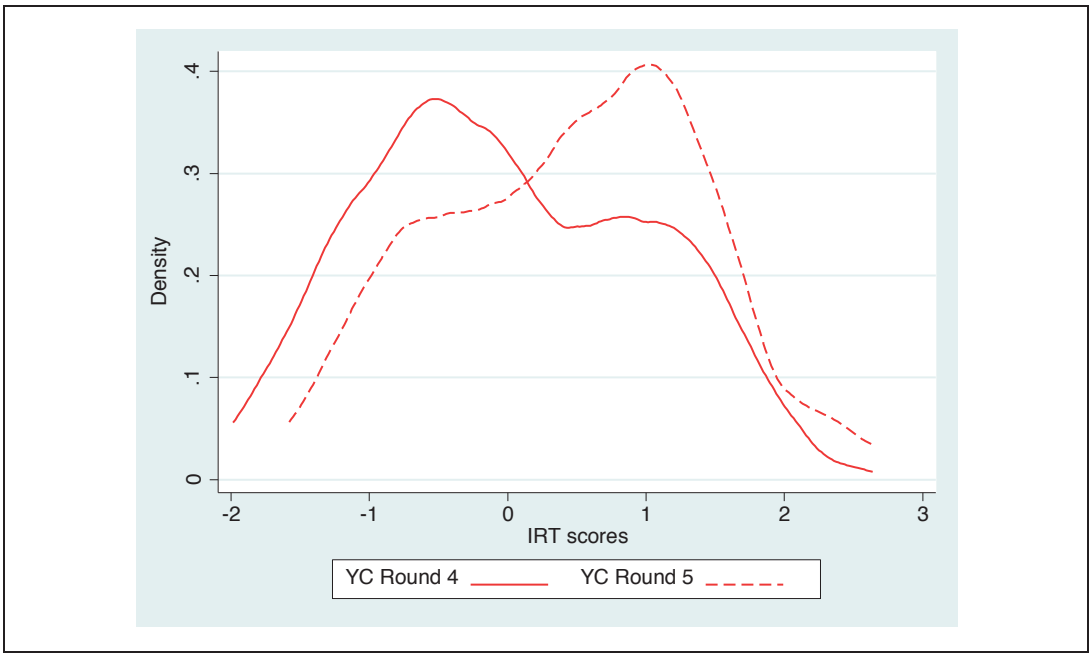
Source: Young Lives main survey, Rounds 4 and 5.

Figure 15. *Distribution of IRT scores for Oromifa by round, Younger Cohort*



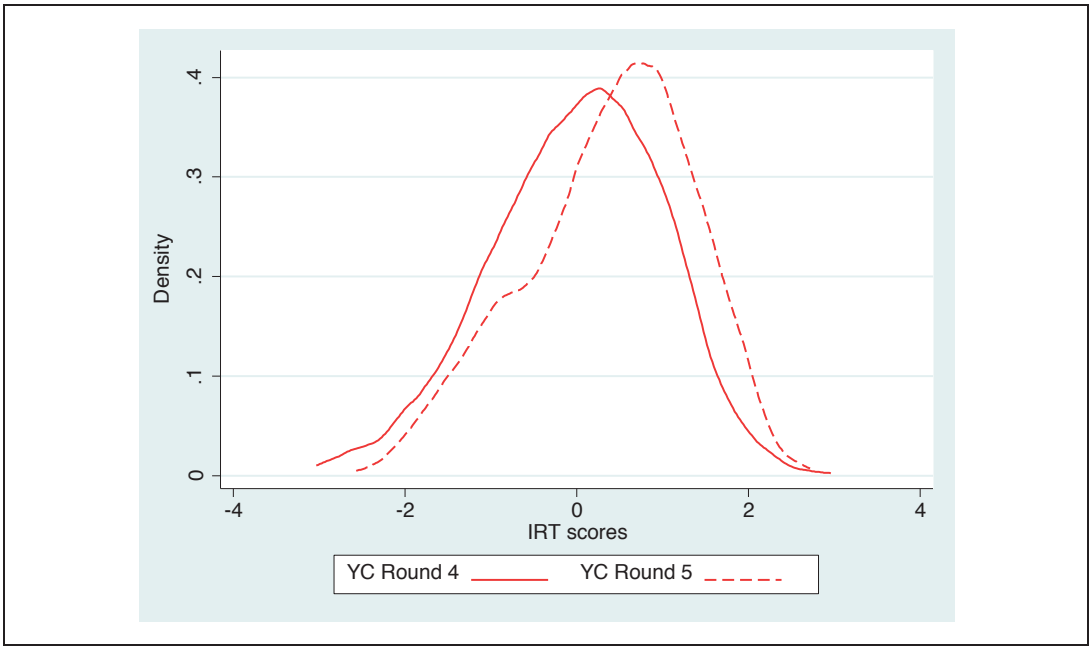
Source: Young Lives main survey, Rounds 4 and 5.

Figure 16. *Distribution of IRT scores for Tigrinya by round, Younger Cohort*



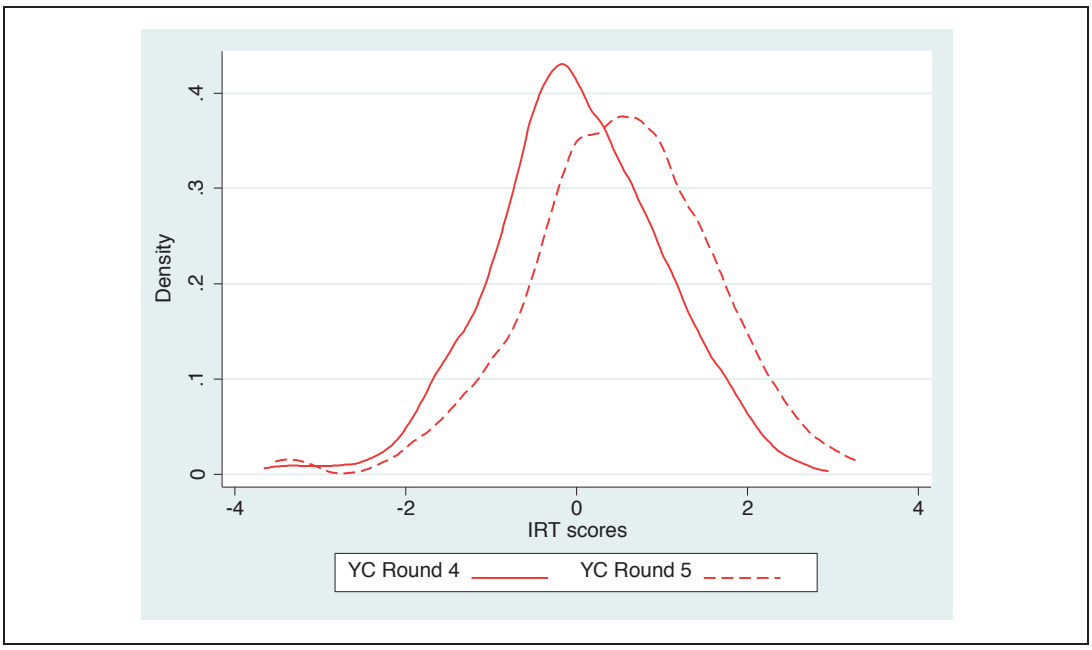
Source: Young Lives main survey, Rounds 4 and 5.

Figure 17. *Distribution of IRT scores for Telugu by round, Younger Cohort*



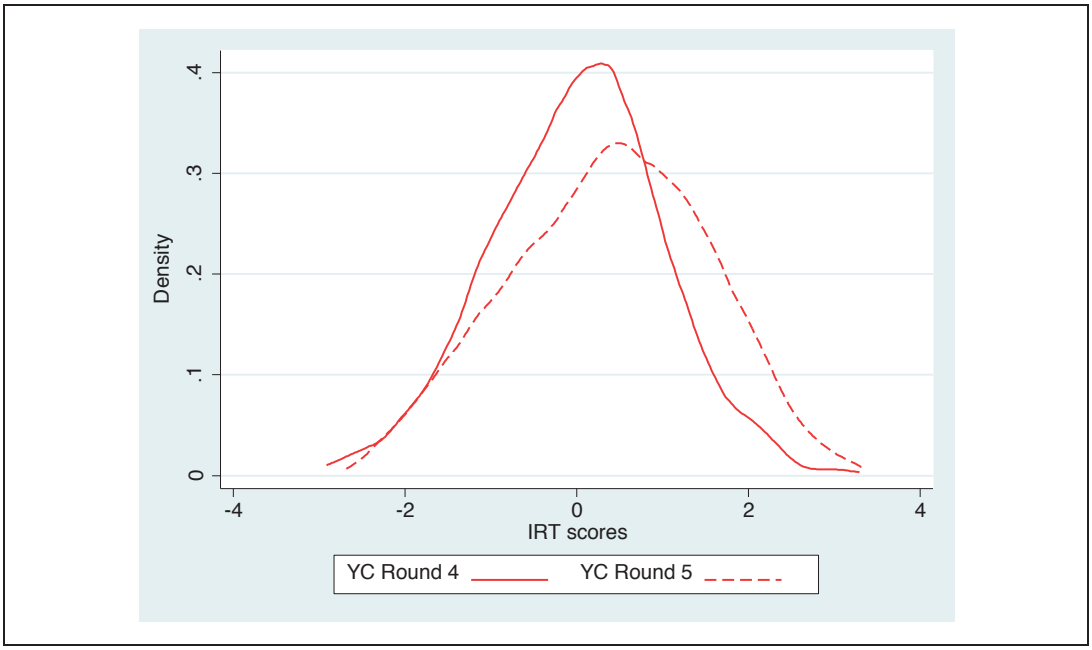
Source: Young Lives main survey, Rounds 4 and 5.

Figure 18. *Distribution of IRT scores for Spanish by round, Younger Cohort*



Source: Young Lives main survey, Rounds 4 and 5.

Figure 19. *Distribution of IRT scores for Vietnamese by round, Younger Cohort*



Source: Young Lives main survey, Rounds 4 and 5.

Finally, for Ethiopia, India, Peru and Vietnam, we equated the PPVT (or TVIP) scores of the siblings with the scores of the Young Lives children in order to have comparable measures. In addition, for India, we equated the maths scores with the scores of the Young Lives children.

5. Final remarks

This technical note provides details of the procedures followed to equate the PPVT/TVIP, maths and reading comprehension scores for the main languages in the four Young Lives study countries (Ethiopia, India, Peru and Vietnam). The key findings were that:

- The new approach followed since Round 4 was suitable since most of the items selected had a good item fit and did not show DIF by groups (cohort or round) for all the main languages in each country, or by country.
- For some languages, such as Tigrigna and Oromifa, the number of items dropped due to poor fit or DIF across all groups (rounds and cohort) was significant, with almost one third of items dropped from the final scale. However, for Spanish, since the TVIP is in the same language, no item was dropped from the analysis.
- It was possible to ensure adequate equating of the PPVT/TVIP scores for all the main languages across rounds and cohorts. Our results show that PPVT/TVIP scores for the Younger and Older Cohort increased over time for all the main languages, and these increments are statistically significant. There were similar results for maths achievement scores. In reading comprehension, the Younger Cohort improved their scores between Rounds 4 and 5.

- Results from the Younger Cohort show a curvilinear (concave) trend in the vocabulary acquisition of all children as the highest increment was between Rounds 2 and 3, and later increments were progressively smaller.
- Even though it was possible to equate the maths scores for each country, it is important to take into consideration that test administration was different across rounds, which could have introduced measurement error, and this could explain the results obtained for the Older Cohort. Therefore, these scores have to be taken as referential.
- The reading comprehension scores demonstrate that Younger Cohort children in all languages and countries show a significant increment over time, with Spanish (0.51 SD) and Oromifa (0.53 SD) children having the highest increment, and Amharic children the lowest (0.36 SD).
- Comparing the level of vocabulary between the Younger Cohort and Older Cohort at the same age, no significant increments over time could be seen for most languages. Oromifa was the only language that had a significant change over time, with the Younger Cohort having higher scores at age 15.

6. References

Bateman, I., S. Dent, E. Peters, P. Slovic, and C. Starmer (2007) 'The Affect Heuristic and the Attractiveness of Simple Gambles', *Journal of Behavioral Decision Making* 20.4: 365–380.

Campbell, J.M. (1998) 'Review of the Peabody Picture Vocabulary Test – Third Edition', *Journal of Psychoeducational Assessment* 16.4: 334–338.

Campbell, J.M., S.K. Bell and L.K. Keith (2001) 'Concurrent Validity of the Peabody Picture Vocabulary Test – Third Edition as an Intelligence and Achievement Screener for Low SES African American Children', *Assessment* 8.1: 85–94.

Cueto, S., and J. Leon (2012) *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 3 of Young Lives*, Technical Note 25, Oxford: Young Lives.

Cueto, S., J. Leon, G. Guerrero, and I. Munoz (2009) *Psychometric Characteristics of Cognitive Development and Achievement Instruments in Round 2 of Young Lives*, Technical Note 15, Oxford: Young Lives.

Dorans, N.J., and P.W. Holland (1993) 'DIF Detection and Description: Mantel-Haenzel and Standardization', in P.W. Holland and H. Wainer (eds) *Differential Item Functioning*, 35-66, Hillsdale, NJ: Lawrence Erlbaum Associates.

Dunn, L.M., E.R. Padilla, D.E. Lugo, and L.M. Dunn. 1986. 'Test de Vocabulario en Imagenes Peabody' [Peabody Picture Vocabulary Test], Circle Pines, MN: American Guidance Service.

Dunn, L. and L. Dunn (1997) *Examiner's Manual for the PPVT-III. Form IIIA and IIIB*, Minnesota: AGS.

Gray, S., E. Plante, R. Vance, and M. Henrichsen (1999) 'The Diagnostic Accuracy of Four Vocabulary Tests Administered to Preschool-Age Children', *Language, Speech and Hearing Services in Schools* 30.2: 196–206.

Hambleton, R.K. (1989) 'Principles and Selected Applications of Item Response Theory', In R.L. Linn (ed.) *Educational Measurement*, 147-200, New York: Macmillan.

Hambleton, R.K., and H. Swaminathan (1985) *Item Response Theory: Principle and Applications*, Boston, MA: Kluwer Nijhoff.

Linacre, J.M. (2008) 'Winsteps: A Rasch Analysis Computer Program (Version 3.68)', http://www.winsteps.com (accessed 29 May 2020).

McMahon, W. (1997) 'Recent Advances in Measuring the Social and Individual Benefits of Education', *International Journal of Educational Research* 27.6: 449-481.

Wolfe, B., and S. Zuvekas (1997) 'Non-Market Effects of Education', *International Journal of Education Research* 27.6: 494-502.

Wright, B.D., and G.A. Douglas (1976) 'Rasch Item Analysis by Hand', MESA Research Memorandum Number 21, Chicago: University of Chicago.

Appendices

Appendix A.
Item parameters for each cognitive and achievement test

Table 1. *Item parameters for the Amharic PPVT analysis*

Item	Discrimination	Difficulty	Guessing
1	1.05	-1.70	0.78
2	1.27	-2.48	0.51
3	1.22	-2.86	0.41
4	1.49	-2.32	0.17
5	0.78	-1.44	0.14
6	1.18	-2.43	0.24
7	3.11	0.21	0.37
8	2.05	-1.24	0.20
9	1.15	-1.39	0.23
10	0.54	-0.92	0.12
11	1.02	-2.02	0.10
12	1.18	-0.76	0.54
13	1.57	-1.68	0.56
14	0.78	-1.02	0.57
15	1.35	-1.67	0.13
16	1.24	-0.52	0.74
17	1.72	-1.21	0.18
18	0.70	0.12	0.23
19	1.20	-0.86	0.57
20	0.73	-0.84	0.05
21	0.67	-1.68	0.19
22	0.59	-1.30	0.12
23	0.89	-0.36	0.18
24	0.76	-1.02	0.21
25	1.49	-1.51	0.17
26	1.60	-1.10	0.17
27	1.27	-0.54	0.16
28	1.22	-1.17	0.28
29	0.57	-1.57	0.08
30	0.64	-1.08	0.05
31	1.19	-0.94	0.11
32	2.02	-1.03	0.34
33	1.33	-0.87	0.29
34	0.74	-0.17	0.26
35	0.77	-0.72	0.18
36	0.86	-0.57	0.31
37	0.71	-0.28	0.04
38	1.95	-1.18	0.47
39	0.57	-1.15	0.08
40	0.90	0.78	0.12
41	0.60	-0.47	0.13
42	0.47	0.26	0.07
43	0.61	0.27	0.10
44	0.44	0.81	0.03
45	0.78	-0.65	0.17
46	1.82	-0.57	0.11
47	1.22	-1.38	0.07
48	0.72	-1.77	0.14
49	0.96	-0.27	0.11
50	1.52	-0.60	0.42
51	0.89	-0.16	0.26
52	1.36	0.38	0.27
53	0.86	1.04	0.24
54	0.88	0.42	0.35
55	1.34	0.20	0.25
56	1.53	0.05	0.15
57	1.23	-0.06	0.33
58	0.70	0.53	0.05
59	1.31	-0.90	0.62
60	1.48	-0.60	0.06
61	1.29	-1.50	0.30
62	1.22	-1.45	0.13
63	1.28	-0.96	0.49
64	1.31	-1.02	0.12
65	0.92	1.23	0.41
66	0.56	0.45	0.06
67	0.63	-0.32	0.13
68	0.89	-0.71	0.26
69	0.86	-0.69	0.12
70	0.87	-0.55	0.08
71	0.53	-0.27	0.05
72	1.07	-0.62	0.15
73	0.57	-1.43	0.42
74	1.60	-0.58	0.20
75	1.69	-0.04	0.42
76	1.22	0.04	0.07
77	1.55	-0.42	0.44
78	0.95	-0.51	0.14
79	0.90	-1.37	0.40
80	1.50	0.18	0.21
82	1.03	-0.34	0.12
83	0.81	0.10	0.05
84	0.65	0.19	0.22
85	0.71	-1.40	0.16
86	0.69	2.27	0.09
87	1.08	0.43	0.39
88	0.98	1.00	0.12
90	1.42	0.00	0.27
91	1.02	-0.01	0.23
92	0.97	0.89	0.12

Item	Discrimination	Difficulty	Guessing
93	1.19	-0.34	0.49
94	0.98	0.18	0.37
95	1.00	-0.09	0.14
96	0.47	1.04	0.26
97	0.75	0.83	0.07
98	1.20	0.69	0.14
99	1.08	0.33	0.29
101	0.82	-0.08	0.03
103	1.31	-0.17	0.42
104	1.42	-0.32	0.48
105	0.43	-0.58	0.53
106	0.71	1.33	0.20
107	0.62	1.34	0.53
108	0.68	-0.15	0.25
110	1.18	0.46	0.06
111	0.78	0.39	0.18
112	1.49	0.11	0.32
114	0.68	-0.70	0.34
115	0.53	-1.71	0.14
116	0.77	-0.94	0.14
117	0.86	0.77	0.12
118	1.33	-0.69	0.28
120	0.92	0.39	0.12
121	0.55	-0.84	0.32
122	1.46	0.17	0.05
123	1.31	0.62	0.17
124	1.33	0.80	0.12
125	0.53	1.36	0.16
126	0.70	1.72	0.30
127	1.18	-0.03	0.64
128	0.93	-0.31	0.20
129	0.97	0.12	0.04
130	2.21	0.01	0.44
131	1.25	0.95	0.14
132	1.36	-0.10	0.59
133	0.68	1.03	0.29
134	0.83	-0.80	0.33
135	0.89	0.55	0.11
136	0.73	-0.62	0.05
137	0.93	0.19	0.08
138	1.00	0.75	0.17
140	0.80	0.97	0.04
141	1.37	0.16	0.35
142	1.10	0.60	0.03
143	1.47	-0.27	0.04
144	0.66	0.35	0.41

Item	Discrimination	Difficulty	Guessing
145	0.96	-1.41	0.28
146	1.06	0.83	0.05
148	0.73	1.35	0.06
149	0.91	0.80	0.53
151	0.86	-0.38	0.06
154	0.65	1.68	0.18
155	1.05	-0.61	0.40
156	1.24	-0.06	0.13
157	0.50	-2.04	0.33
160	0.34	-0.48	0.43
161	0.56	-0.51	0.33
162	0.70	0.02	0.13
163	1.60	0.60	0.16
164	0.84	0.21	0.19
167	0.38	-1.45	0.17
168	0.76	1.87	0.63
169	0.68	-1.25	0.28
170	1.11	-0.20	0.41
171	0.48	0.52	0.30
174	0.74	-0.27	0.44
176	0.96	0.96	0.21
177	0.60	1.91	0.22
178	0.44	-0.82	0.36
179	2.20	0.14	0.24
180	0.65	2.18	0.07
181	0.92	0.92	0.51
182	0.66	0.80	0.22
185	1.01	2.48	0.16
186	0.66	-0.84	0.32
189	0.44	0.82	0.18
190	0.59	-1.32	0.28
191	1.21	0.30	0.38
192	0.85	0.67	0.33
195	1.64	0.02	0.38
196	0.34	-0.74	0.15
197	1.97	0.39	0.20
199	0.53	0.73	0.57
203	0.79	0.95	0.17
204	1.54	0.82	0.28
205	0.92	-0.83	0.13
206	1.50	-1.12	0.28
207	0.67	-0.55	0.29
208	0.73	-0.15	0.26

Table 2. *Item parameters for the Oromifa PPVT analysis*

Item	Discrimination	Difficulty	Guessing
1	1.18	-2.20	0.47
2	1.21	-2.59	0.38
3	1.37	-2.25	0.60
4	1.15	-2.17	0.32
5	0.71	0.40	0.39
6	1.48	-2.06	0.26
7	1.12	-0.99	0.18
8	1.81	-0.75	0.24
9	1.34	-0.98	0.27
11	1.07	-1.68	0.29
12	0.80	-0.89	0.39
13	1.14	-1.63	0.28
14	0.97	-1.60	0.40
15	0.94	-1.25	0.31
16	0.72	-1.78	0.20
17	1.50	-0.72	0.25
18	0.61	1.64	0.25
19	0.83	-1.02	0.43
20	1.42	-0.71	0.23
21	0.82	-1.23	0.34
22	1.25	-0.90	0.32
23	0.92	0.60	0.20
24	0.93	0.12	0.37
25	1.60	-0.99	0.31
26	1.49	-0.62	0.27
27	0.71	0.98	0.06
28	1.37	-0.26	0.39
29	1.02	-0.98	0.23
30	0.72	-0.99	0.20
31	1.06	-0.20	0.26
32	1.62	-0.77	0.36
33	1.28	-0.37	0.38
34	0.88	0.35	0.22
35	0.72	0.62	0.23
36	0.82	-0.06	0.34
38	1.05	-1.88	0.17
39	0.70	-0.18	0.14
41	0.81	1.16	0.40
43	0.90	0.13	0.15
45	1.02	-0.12	0.28
46	1.26	0.00	0.15
47	1.18	-1.08	0.27
48	0.65	-1.72	0.35
49	0.70	1.13	0.19

Item	Discrimination	Difficulty	Guessing
50	1.10	-0.76	0.36
51	0.67	0.59	0.27
52	1.27	0.32	0.20
53	1.15	-0.27	0.16
54	0.62	0.04	0.19
55	1.20	0.89	0.40
56	0.79	1.84	0.25
57	1.24	-0.17	0.32
58	0.73	0.37	0.19
59	1.27	-0.88	0.39
60	1.30	0.42	0.25
61	1.28	-1.78	0.20
62	1.41	-1.34	0.13
63	1.06	-1.33	0.10
64	1.39	-0.60	0.33
65	0.84	1.78	0.54
67	0.80	0.41	0.35
68	0.70	-0.49	0.25
69	0.98	-0.02	0.34
70	0.55	1.11	0.23
71	0.67	0.88	0.23
74	1.13	-0.33	0.14
75	0.86	-0.15	0.23
77	1.13	-0.41	0.46
78	0.60	0.46	0.18
79	0.63	-0.83	0.17
80	0.94	0.80	0.21
81	1.10	1.68	0.21
82	1.26	0.77	0.21
83	1.22	1.04	0.38
84	0.83	0.79	0.38
85	0.58	0.63	0.34
86	0.72	1.53	0.37
87	1.06	1.17	0.17
88	1.17	2.12	0.24
90	0.97	-0.30	0.35
91	0.97	0.87	0.15
92	0.96	1.87	0.20
93	0.49	0.29	0.59
96	0.86	2.26	0.27
97	0.92	1.44	0.08
98	0.98	0.99	0.20
100	0.75	-1.42	0.36
101	0.85	-1.11	0.18

Item	Discrimination	Difficulty	Guessing
103	1.28	0.04	0.38
105	0.70	-0.17	0.20
106	1.23	1.00	0.25
107	0.66	-0.06	0.33
108	1.24	-0.75	0.17
109	1.31	-0.56	0.13
110	1.00	1.18	0.05
111	0.89	0.27	0.07
112	0.62	-0.21	0.19
113	0.81	1.24	0.08
114	0.98	-0.18	0.34
115	0.72	-0.06	0.11
116	0.92	-1.10	0.34
117	0.77	1.67	0.20
120	0.94	0.89	0.15
122	0.53	0.01	0.13
123	0.70	1.10	0.22
128	1.06	1.06	0.29
129	0.90	1.48	0.41
130	1.00	-0.18	0.14
131	0.76	1.51	0.16
133	0.75	0.74	0.17
134	1.18	-0.11	0.64
135	0.71	1.18	0.32
140	0.96	1.91	0.07

Item	Discrimination	Difficulty	Guessing
142	0.99	1.16	0.23
143	0.71	-1.44	0.43
145	1.02	-0.02	0.43
148	0.62	1.04	0.70
149	0.72	0.38	0.36
151	0.96	-0.10	0.08
152	0.84	-0.36	0.13
154	0.71	1.58	0.11
155	0.95	-0.15	0.55
156	0.82	0.46	0.12
157	0.73	-0.23	0.56
160	0.82	1.66	0.12
163	0.99	0.85	0.27
164	0.59	0.81	0.04
165	0.89	-0.02	0.14
168	0.86	0.64	0.16
170	0.51	-0.31	0.42
176	0.84	1.67	0.17
178	0.85	-1.15	0.52
179	0.94	0.79	0.32
182	0.90	2.07	0.22
197	0.91	0.88	0.21
201	0.78	-0.16	0.07
202	0.74	1.02	0.08
205	0.79	0.67	0.42

Table 3. *Item parameters for the Tigrigna PPVT analysis*

Item	Discrimination	Difficulty	Guessing
1	0.98	-3.00	0.49
2	0.91	-2.55	0.58
3	1.42	-2.12	0.55
4	1.59	-1.72	0.34
5	0.62	-1.67	0.19
6	1.50	-1.78	0.62
7	0.54	-0.58	0.11
8	1.57	-0.72	0.31
9	1.88	-0.83	0.34
10	0.65	0.03	0.32
11	0.91	-1.46	0.47
12	1.22	-0.21	0.67
13	1.89	-1.29	0.55
14	0.85	-1.35	0.70
15	1.13	-1.72	0.16
16	0.87	-1.51	0.30
17	1.05	-1.00	0.09
18	0.84	0.41	0.18
19	0.70	-1.25	0.20
20	0.97	-0.80	0.12
21	0.84	-1.15	0.18
22	1.26	-0.95	0.26
23	0.37	0.42	0.06
24	0.75	-0.34	0.33
25	1.51	-1.08	0.27
26	1.42	-0.78	0.11
27	1.89	-0.16	0.34
28	1.43	-0.82	0.36
29	1.38	-1.09	0.26
30	1.02	-0.10	0.35
31	0.94	-0.44	0.17
32	1.34	-0.94	0.12
33	1.08	-0.94	0.19
34	0.75	-0.09	0.12
35	0.73	-0.74	0.08
36	1.00	-0.90	0.26
38	1.35	-1.20	0.29
39	1.45	-1.27	0.33
40	0.77	-0.22	0.09
41	0.87	0.28	0.23
42	0.94	0.17	0.06
43	1.01	0.20	0.12
44	1.03	0.04	0.17
45	0.88	-0.28	0.22
46	1.23	0.06	0.14
47	1.62	-1.02	0.08
48	1.57	-1.06	0.10
49	0.81	0.29	0.06
50	0.95	-0.36	0.19
51	0.71	-0.30	0.08
52	0.60	0.78	0.09
53	0.76	0.49	0.04
54	0.65	-0.03	0.15
55	0.66	0.64	0.15
56	0.61	0.03	0.28
57	1.02	0.17	0.16
58	0.94	0.40	0.16
59	1.71	-0.41	0.71
60	1.44	-0.14	0.16
61	0.93	-1.03	0.81
62	0.64	-1.39	0.46
63	1.22	-0.87	0.18
64	1.24	-0.63	0.24
66	1.12	0.35	0.16
67	1.31	0.12	0.24
68	0.74	-0.34	0.16
69	0.87	-0.15	0.15
71	0.33	0.60	0.39
72	0.61	0.18	0.26
74	1.44	0.26	0.30
75	0.54	0.21	0.40
76	1.11	0.56	0.13
77	1.00	-0.64	0.37
78	0.90	-0.21	0.36
79	0.97	-1.32	0.21
80	0.69	0.60	0.09
81	1.12	0.55	0.07
82	0.75	0.09	0.13
83	0.99	-0.12	0.06
84	0.48	0.97	0.21
85	0.94	-0.19	0.18
86	1.11	1.02	0.04
87	0.97	0.41	0.27
88	0.60	1.82	0.15
89	0.84	1.19	0.12
90	1.07	0.34	0.20
91	1.37	0.33	0.54
92	0.45	2.10	0.05

Item	Discrimination	Difficulty	Guessing
94	0.79	0.09	0.38
95	1.27	0.12	0.11
96	0.89	0.67	0.08
97	0.94	0.90	0.09
99	0.67	1.57	0.43
101	0.44	0.42	0.05
102	0.97	1.62	0.17
103	0.77	-0.70	0.16
104	0.98	0.24	0.73
105	0.57	0.20	0.30
106	0.88	0.99	0.15
107	0.69	0.17	0.34
110	0.99	0.89	0.05
111	1.09	1.04	0.31
112	0.75	0.27	0.30
113	1.31	0.95	0.24
114	0.64	-0.71	0.36
115	0.92	1.64	0.69
116	0.83	-0.58	0.38
117	1.07	1.09	0.07
118	0.78	-0.14	0.50
120	0.77	0.16	0.07
121	0.62	0.69	0.38
122	0.83	1.75	0.17
123	2.05	1.08	0.21
124	1.07	1.67	0.10

Item	Discrimination	Difficulty	Guessing
125	1.02	0.87	0.21
126	0.50	-0.16	0.33
128	0.70	0.66	0.12
129	0.35	0.76	0.11
130	1.25	0.51	0.30
131	0.58	0.36	0.06
132	0.91	1.10	0.74
133	0.96	1.42	0.39
135	0.93	1.50	0.32
138	0.86	1.83	0.28
139	0.72	0.42	0.69
143	0.61	0.59	0.47
146	0.87	1.57	0.14
147	0.93	0.80	0.68
149	0.81	1.51	0.36
151	0.62	-0.18	0.07
152	0.97	1.17	0.24
154	0.51	1.26	0.10
155	1.03	0.50	0.38
157	0.72	-0.23	0.19
163	1.25	0.99	0.24
171	0.49	0.67	0.15
173	0.89	2.25	0.23
179	0.76	0.97	0.25
189	0.69	1.34	0.25

Table 4. *Item parameters for the Telugu PPVT analysis*

Item	Discrimination	Difficulty	Guessing
1	0.56	-3.30	0.69
2	0.62	-3.98	0.13
3	1.56	-3.23	0.17
4	1.17	-2.81	0.29
5	0.68	-1.07	0.21
6	0.69	-3.69	0.18
7	0.86	-2.60	0.10
8	0.80	-0.80	0.16
9	0.83	-1.97	0.32
10	0.97	-0.80	0.39
11	1.72	-1.51	0.48
12	1.34	-1.64	0.30
13	0.91	-1.35	0.25
14	1.41	-1.24	0.57
15	0.90	-2.13	0.12
16	1.74	-0.95	0.48
17	1.08	-1.64	0.19
18	0.97	-0.45	0.22
19	1.27	-0.41	0.45
20	1.57	-0.34	0.31
21	0.61	-0.06	0.13
22	0.52	-1.22	0.04
23	0.87	-0.55	0.21
24	0.54	-0.06	0.23
25	1.18	-1.85	0.10
26	1.12	-1.65	0.04
27	1.27	0.39	0.18
28	1.20	-1.30	0.19
29	0.50	-1.08	0.18
30	0.42	0.15	0.42
31	1.62	-0.50	0.31
32	0.54	-0.69	0.31
33	1.10	-1.13	0.24
34	1.34	-0.14	0.30
35	1.90	-0.17	0.34
36	0.68	-1.50	0.03
37	0.95	0.47	0.33
38	0.74	-2.36	0.04
39	0.91	-0.03	0.25
40	0.48	0.99	0.03
41	0.97	-0.68	0.18
42	2.08	1.66	0.13
43	0.62	0.95	0.16
44	0.47	1.54	0.17
45	1.27	1.37	0.20
46	1.24	0.28	0.20
47	1.19	-1.56	0.04
48	0.67	-1.88	0.40
49	1.06	1.20	0.22
50	0.79	0.67	0.21
51	1.29	0.38	0.20
52	1.04	-0.70	0.10
53	1.13	1.37	0.14
54	0.50	1.08	0.33
55	0.62	-0.71	0.04
56	1.02	0.27	0.20
57	0.44	1.02	0.18
58	0.74	0.24	0.02
59	0.92	-1.32	0.04
60	0.77	-0.59	0.17
61	2.31	-0.17	0.24
62	0.99	0.48	0.15
63	0.45	-1.10	0.14
64	1.42	-1.11	0.08
65	0.73	2.37	0.31
67	0.68	0.74	0.21
68	1.66	-0.58	0.36
69	0.66	-1.23	0.04
70	1.28	-0.61	0.32
71	0.79	-0.91	0.24
72	2.00	0.41	0.29
73	0.44	0.29	0.16
74	1.31	-1.36	0.05
75	0.87	-0.35	0.19
76	1.46	0.80	0.26
77	1.03	-0.58	0.22
78	1.10	-1.18	0.09
79	0.58	-0.15	0.30
80	1.25	0.24	0.28
81	0.80	0.77	0.24
82	0.77	-0.47	0.31
83	1.24	-0.14	0.24
84	0.86	1.45	0.37
86	1.07	1.60	0.16
87	0.77	-0.41	0.37
88	0.93	-0.31	0.07
89	0.99	1.61	0.27
90	1.52	0.30	0.25

Item	Discrimination	Difficulty	Guessing
91	0.78	0.23	0.18
92	0.51	1.16	0.20
93	0.99	-1.11	0.34
95	2.28	0.31	0.35
96	1.10	-0.22	0.11
97	1.10	1.00	0.22
98	0.55	0.65	0.12
99	0.84	-0.18	0.16
100	2.13	1.14	0.10
101	1.10	-0.37	0.24
103	0.96	1.91	0.18
105	1.51	0.07	0.34
106	0.99	0.30	0.18
107	1.13	1.52	0.30
108	0.54	-0.37	0.16
109	1.50	-0.56	0.69
110	0.40	-1.27	0.20
111	1.51	0.61	0.29
112	1.67	0.68	0.39
113	0.60	1.98	0.22
114	1.04	0.89	0.21
115	0.72	-0.93	0.08
116	1.14	0.24	0.28
117	1.25	1.16	0.10
118	1.22	0.35	0.20
119	2.27	0.31	0.21
120	1.09	0.98	0.13
121	0.89	0.23	0.09
122	0.80	0.63	0.17
123	1.53	0.39	0.39
125	0.99	1.00	0.37
126	1.20	-1.07	0.44
127	1.07	2.48	0.11
128	0.64	-1.64	0.29
129	0.55	0.77	0.14
130	1.10	0.77	0.27
131	0.56	-0.33	0.27
132	0.45	-1.94	0.36
133	1.04	-0.92	0.19

Item	Discrimination	Difficulty	Guessing
134	1.14	1.78	0.06
135	0.79	1.96	0.23
140	0.63	0.55	0.17
141	0.65	1.22	0.26
142	0.62	0.27	0.03
143	1.62	0.86	0.30
144	1.39	1.12	0.17
146	1.12	1.83	0.39
148	1.69	1.19	0.13
149	0.60	0.75	0.19
150	0.92	0.59	0.30
151	0.36	0.98	0.19
153	0.54	-0.93	0.19
154	0.77	0.45	0.22
155	1.65	0.43	0.28
156	1.13	0.47	0.29
157	1.28	-0.18	0.40
159	0.87	2.31	0.22
160	1.22	1.33	0.24
161	0.42	0.26	0.22
163	0.70	1.69	0.13
164	0.49	-1.07	0.08
168	0.98	1.09	0.11
169	1.23	1.43	0.16
170	1.61	1.92	0.18
171	0.45	1.78	0.08
172	1.10	2.29	0.20
175	1.42	1.16	0.27
176	0.97	1.86	0.30
178	1.85	0.76	0.29
180	1.09	1.34	0.11
189	0.71	1.45	0.30
190	0.40	-0.25	0.08
205	0.83	0.01	0.40
206	1.28	1.06	0.11
207	0.64	-0.21	0.31
208	1.03	0.84	0.28

Table 5. *Item parameters for the Spanish TVIP analysis*

Item	Discrimination	Difficulty	Guessing
1	1.58	-2.52	0.33
2	2.79	-1.45	0.32
3	1.58	-2.38	0.74
4	1.45	-2.94	0.57
5	2.34	-1.51	0.39
6	1.56	-2.05	0.60
7	2.36	-1.60	0.38
8	2.21	-1.68	0.27
9	1.74	-2.20	0.18
10	1.59	-2.15	0.22
11	1.80	-2.66	0.04
12	2.16	-1.42	0.22
13	1.40	-2.27	0.04
14	1.61	-2.27	0.01
15	1.85	-2.12	0.01
16	1.77	-1.81	0.01
17	1.86	-1.95	0.01
18	1.96	-1.70	0.01
19	1.96	-1.93	0.01
20	1.86	-1.44	0.03
21	1.62	-1.69	0.01
22	1.72	-1.81	0.01
23	1.93	-1.66	0.00
24	2.09	-1.58	0.00
25	2.28	-1.52	0.00
26	2.06	-1.62	0.00
27	2.39	-1.67	0.00
28	2.09	-1.60	0.00
29	2.02	-1.66	0.00
30	2.04	-1.16	0.00
31	2.12	-1.34	0.00
32	1.63	-1.52	0.00
33	2.13	-1.24	0.00
34	6.25	-1.72	0.00
35	1.97	-1.00	0.00
36	1.74	-1.17	0.00
37	1.89	-1.02	0.00
38	1.69	-0.94	0.00
39	2.21	-1.09	0.00
40	2.01	-0.96	0.00
41	1.98	-1.17	0.00
42	1.86	-0.76	0.00
43	2.24	-1.00	0.00
44	2.03	-0.75	0.00
45	1.81	-0.65	0.00
46	2.10	-0.97	0.00
47	2.37	-0.93	0.00
48	2.14	-0.80	0.00
49	1.79	-0.89	0.00
50	1.88	-0.73	0.00
51	1.67	-0.44	0.00
52	1.97	-0.75	0.00
53	2.33	-0.66	0.00
54	1.73	-0.51	0.00
55	2.04	-0.53	0.00
56	1.99	-0.35	0.00
57	1.59	-0.55	0.00
58	1.77	0.02	0.00
59	2.47	-0.65	0.00
60	2.84	-0.54	0.00
61	1.98	0.27	0.00
62	1.50	0.21	0.00
63	1.82	-0.11	0.00
64	2.63	-0.34	0.00
65	2.22	-0.43	0.00
66	1.72	-0.22	0.00
67	1.79	0.03	0.00
68	1.62	-0.08	0.00
69	1.91	0.04	0.00
70	1.97	-0.14	0.00
71	2.18	0.09	0.00
72	1.36	0.42	0.00
73	1.87	-0.04	0.00
74	2.01	0.11	0.00
75	1.39	0.58	0.00
76	1.74	0.12	0.00
77	1.30	0.51	0.00
78	1.45	0.70	0.00
79	1.58	0.25	0.00
80	1.25	0.54	0.00
81	1.03	1.05	0.00
82	1.90	0.28	0.00
83	1.74	0.82	0.00
84	1.49	0.73	0.00
85	1.52	0.72	0.00
86	2.21	0.62	0.00
87	1.64	0.96	0.00
88	2.17	0.67	0.00

Item	Discrimination	Difficulty	Guessing
89	1.63	0.99	0.00
90	1.59	1.21	0.00
91	1.59	1.26	0.00
92	2.04	1.13	0.00
93	1.99	1.20	0.00
94	2.79	1.09	0.00
95	3.59	0.87	0.00
96	2.49	1.28	0.00
97	2.53	1.20	0.00
98	2.30	1.32	0.00
99	5.10	0.96	0.00
100	3.73	1.09	0.00
101	2.25	1.58	0.00
102	2.51	1.65	0.00
103	2.89	1.43	0.00
104	3.96	1.20	0.00
105	2.90	1.52	0.00
106	2.60	1.48	0.00
107	3.62	1.38	0.00

Item	Discrimination	Difficulty	Guessing
108	4.50	1.37	0.00
109	4.61	1.44	0.00
110	2.99	1.69	0.00
111	8.26	1.29	0.00
112	4.43	1.52	0.00
113	4.87	1.48	0.00
114	3.08	1.68	0.00
115	3.55	1.68	0.00
116	3.99	1.62	0.00
117	9.08	1.44	0.00
118	4.73	1.66	0.00
119	6.33	1.52	0.00
120	4.06	1.70	0.00
121	7.90	1.52	0.00
122	3.77	1.79	0.00
123	6.76	1.58	0.00
124	5.32	1.70	0.00
125	3.91	1.82	0.00

Table 6. *Item parameters for the Vietnamese PPVT analysis*

Item	Discrimination	Difficulty	Guessing
1	1.41	-3.09	0.48
2	1.92	-3.61	0.20
3	2.06	-3.85	0.41
4	1.52	-3.02	0.16
5	0.81	-2.27	0.06
6	1.14	-3.24	0.27
7	1.32	-2.95	0.10
8	0.75	-2.54	0.08
9	1.40	-2.35	0.35
11	1.25	-1.91	0.49
12	0.95	-2.37	0.10
13	1.09	-2.73	0.37
14	1.24	-1.23	0.59
15	2.07	-2.18	0.48
16	0.84	-1.92	0.11
17	0.99	-2.14	0.04
19	0.70	-1.35	0.65
20	0.62	-1.80	0.03
21	0.69	-2.40	0.37
22	0.55	-1.51	0.12
23	0.93	-2.08	0.05
24	1.05	-2.03	0.12
25	0.54	-1.55	0.04
26	1.73	-1.96	0.51
27	1.81	-1.82	0.40
28	0.82	-2.99	0.05
29	1.20	-1.79	0.45
30	1.20	-1.97	0.52
31	0.99	-2.15	0.02
32	1.46	-2.39	0.42
33	0.70	-2.61	0.05
34	1.43	-1.70	0.29
35	0.98	-1.81	0.02
36	1.23	-1.06	0.37
37	0.47	-0.38	0.13
38	1.33	-2.53	0.41
39	1.44	-0.90	0.26
40	1.31	-0.37	0.21
41	0.71	-0.58	0.18
42	0.66	0.63	0.04
43	1.25	-1.26	0.17
44	0.70	-1.24	0.01
45	1.10	-1.09	0.11
46	2.04	-1.22	0.31

Item	Discrimination	Difficulty	Guessing
47	1.75	-1.86	0.31
48	0.90	-1.19	0.10
49	1.07	-1.77	0.49
50	1.24	-0.65	0.24
51	0.95	-1.70	0.19
52	1.01	-0.92	0.03
54	0.63	-0.05	0.34
55	1.15	-0.95	0.23
56	0.80	-0.23	0.17
57	0.89	-1.37	0.22
58	2.25	-0.62	0.41
59	0.82	-1.09	0.02
60	1.67	-0.91	0.08
61	0.60	-1.92	0.18
62	0.82	0.21	0.12
63	2.70	-1.03	0.43
64	0.88	-1.66	0.03
65	0.28	-1.63	0.02
66	0.74	-0.25	0.07
68	1.12	-1.58	0.11
69	1.05	-1.38	0.17
70	0.76	-1.42	0.04
71	1.90	-0.55	0.55
72	1.47	-1.19	0.44
74	1.29	-1.55	0.07
75	0.73	-2.06	0.04
76	1.03	-0.63	0.05
77	1.15	-1.40	0.09
78	0.82	-1.34	0.03
79	1.44	-1.27	0.26
80	1.16	-0.76	0.06
81	1.00	-0.75	0.03
82	1.20	-0.90	0.08
83	1.33	-0.71	0.09
84	0.79	-0.65	0.04
86	0.93	2.11	0.18
87	0.70	0.80	0.12
88	0.63	-0.74	0.02
90	0.72	-0.99	0.02
91	1.02	1.54	0.12
92	1.05	0.95	0.25
94	0.89	0.81	0.14
95	1.73	-0.12	0.24
96	1.18	-0.43	0.05

Item	Discrimination	Difficulty	Guessing
97	0.84	-0.70	0.09
98	0.77	-0.12	0.05
99	0.75	-1.14	0.11
100	1.36	-0.10	0.18
101	0.80	-1.33	0.19
102	0.52	1.63	0.15
103	1.03	-1.19	0.24
104	0.78	-1.47	0.27
105	1.52	-0.43	0.22
106	1.14	-0.24	0.14
107	1.41	0.09	0.32
108	0.81	-0.91	0.03
109	0.88	0.33	0.12
110	0.70	-0.13	0.42
111	0.88	0.13	0.21
112	0.74	-0.66	0.05
114	0.93	-0.90	0.15
115	1.05	0.15	0.06
116	1.02	-0.48	0.10
117	1.79	0.20	0.16
118	0.83	-2.14	0.11
119	0.76	-2.09	0.52
120	0.56	0.10	0.05
121	1.28	-0.17	0.21
122	0.57	0.07	0.14
123	1.13	-0.28	0.16
124	0.98	1.82	0.12
125	0.81	0.16	0.11
126	0.93	0.52	0.14
128	0.69	-0.21	0.03
129	0.74	-1.11	0.34
130	1.19	-0.25	0.16
131	0.93	0.06	0.07
132	1.90	-0.47	0.43
133	0.70	-1.27	0.26
134	1.13	-0.85	0.30
135	0.90	0.22	0.20
136	0.59	2.02	0.06
137	0.90	1.45	0.19
138	1.05	1.01	0.21
139	0.60	0.20	0.10
140	0.38	0.93	0.10
141	0.62	0.57	0.31

Item	Discrimination	Difficulty	Guessing
142	0.86	0.77	0.16
144	1.48	-0.11	0.24
145	0.59	0.26	0.18
146	1.38	0.90	0.27
149	0.63	0.10	0.06
151	0.74	-0.57	0.52
152	0.77	0.09	0.25
153	0.51	-0.47	0.22
154	0.92	0.38	0.23
155	1.32	0.34	0.06
156	1.26	0.23	0.11
157	0.97	0.72	0.31
158	1.12	1.67	0.17
159	0.77	1.86	0.21
161	0.33	-1.20	0.48
162	0.71	2.48	0.36
163	1.59	1.66	0.20
164	1.00	1.26	0.20
165	0.82	1.04	0.36
166	0.67	-0.58	0.24
167	0.58	-1.43	0.08
168	0.69	0.48	0.04
169	1.11	-0.28	0.37
170	0.40	-1.02	0.13
171	0.75	-0.18	0.04
172	0.68	1.42	0.17
173	0.67	2.37	0.19
175	0.64	-0.05	0.05
176	0.62	-0.42	0.27
178	0.81	0.06	0.04
179	0.63	-0.58	0.25
180	0.43	1.17	0.08
181	0.95	0.81	0.25
182	0.41	-0.89	0.32
189	0.89	0.04	0.47
192	0.35	0.58	0.14
196	1.50	-0.43	0.08
197	0.80	0.32	0.30
199	0.67	-1.56	0.13
204	0.80	0.59	0.51
205	0.78	0.04	0.37
206	0.64	-1.24	0.24

Table 7. *Item parameters for the Ethiopia maths analysis*

Item	Discrimination	Difficulty
1	1.43	0.07
2	1.75	0.03
3	1.42	-0.72
5	2.88	-0.99
6	2.26	-0.60
7	3.05	-0.35
8	2.75	-0.36
9	3.18	-0.41
10	2.92	0.26
11	2.79	-0.15
12	2.20	-0.09
13	3.19	0.37
14	1.77	0.82
15	1.75	0.51
16	2.28	0.43
17	2.67	0.77
19	1.55	0.25
20	1.66	0.52
21	1.34	-0.17
22	1.75	0.50
23	1.54	0.49
24	1.01	0.01
25	1.12	-0.16
26	1.65	1.02
27	0.91	1.41
28	1.25	0.42
29	1.80	1.00
30	1.48	0.86
31	2.38	1.04
32	1.50	1.66
33	1.23	1.23
34	1.48	2.26
35	1.39	1.94
36	2.20	2.20
37	2.15	2.13
38	1.33	2.21
39	2.44	1.97
44	1.02	-0.23
45	1.34	-0.54
46	0.75	0.13

Item	Discrimination	Difficulty
47	0.73	0.15
48	1.05	0.61
49	0.54	-0.33
50	1.08	2.07
51	0.86	1.06
53	0.33	1.67
54	0.84	2.00
56	0.37	1.66
57	0.36	2.41
59	0.97	3.14
60	1.33	2.47
61	1.04	2.68
62	0.46	0.76
63	0.96	1.16
69	0.36	1.46
71	0.34	2.65
84	0.57	0.88
85	0.55	1.06
86	0.27	4.16
88	0.76	0.55
92	0.35	3.10
94	0.28	2.23
95	0.20	2.50
96	2.25	-1.05
97	1.96	-0.87
98	1.52	-0.94
99	1.42	-2.10
100	1.41	-0.02
101	1.23	0.35
102	1.43	-1.64
103	1.57	-0.63
104	0.58	1.11
105	0.48	1.16
106	0.26	3.74
107	0.21	3.52
108	0.33	1.72
111	0.39	1.51
200	0.68	1.04

Table 8. *Item parameters for the India maths analysis*

Item	Discrimination	Difficulty
1	1.13	-0.73
2	1.74	-0.24
3	1.75	-1.75
4	1.16	-1.42
5	1.50	-1.97
6	1.70	-1.34
7	2.32	-0.96
8	2.03	-1.28
9	2.45	-1.08
10	2.43	-0.53
11	1.71	-0.65
12	1.92	-0.66
13	2.71	-0.50
14	1.38	0.13
15	1.73	-0.07
16	1.87	0.02
17	1.85	0.31
18	1.70	-0.19
19	2.74	0.01
20	2.10	-0.04
21	2.54	-0.01
22	2.73	0.13
23	2.23	0.00
24	1.03	-0.84
25	0.89	-1.18
26	2.34	0.32
27	1.92	0.41
28	2.18	0.10
29	1.33	0.29
30	1.45	0.28
31	1.31	1.65
32	1.17	1.19
33	1.33	0.99
34	1.06	1.27
35	1.58	1.36
36	2.00	1.68
37	1.66	1.41
38	1.14	2.13
39	1.54	2.13
40	1.17	-0.03
41	1.85	-0.09
42	0.42	1.82
44	1.02	-0.32
45	1.20	-0.80
46	1.11	0.05
47	1.05	-0.31
48	1.05	0.86
49	1.07	-0.45
50	1.14	0.71
51	0.87	0.80
53	0.42	1.10
54	0.91	0.36
55	0.42	2.92
56	0.62	0.89
57	0.49	1.55
59	1.27	1.10
60	1.64	1.06
61	1.15	1.78
62	0.66	0.94
63	1.10	0.66
64	0.34	2.63
69	0.44	0.75
71	0.50	1.76
74	0.48	2.22
84	1.03	0.46
85	0.91	0.77
86	0.47	2.65
87	0.37	4.48
92	0.45	1.89
93	0.27	4.18
94	0.52	1.38
95	0.29	2.86
96	1.94	-2.40
97	2.04	-1.89
98	1.66	-1.89
99	1.34	-2.87
100	1.26	-0.43
101	1.38	-0.02
102	1.19	-2.04
103	1.70	-1.40
104	0.76	1.04
105	0.94	0.42
106	0.35	2.29
107	0.33	2.43
108	0.39	1.79
109	0.33	2.89
111	0.59	1.22
112	0.60	2.39
114	0.40	1.99
200	0.95	0.44

Table 9. *Item parameters for the Peru maths analysis*

Item	Discrimination	Difficulty
2	1.55	-0.56
3	1.57	-1.40
5	0.93	-3.21
6	0.78	-2.26
7	1.09	-1.95
8	1.25	-1.95
9	1.43	-1.64
10	1.52	-1.19
11	1.43	-1.19
12	1.70	-1.10
13	1.78	-1.09
14	1.21	-0.14
15	1.00	-0.78
16	1.73	-0.13
17	1.65	0.11
18	2.64	-0.13
19	2.82	-0.17
20	2.32	-0.32
21	2.78	-0.22
22	2.34	0.00
23	2.42	-0.18
25	0.70	-2.08
26	1.62	-0.22
27	1.72	-0.29
28	1.35	-0.28
29	0.73	0.04
30	1.03	0.17
31	1.14	0.51
32	1.04	0.88
33	0.79	1.25
34	0.73	1.66
35	1.40	1.19
36	1.56	1.41
37	1.10	2.31
38	1.26	1.65
39	1.78	1.54
40	0.99	-0.45
41	1.52	-0.53
42	0.33	2.05
44	1.07	-0.45
45	1.44	-1.24
46	1.12	-0.33
47	0.79	-0.90
48	1.68	0.20
49	0.76	-0.36
50	1.01	1.12
51	0.75	0.63
54	0.75	0.30
55	0.69	0.97
56	0.69	0.57
57	0.54	1.49
59	1.00	1.37
60	1.48	1.35
61	1.23	1.88
62	0.75	0.53
63	1.53	0.04
64	1.74	0.43
69	0.31	0.80
71	0.60	1.54
74	0.45	1.96
84	1.12	0.21
85	1.32	0.19
86	0.85	1.23
92	0.86	1.29
93	0.56	2.25
94	0.31	2.02
95	0.36	1.25
96	1.60	-2.77
97	1.81	-2.37
98	1.97	-2.48
99	0.86	-3.81
100	1.28	-1.00
101	1.47	-0.47
102	1.06	-2.31
103	1.32	-1.01
104	1.29	0.37
105	1.07	-0.16
106	0.69	1.33
107	0.47	1.05
108	0.51	1.38
109	0.26	2.90
111	0.82	0.29
112	0.68	2.09
114	0.31	2.66
200	1.16	0.11
201	0.29	3.06
202	0.54	1.84

Table 10. *Item parameters for the Vietnam maths analysis*

Item	Discrimination	Difficulty
1	0.93	-1.94
2	0.92	-1.29
3	1.03	-1.88
6	0.71	-3.82
7	0.66	-3.11
8	0.68	-2.53
9	0.69	-2.42
10	0.68	-0.41
11	0.81	-2.25
12	0.91	-2.05
13	0.83	-0.22
14	0.81	0.64
15	0.83	0.16
16	1.04	-0.13
17	1.15	0.09
18	2.11	-0.10
19	2.82	-0.22
20	2.65	0.36
21	2.90	-0.10
22	1.57	-1.11
23	0.98	-1.62
24	0.66	-2.53
25	0.68	-2.80
26	1.59	-0.94
27	1.38	-1.06
28	1.08	-1.40
29	0.69	-0.70
30	0.77	-1.04
31	1.37	-0.94
32	0.89	-0.81
33	1.03	-0.46
34	1.21	-0.01
35	1.22	0.09
36	1.57	0.43
37	1.24	0.75
38	1.00	0.51
39	1.64	0.15
42	0.68	-0.21
43	0.81	0.11
44	0.87	-1.92
45	1.61	-2.19
46	1.15	-1.19
47	0.71	-0.94
48	1.31	-0.95

Item	Discrimination	Difficulty
49	0.89	-1.66
50	1.04	-0.27
51	0.88	-0.52
52	0.62	0.87
53	0.54	-2.36
54	0.70	-1.09
55	0.51	-0.24
56	0.81	-0.68
57	0.64	0.10
58	0.69	0.06
59	1.20	0.34
60	1.59	0.34
61	1.28	0.25
62	1.00	-0.59
65	1.21	0.56
66	0.38	2.10
67	0.74	0.48
68	0.34	2.04
69	0.30	-0.25
70	0.45	1.53
71	0.85	0.11
72	0.92	-0.50
73	0.81	-0.54
74	0.72	0.46
76	0.49	1.18
77	0.80	0.82
78	0.67	0.05
80	0.83	1.49
81	1.49	1.90
82	1.58	1.66
83	1.32	1.31
84	1.01	-0.64
85	0.89	-0.63
86	0.64	0.69
87	0.36	2.91
92	0.78	0.33
93	0.53	1.15
94	0.49	0.29
95	0.19	2.46
97	0.71	-3.36
98	1.46	-2.99
100	0.71	-0.43
101	0.76	0.05
102	0.59	-3.00

Item	Discrimination	Difficulty
103	0.80	-1.51
104	1.08	-0.60
105	0.78	-0.58
106	0.81	0.16
107	0.48	0.55
108	0.72	0.10
109	0.72	0.43
110	1.15	1.71
111	1.01	-0.46
112	0.80	0.29

Item	Discrimination	Difficulty
113	0.95	1.41
114	0.49	0.69
115	1.19	1.89
116	1.38	1.45
200	1.07	2.02
201	2.29	0.52
202	0.67	-1.05
203	0.97	-1.86
204	0.76	-1.43

Table 11. *Item parameters for the Amharic reading comprehension analysis*

Item	Discrimination	Difficulty	Guessing
4	0.86	-1.99	0.49
5	0.50	-0.58	0.73
6	0.52	0.08	0.36
8	0.94	-1.85	0.24
9	0.94	1.26	0.53
11	0.67	-1.45	0.09
13	2.15	-0.80	0.11
14	0.73	-0.57	0.12
15	2.36	-0.72	0.13
16	1.86	-0.40	0.22
17	2.38	-0.49	0.20
18	2.00	-0.62	0.07
19	1.84	-1.01	0.16
20	1.38	-0.35	0.20
21	2.02	-0.33	0.17
22	1.41	-0.74	0.08
23	0.98	-0.06	0.14

Item	Discrimination	Difficulty	Guessing
24	0.47	0.22	0.04
31	0.62	-0.51	0.09
32	1.58	-0.05	0.25
33	0.61	1.01	0.15
35	1.03	1.43	0.24
36	1.04	1.78	0.21
37	0.91	0.48	0.14
38	1.08	0.34	0.17
39	0.57	0.40	0.07
40	0.56	0.67	0.08
42	0.79	0.47	0.13
49	0.76	0.41	0.05
50	1.41	1.64	0.17
51	0.98	1.29	0.22
52	0.82	1.45	0.14
53	0.83	1.70	0.12
54	0.58	2.09	0.24

Table 12. *Item parameters for the Oromifa reading comprehension analysis*

Item	Discrimination	Difficulty	Guessing
4	1.19	-0.77	0.68
5	0.78	-0.80	0.43
6	0.49	1.09	0.30
8	0.53	0.27	0.45
9	0.45	-0.70	0.21
11	1.17	-0.51	0.49
13	1.20	0.33	0.15
14	1.58	0.02	0.19
15	1.99	0.16	0.18
16	1.55	0.20	0.13
17	3.07	0.40	0.22
18	2.09	0.42	0.08
19	2.52	-0.38	0.11
20	1.65	0.40	0.22
21	1.43	0.69	0.17

Item	Discrimination	Difficulty	Guessing
22	1.74	0.13	0.13
23	0.94	0.48	0.15
24	0.61	1.84	0.10
31	0.81	0.58	0.05
32	0.91	0.34	0.05
33	0.97	1.82	0.20
34	0.76	2.81	0.05
35	0.83	2.48	0.21
38	1.19	0.76	0.18
39	0.94	2.07	0.15
40	0.99	1.29	0.15
41	0.51	2.37	0.15
42	0.73	1.28	0.05
49	1.04	1.56	0.25
51	1.02	2.39	0.13

Table 13. *Item parameters for the Tigrigna reading comprehension analysis*

Item	Discrimination	Difficulty	Guessing
4	0.99	-1.48	0.35
5	0.64	-1.77	0.46
6	0.93	0.16	0.42
8	0.62	-0.85	0.15
9	0.46	-1.15	0.49
11	0.86	-1.12	0.07
13	2.72	-0.19	0.11
15	2.05	-0.52	0.14
16	2.42	-0.05	0.29
17	2.52	0.03	0.21
18	1.98	-0.19	0.07
19	1.64	-0.67	0.26
20	1.16	-0.05	0.17
21	1.43	-0.15	0.13
22	1.31	-0.66	0.07
23	0.92	0.18	0.15

Item	Discrimination	Difficulty	Guessing
24	0.73	1.57	0.22
31	1.06	0.14	0.17
32	1.09	0.52	0.29
33	0.71	1.20	0.15
35	1.09	2.39	0.21
36	1.18	2.56	0.23
37	1.04	1.47	0.24
38	0.74	0.60	0.11
39	0.72	1.89	0.19
40	0.60	0.83	0.06
41	0.86	2.94	0.28
42	1.15	1.34	0.16
49	1.06	1.36	0.16
51	0.94	2.62	0.16
52	0.75	2.90	0.11
53	1.10	1.72	0.10

Table 14. *Item parameters for the Telugu reading comprehension analysis*

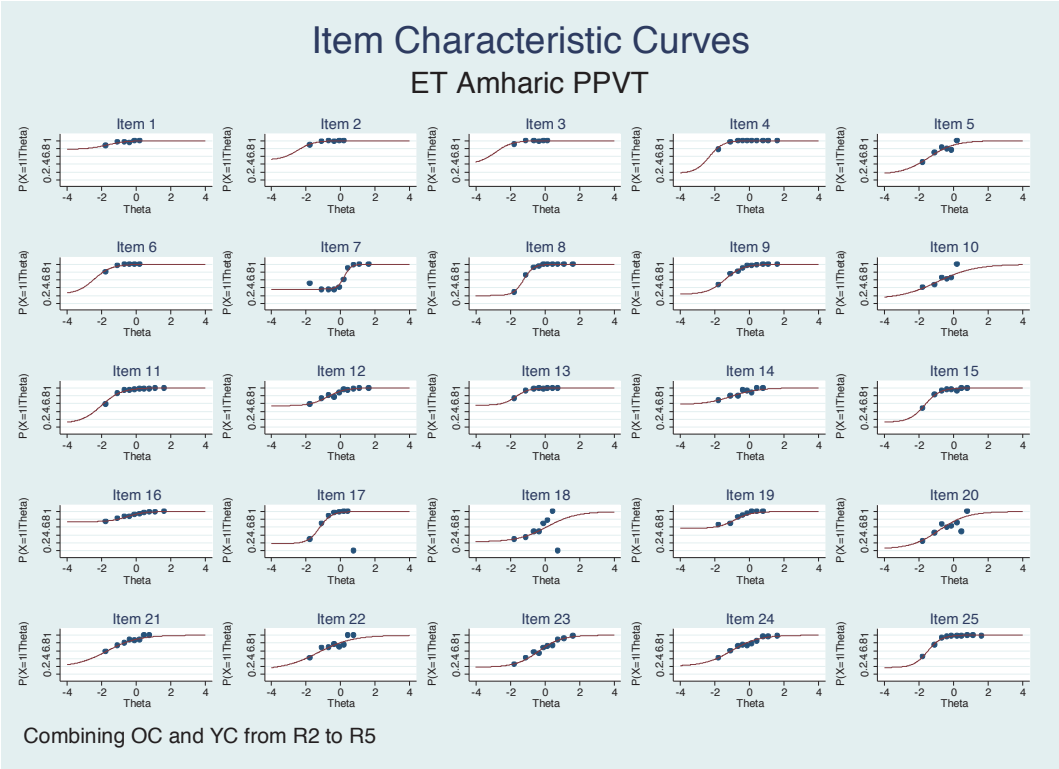
Item	Discrimination	Difficulty	Guessing
1	1.01	-2.20	0.19
2	1.17	-2.17	0.08
3	0.66	-1.67	0.03
8	0.89	-1.86	0.05
9	0.88	-1.48	0.19
10	1.21	-1.74	0.12
11	1.15	-1.64	0.06
19	1.32	-1.62	0.13
20	1.24	-0.74	0.09
21	1.24	-0.58	0.08
22	1.52	-0.90	0.09
23	0.94	-0.17	0.05
24	0.61	0.70	0.03
25	0.74	2.42	0.14
26	0.93	0.87	0.12
27	1.62	-0.29	0.13
28	0.64	0.55	0.17
29	0.40	0.06	0.03
31	0.60	-0.06	0.02
32	1.00	-0.38	0.03
33	0.69	0.96	0.06
34	0.60	0.52	0.02
35	0.78	0.70	0.08
36	0.84	1.55	0.15
37	0.94	1.41	0.11
38	1.12	-0.25	0.06
39	0.61	0.55	0.09
40	0.60	0.38	0.02
41	0.36	1.43	0.06
42	0.61	0.73	0.08
49	0.59	0.88	0.04

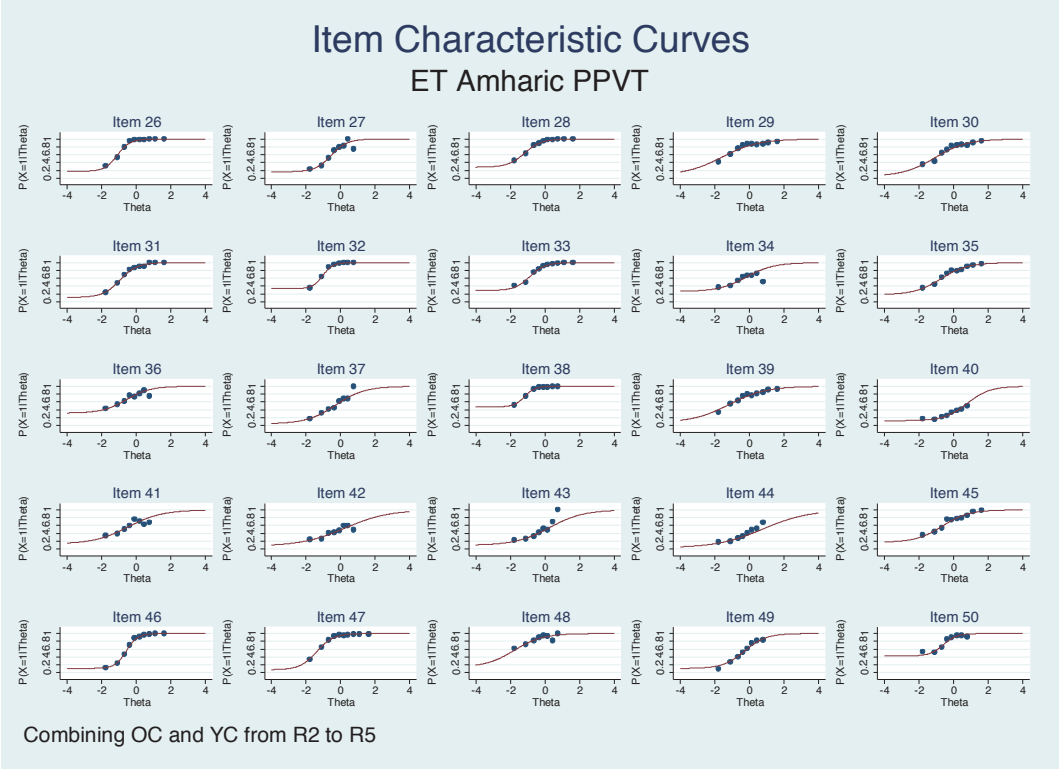
Table 15. *Item parameters for the Spanish reading comprehension analysis*

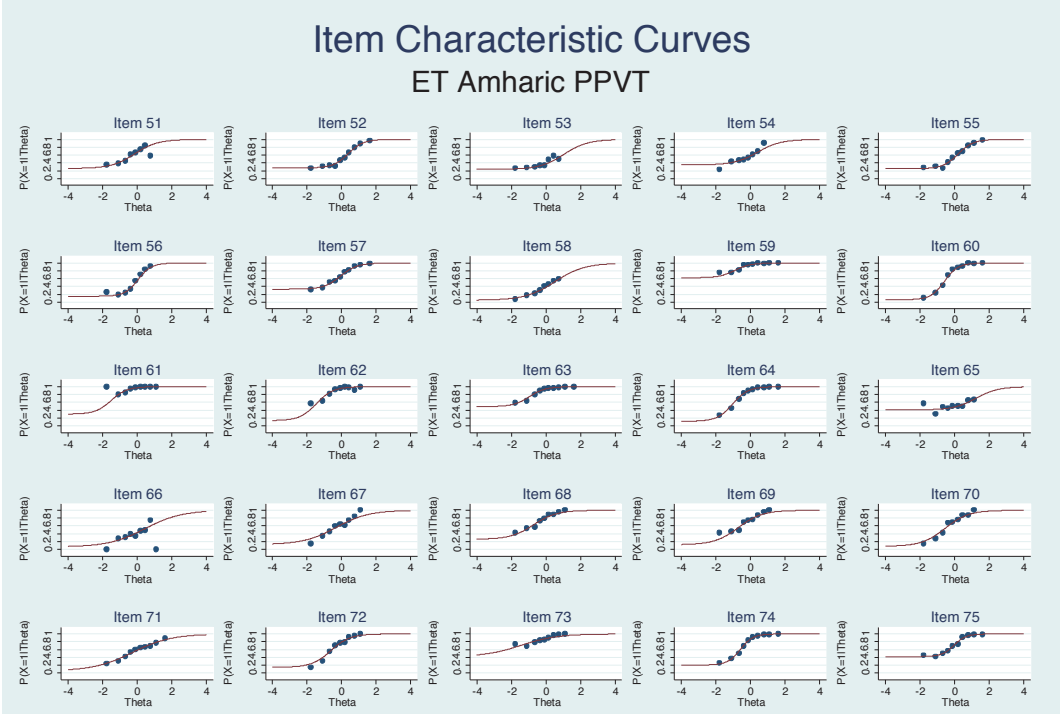
Item	Discrimination	Difficulty	Guessing
1	2.09	-2.27	0.03
2	2.07	-2.61	0.04
3	0.88	-1.61	0.06
7	1.06	-1.96	0.03
11	0.43	-3.37	0.08
12	0.84	-2.72	0.04
19	1.63	-2.10	0.10
20	1.29	-1.15	0.05
21	0.82	-1.33	0.04
22	1.71	-1.31	0.05
23	0.93	-1.16	0.06
24	0.74	-1.19	0.04
25	0.59	-0.36	0.15
26	0.52	-0.14	0.10
27	1.38	-1.49	0.06
28	0.66	0.32	0.23
29	0.47	1.61	0.07
31	0.87	-0.51	0.05
32	1.17	-1.18	0.03
33	0.75	-0.16	0.12
35	0.59	-0.11	0.08
36	1.50	1.18	0.15
37	0.47	0.36	0.06
38	1.14	-0.28	0.30
39	1.18	-0.05	0.21
40	0.36	0.80	0.04
41	0.38	1.46	0.13
42	1.49	0.18	0.12
49	0.62	-0.52	0.04
50	0.90	1.43	0.15
51	0.51	0.53	0.02
52	1.41	1.77	0.12
53	0.34	2.19	0.02

Appendix B.
Item Characteristic Curves

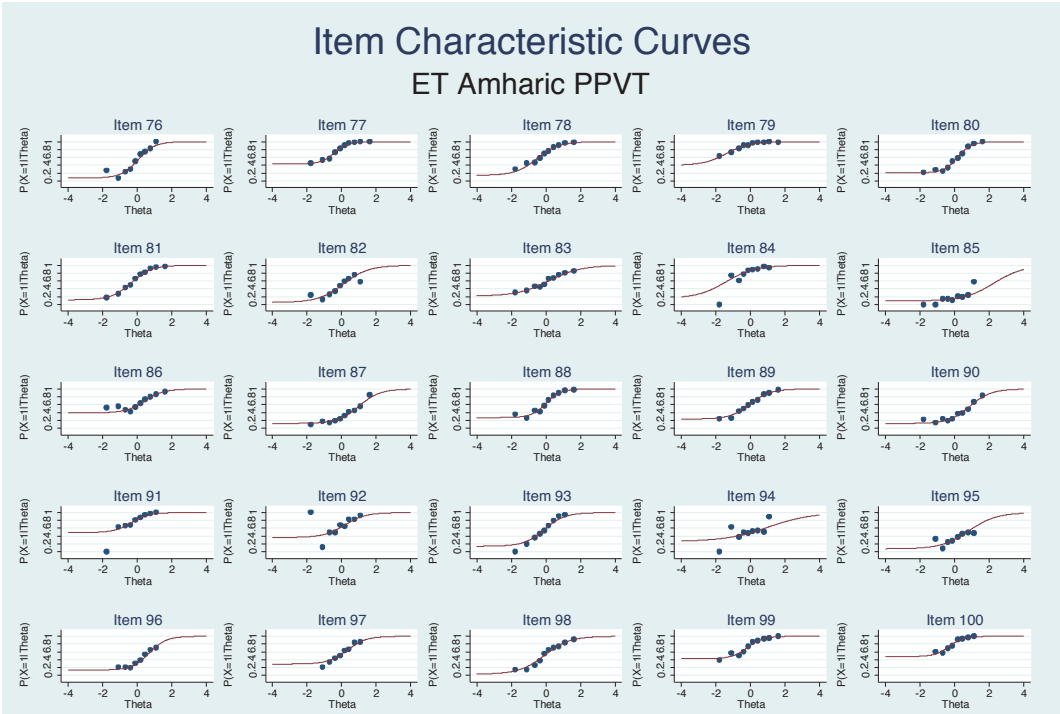
Figure 1. *Item Characteristic Curves for PPVT analysis, Amharic*



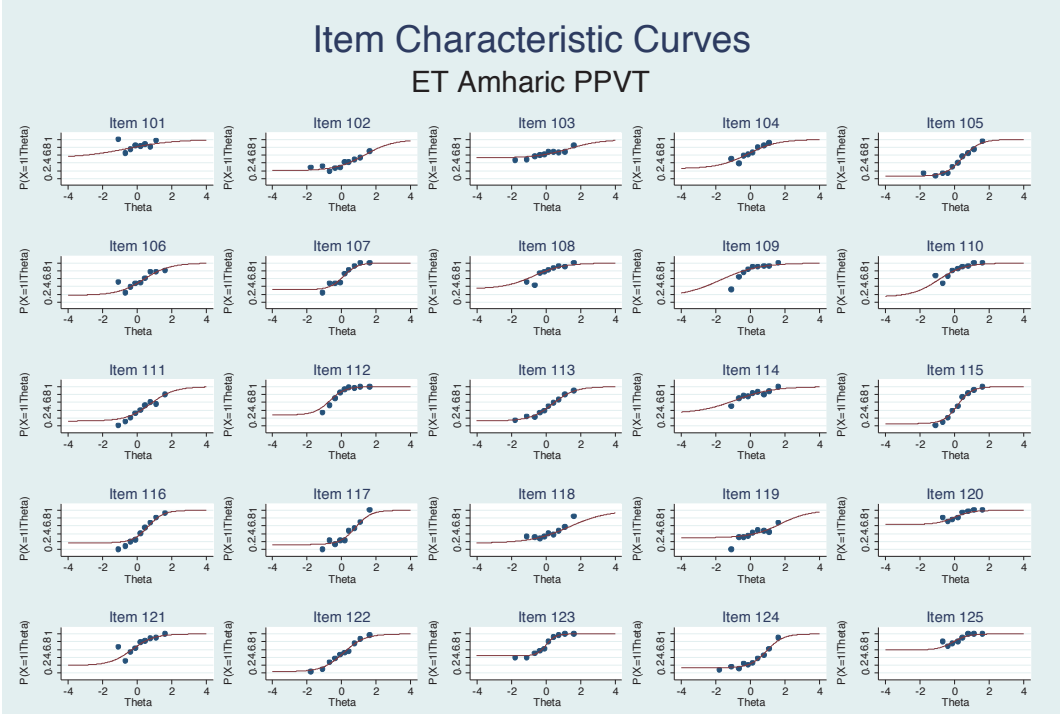


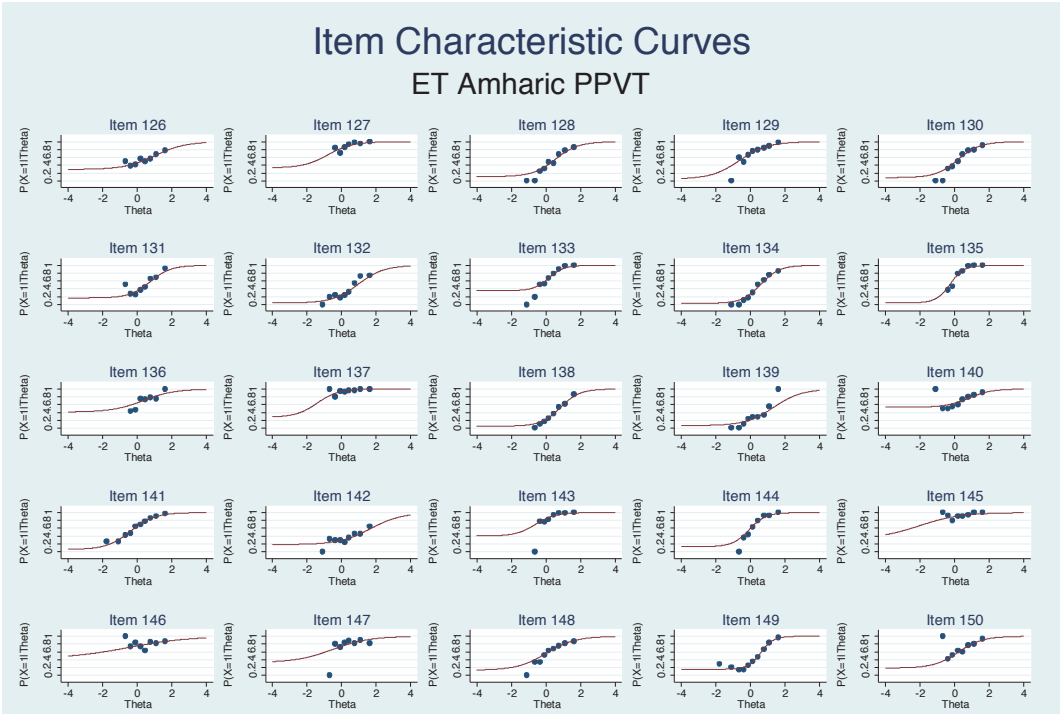


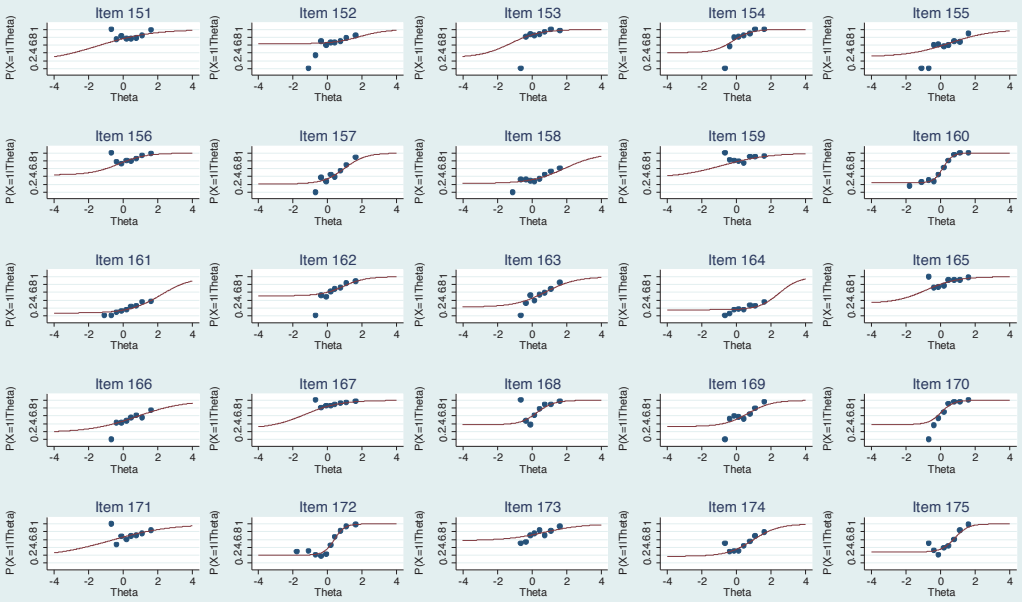
Combining OC and YC from R2 to R5



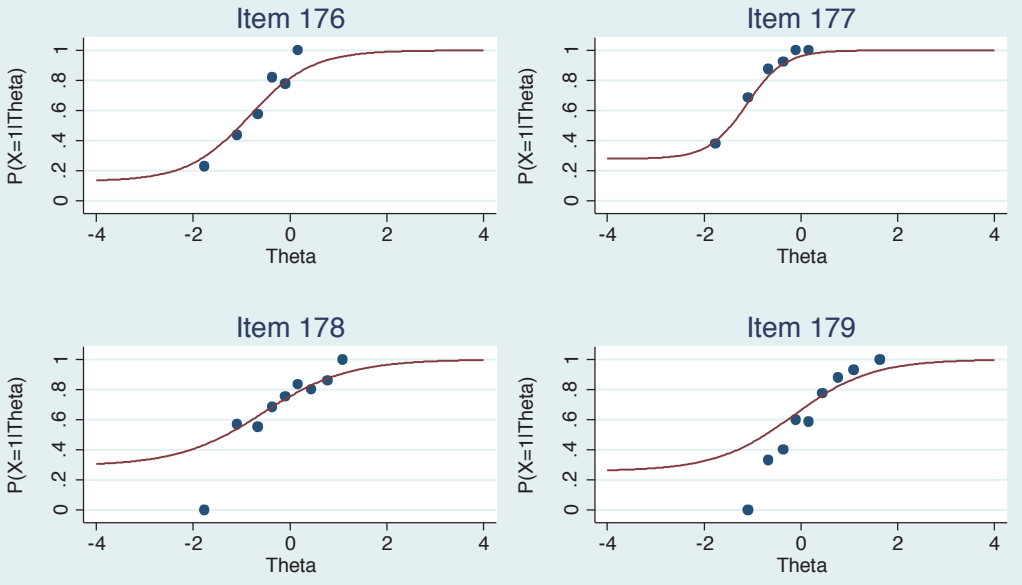
Combining OC and YC from R2 to R5


Item Characteristic Curves
ET Amharic PPVT
Item 101
Item 102
Item 103
Item 104
Item 105
Item 106
Item 107
Item 108
Item 109
Item 110
Item 111
Item 112
Item 113
Item 114
Item 115
Item 116
Item 117
Item 118
Item 119
Item 120
Item 121
Item 122
Item 123
Item 124
Item 125
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5


Item Characteristic Curves
ET Amharic PPVT
Item 126
Item 127
Item 128
Item 129
Item 130
Item 131
Item 132
Item 133
Item 134
Item 135
Item 136
Item 137
Item 138
Item 139
Item 140
Item 141
Item 142
Item 143
Item 144
Item 145
Item 146
Item 147
Item 148
Item 149
Item 150
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

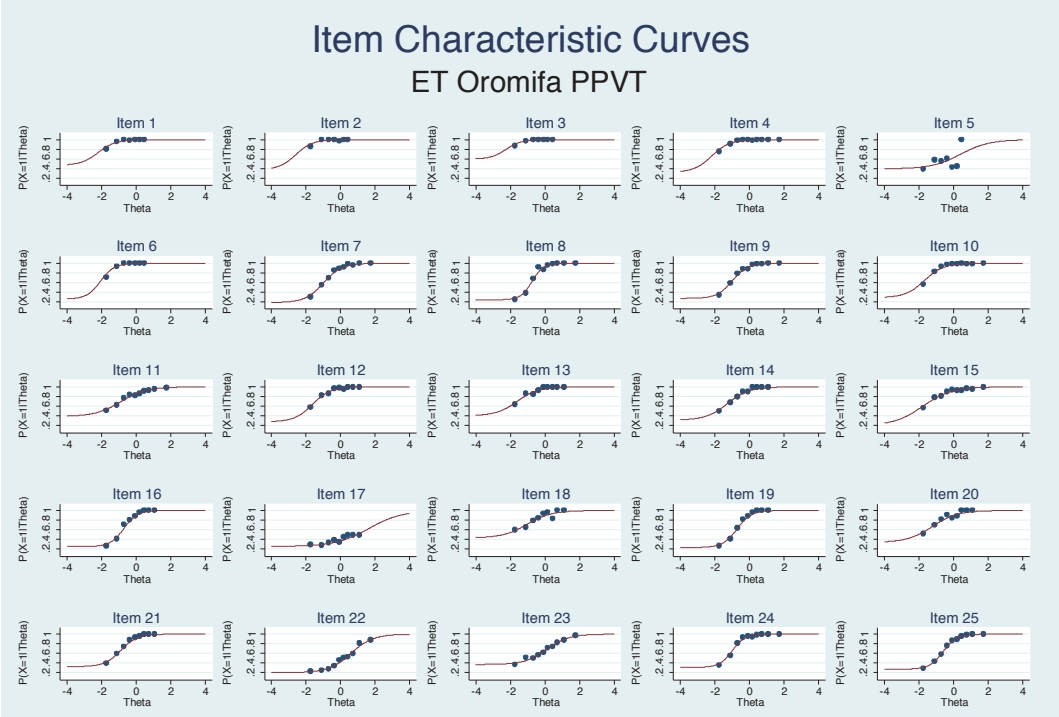


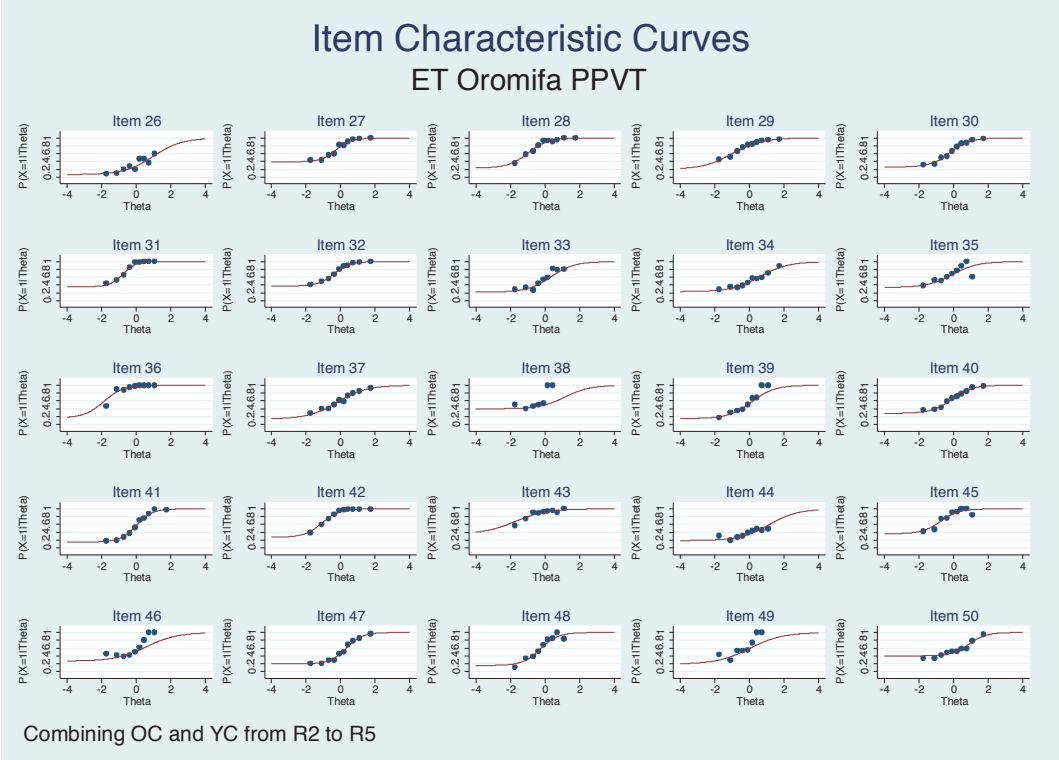
Combining OC and YC from R2 to R5

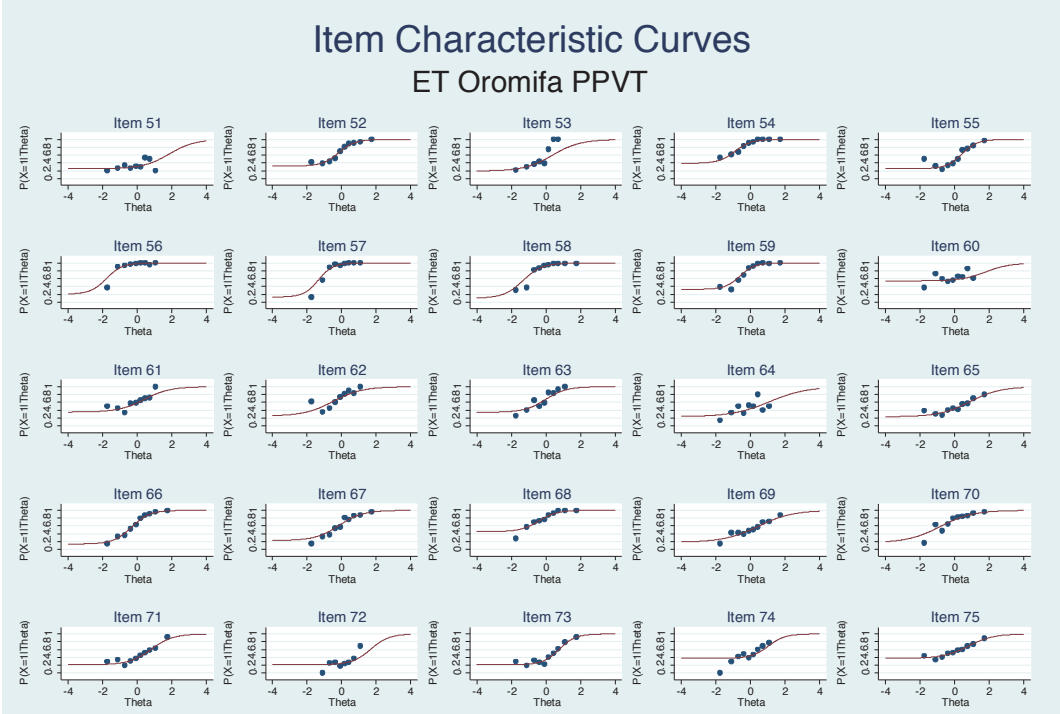


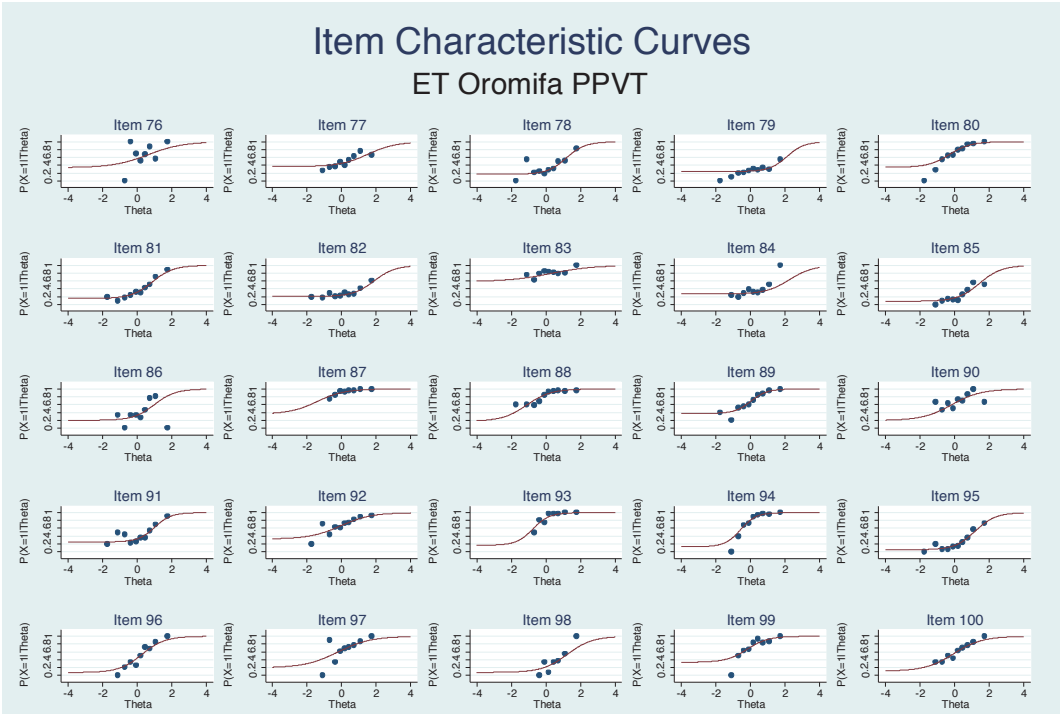
Combining OC and YC from R2 to R5

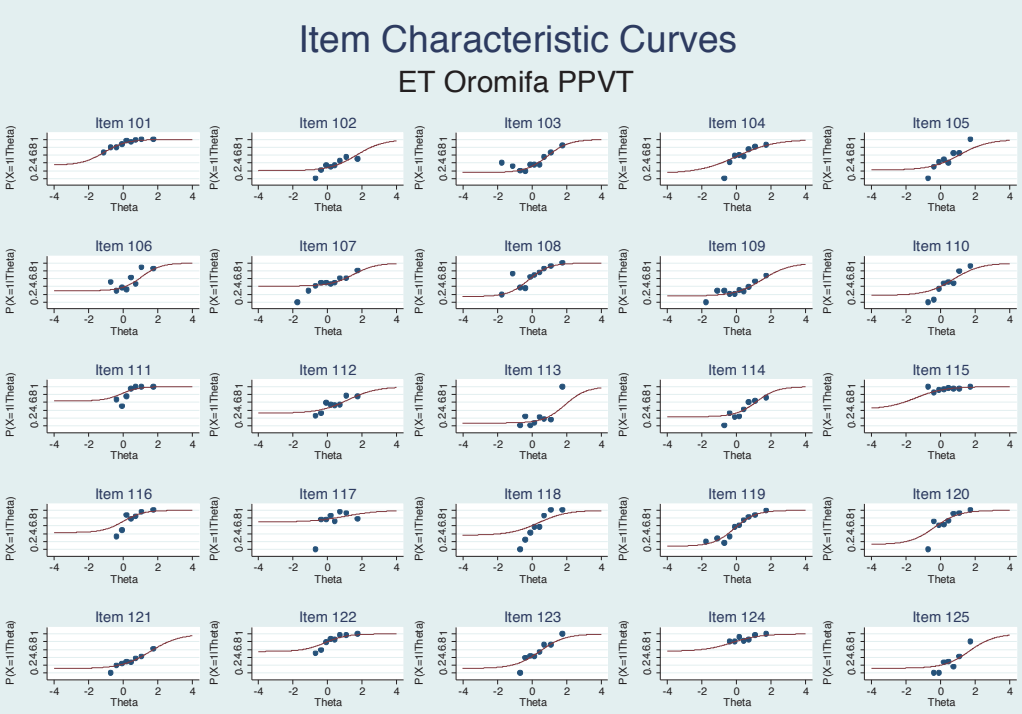
Figure 2. *Item Characteristic Curves for PPVT analysis, Oromifa*



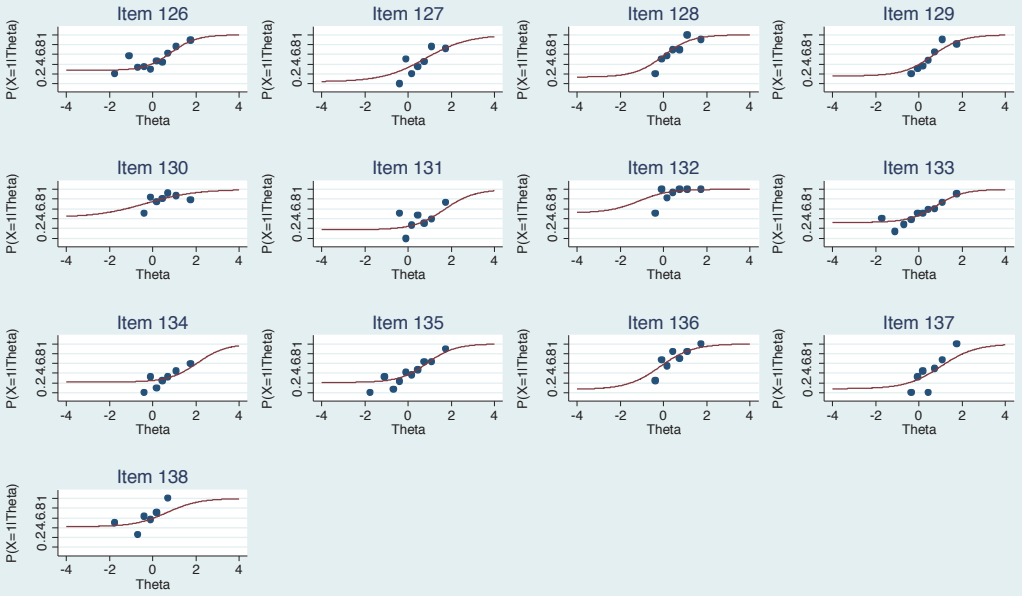



Item Characteristic Curves
ET Oromifa PPVT
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
Item 65
Item 66
Item 67
Item 68
Item 69
Item 70
Item 71
Item 72
Item 73
Item 74
Item 75
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5


Item Characteristic Curves
ET Oromifa PPVT
Item 76
Item 77
Item 78
Item 79
Item 80
Item 81
Item 82
Item 83
Item 84
Item 85
Item 86
Item 87
Item 88
Item 89
Item 90
Item 91
Item 92
Item 93
Item 94
Item 95
Item 96
Item 97
Item 98
Item 99
Item 100
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

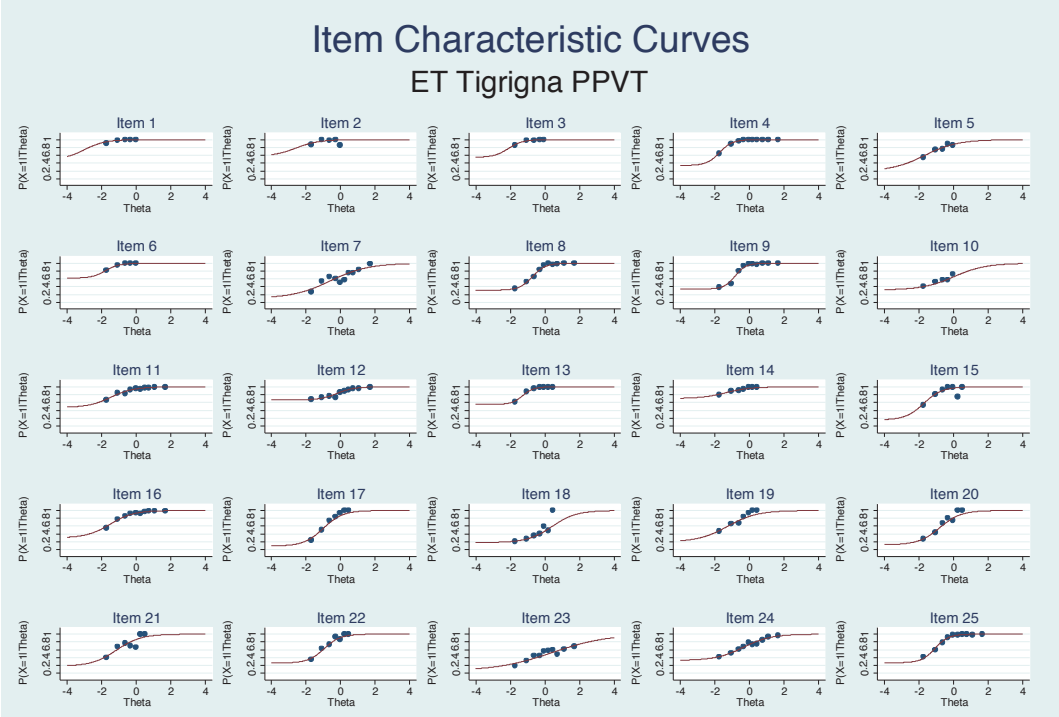


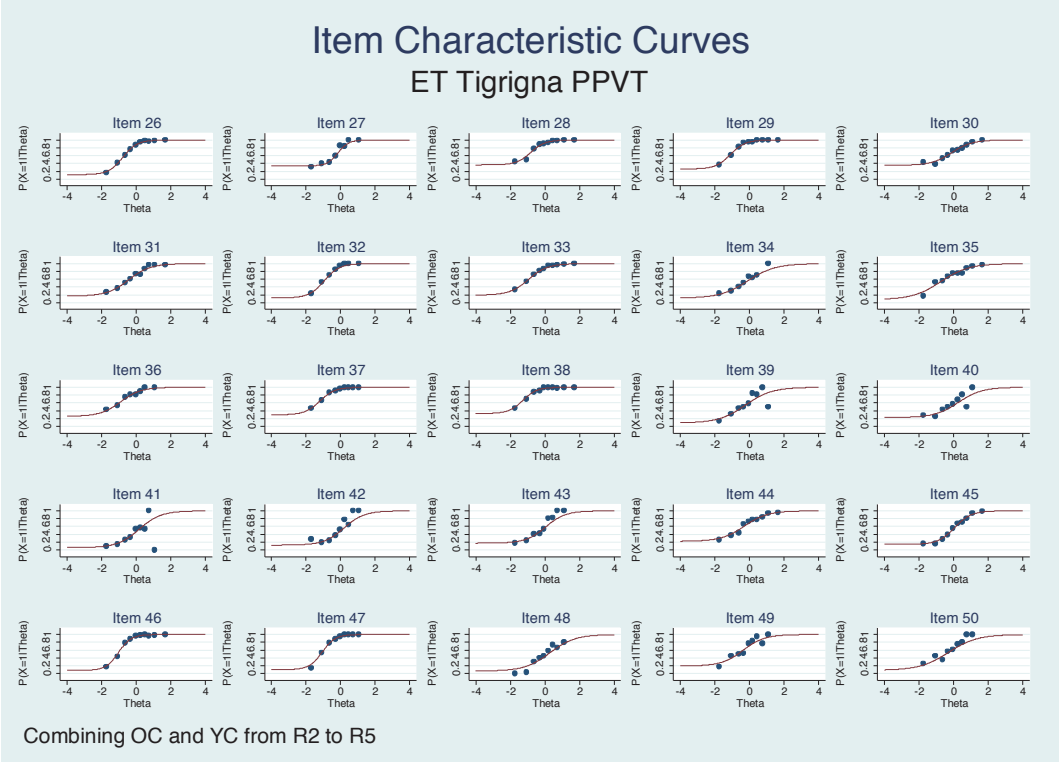
Combining OC and YC from R2 to R5

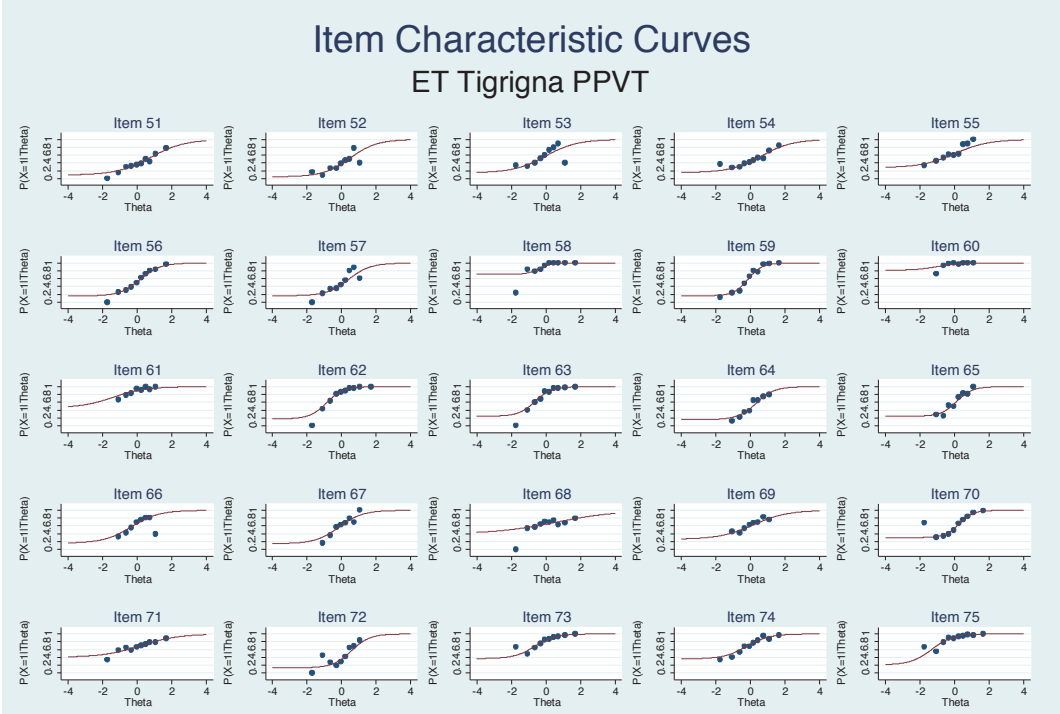


Combining OC and YC from R2 to R5

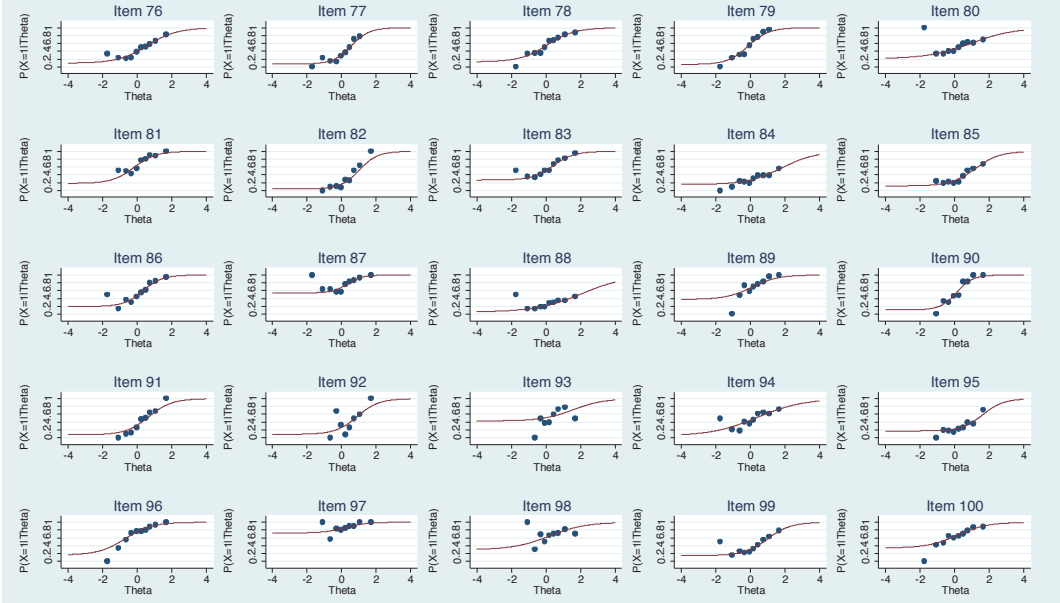
Figure 3. *Item Characteristic Curves for PPVT analysis, Tigrigna*



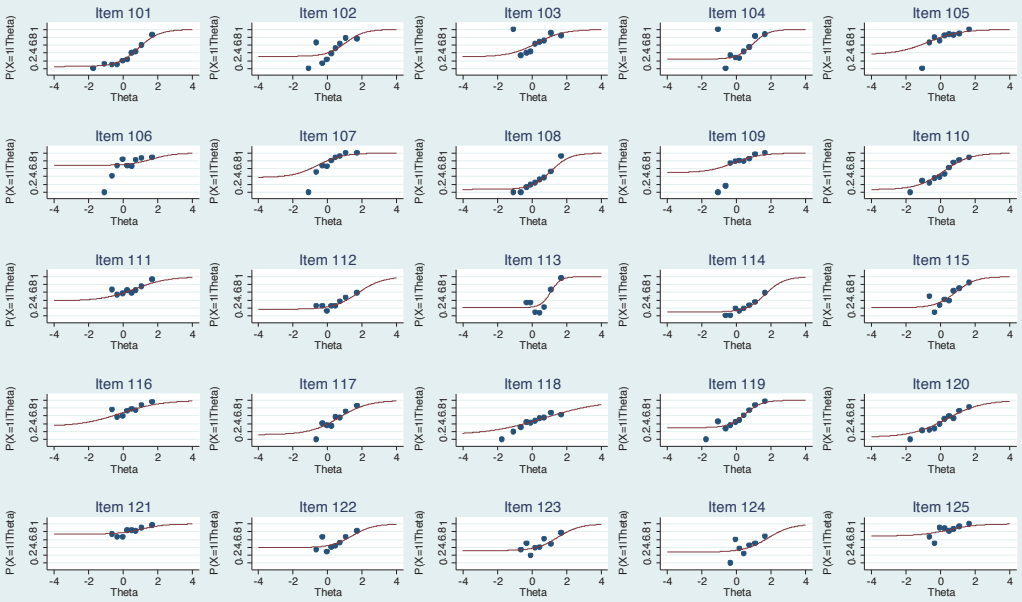




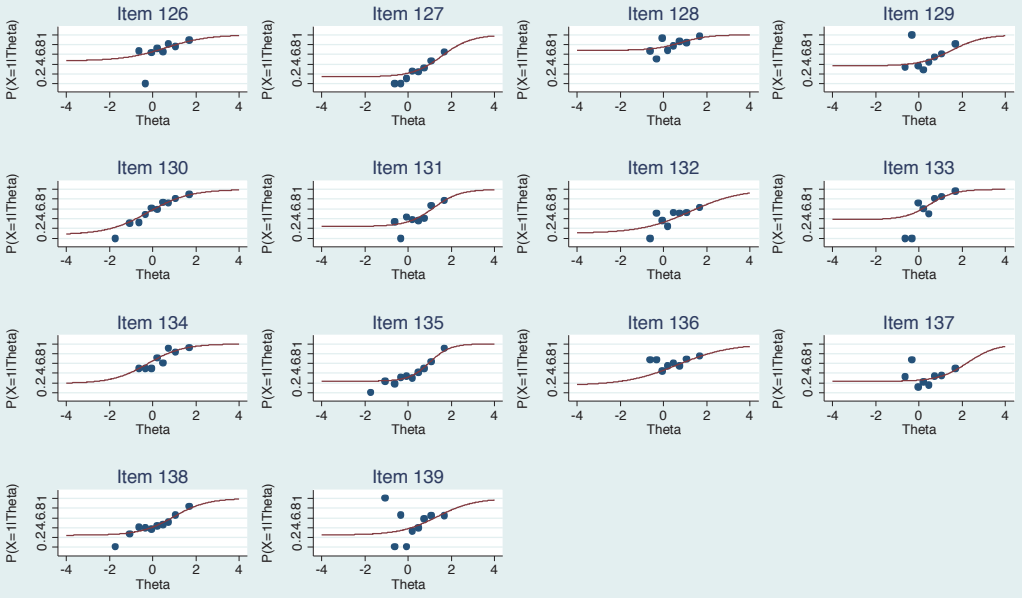
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

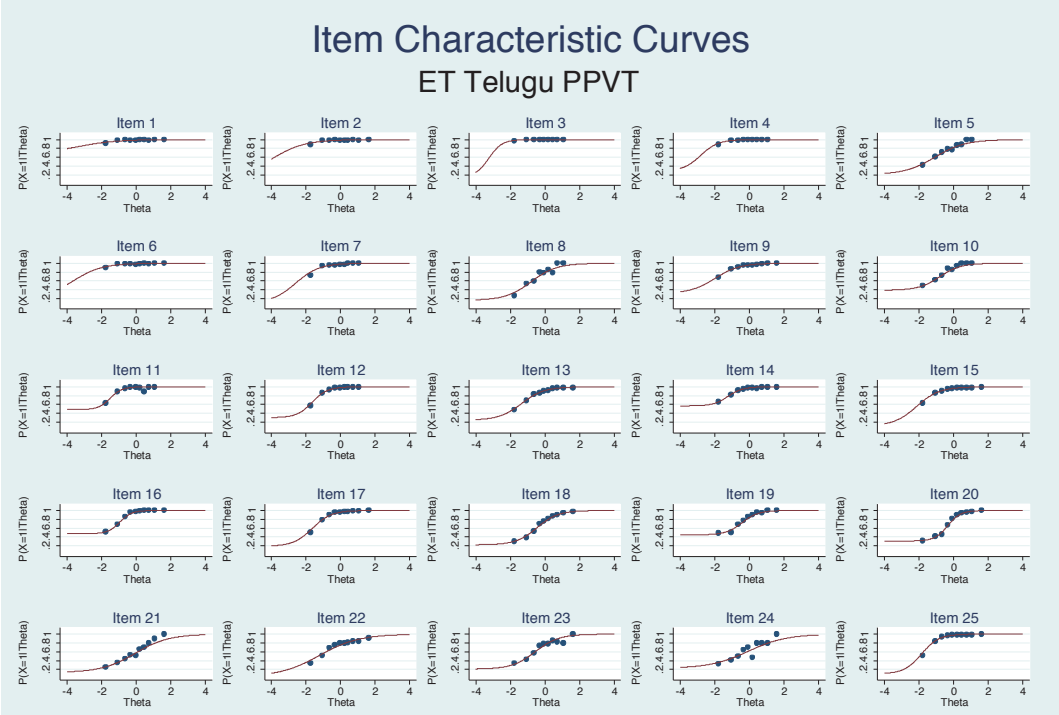


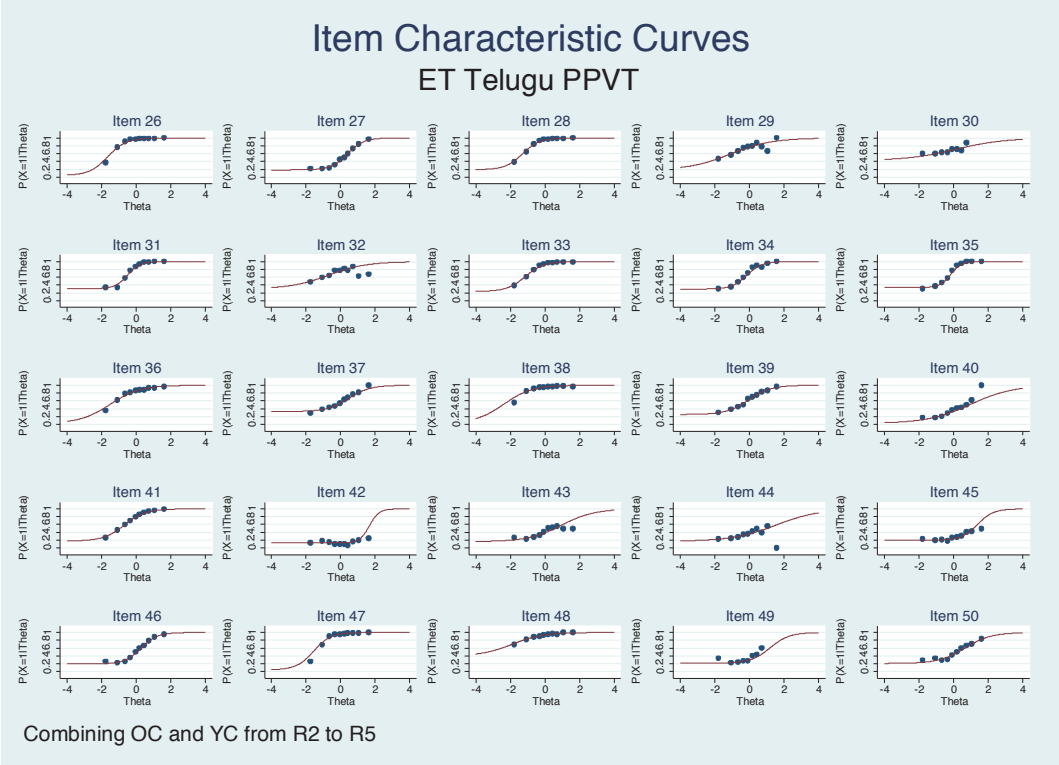
Combining OC and YC from R2 to R5

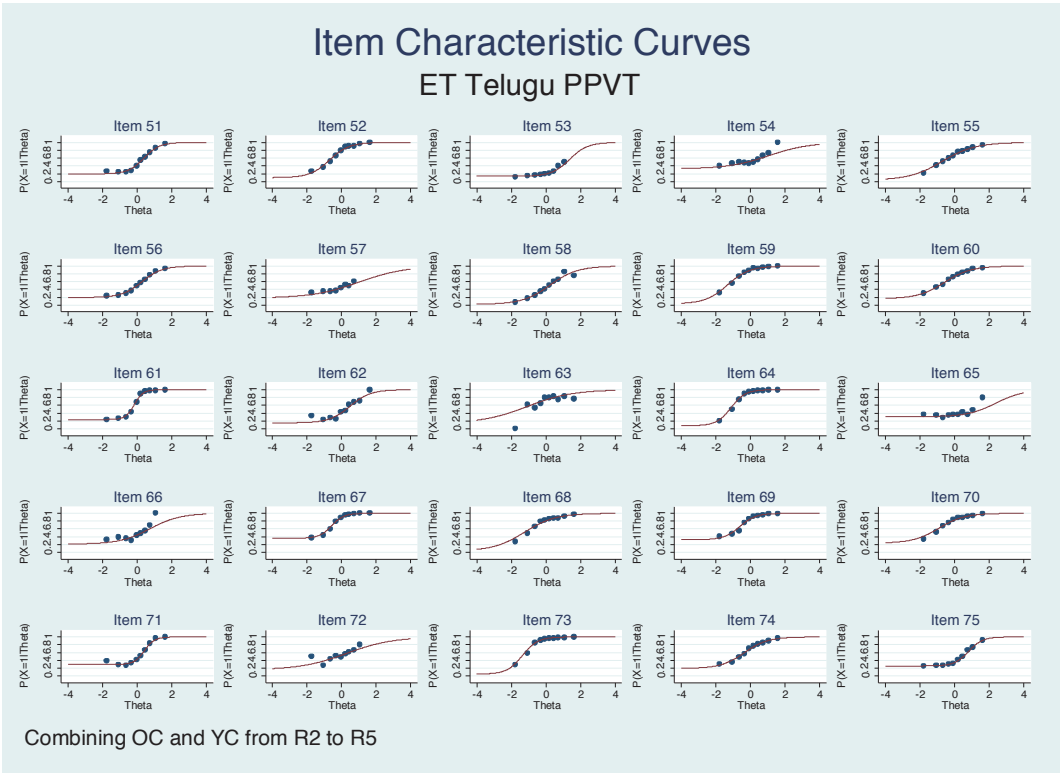


Combining OC and YC from R2 to R5

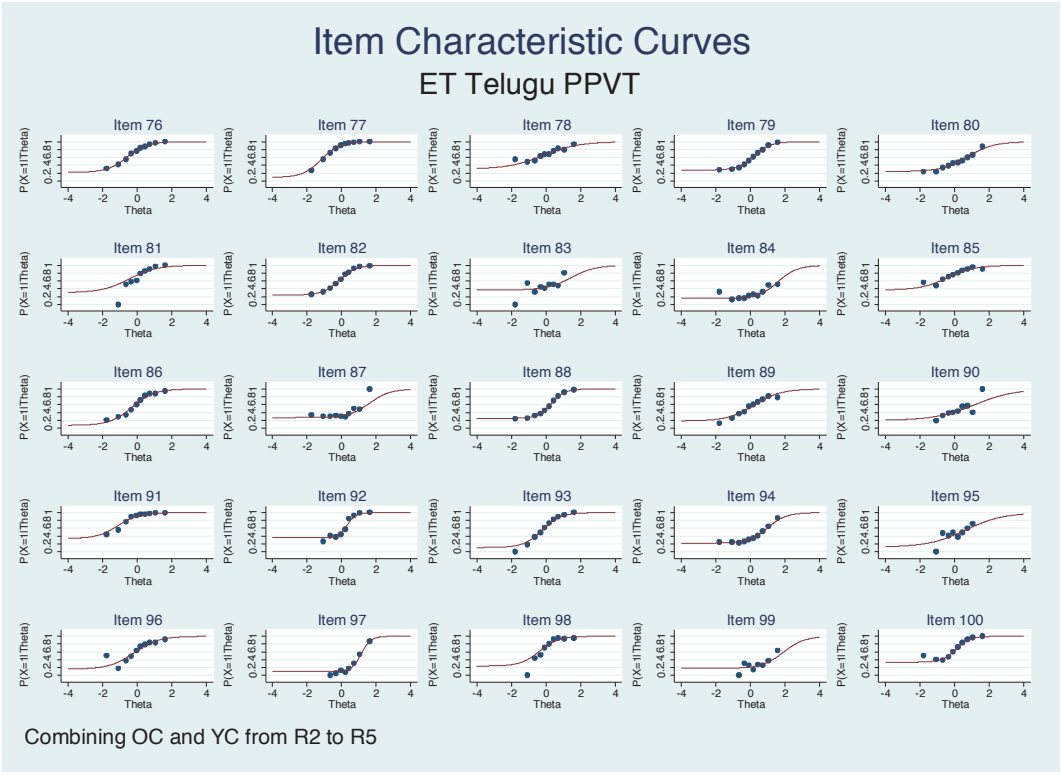
Figure 4. *Item Characteristic Curves for PPVT analysis, Telugu*



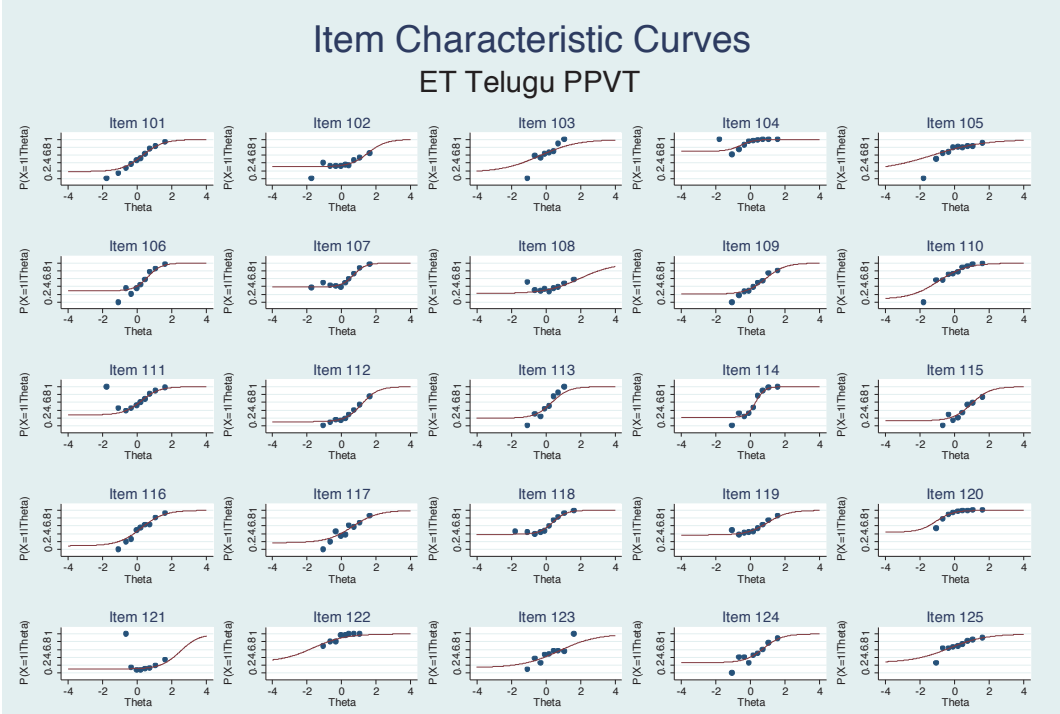


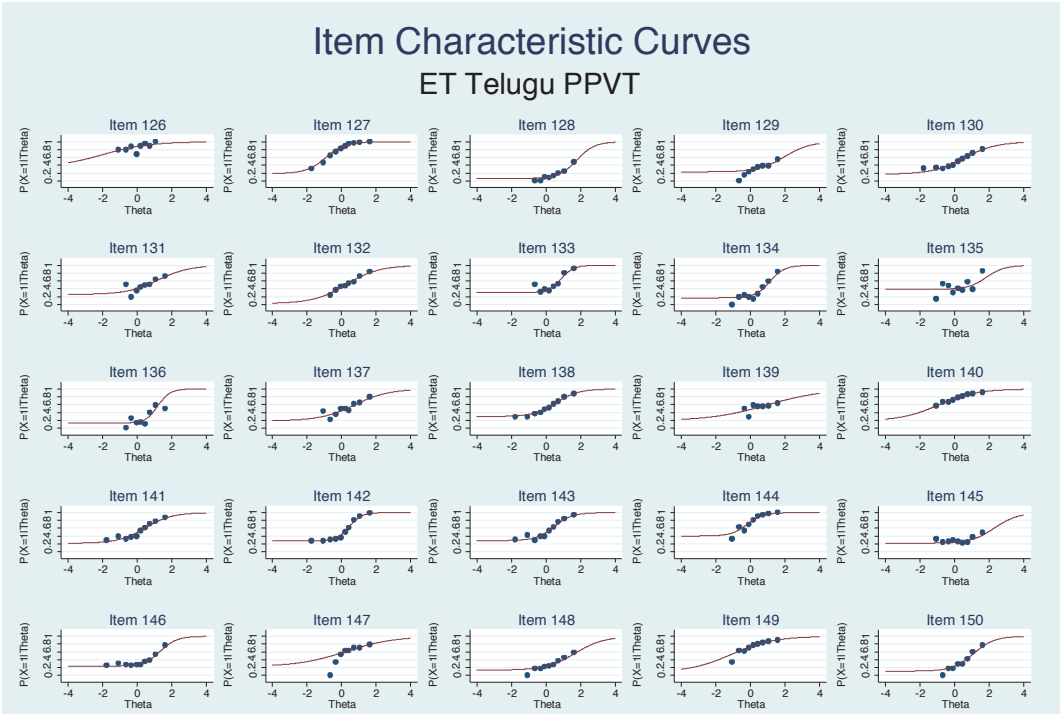


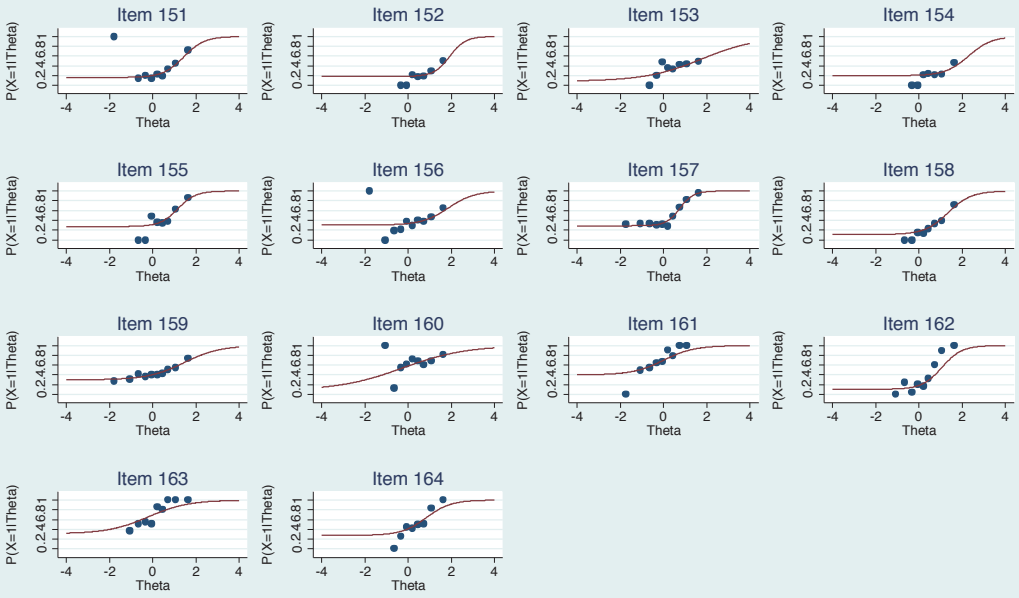
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

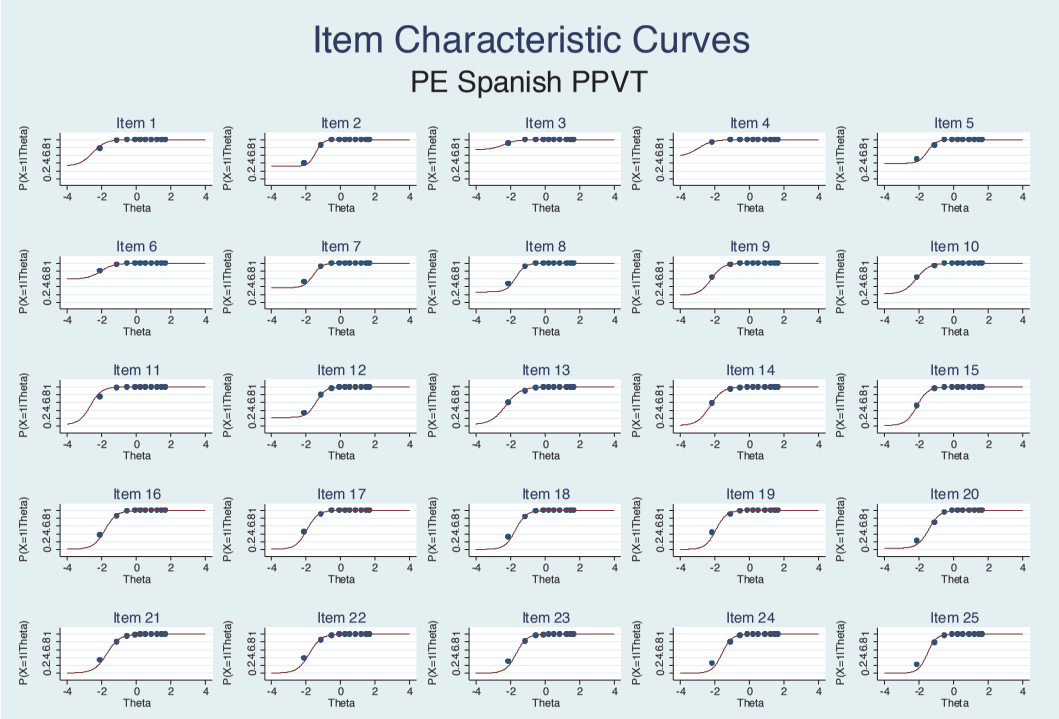

Item Characteristic Curves
ET Telugu PPVT
Item 101
Item 102
Item 103
Item 104
Item 105
Item 106
Item 107
Item 108
Item 109
Item 110
Item 111
Item 112
Item 113
Item 114
Item 115
Item 116
Item 117
Item 118
Item 119
Item 120
Item 121
Item 122
Item 123
Item 124
Item 125
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

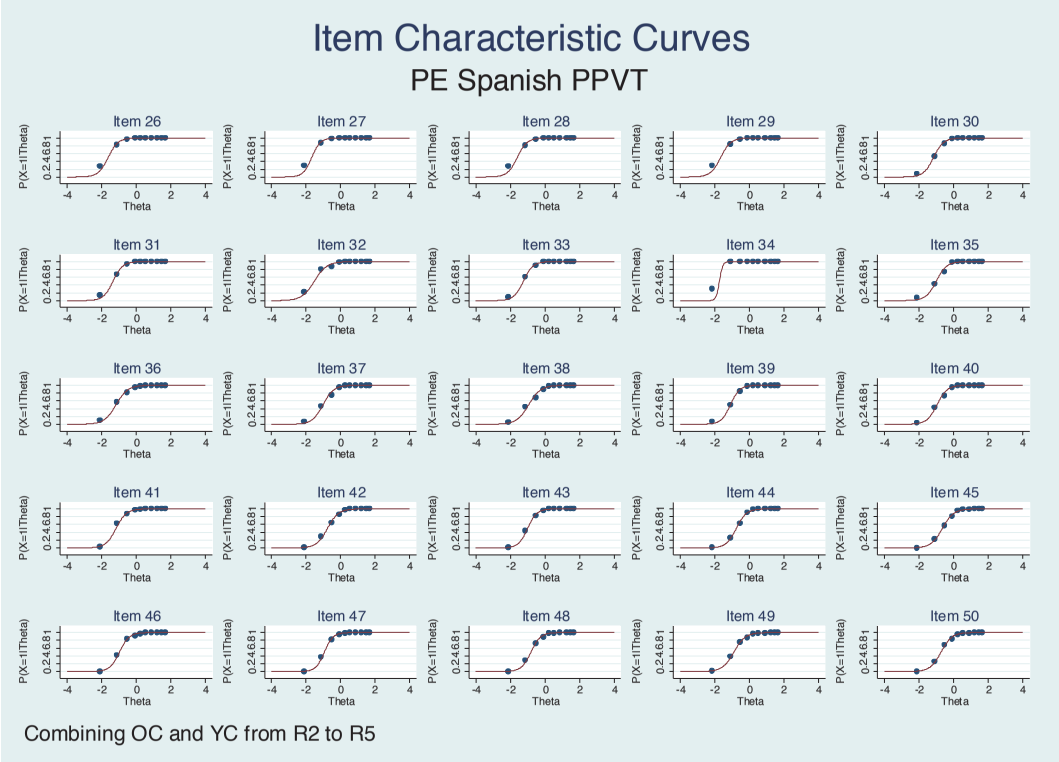

Item Characteristic Curves
ET Telugu PPVT
Item 126
Item 127
Item 128
Item 129
Item 130
Item 131
Item 132
Item 133
Item 134
Item 135
Item 136
Item 137
Item 138
Item 139
Item 140
Item 141
Item 142
Item 143
Item 144
Item 145
Item 146
Item 147
Item 148
Item 149
Item 150
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

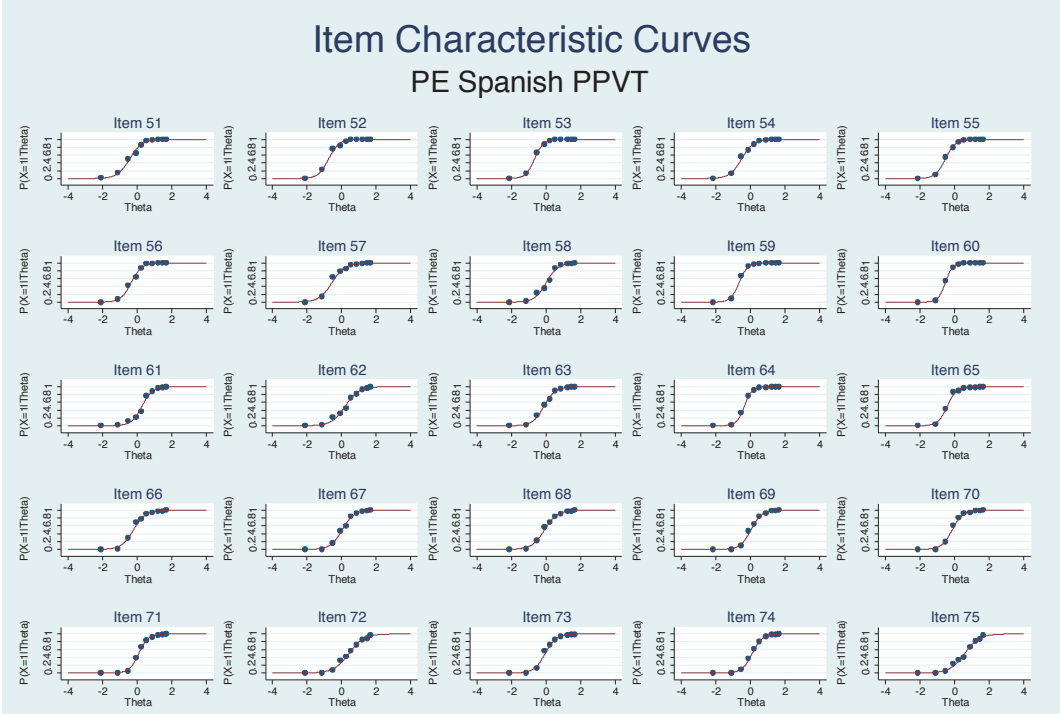


Combining OC and YC from R2 to R5

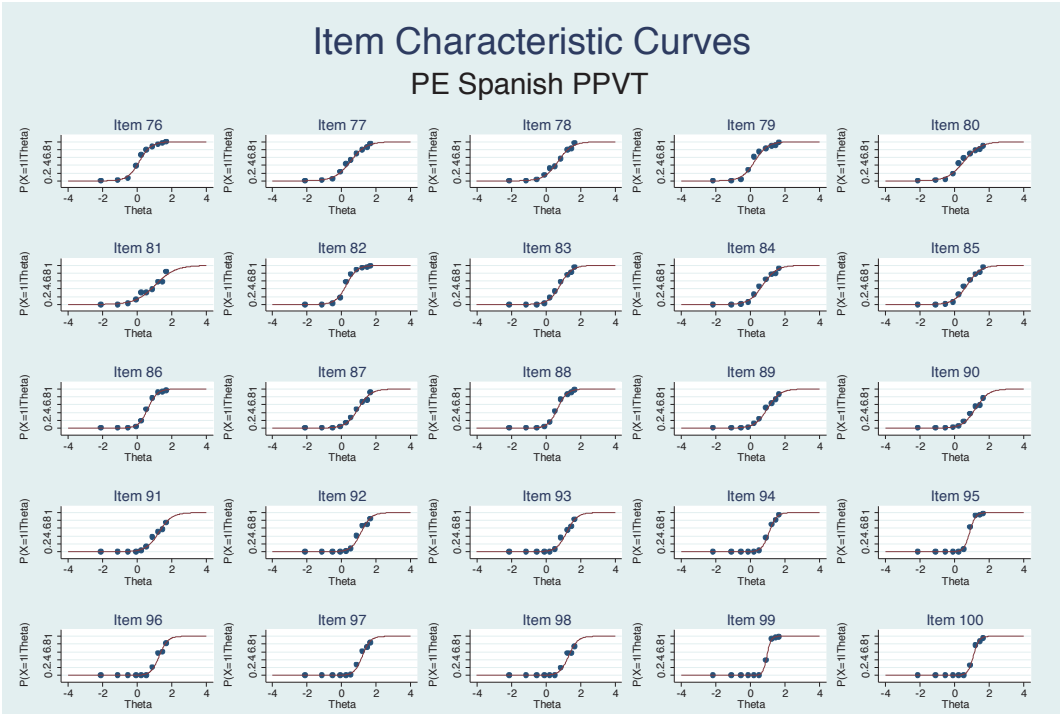
Figure 5. *Item Characteristic Curves for PPVT analysis, Spanish*







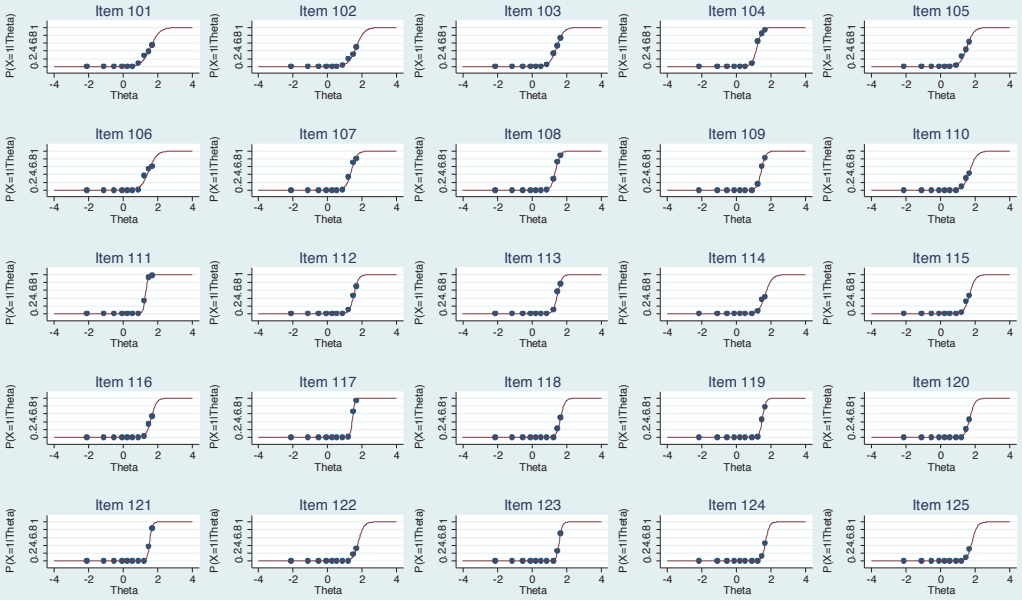
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

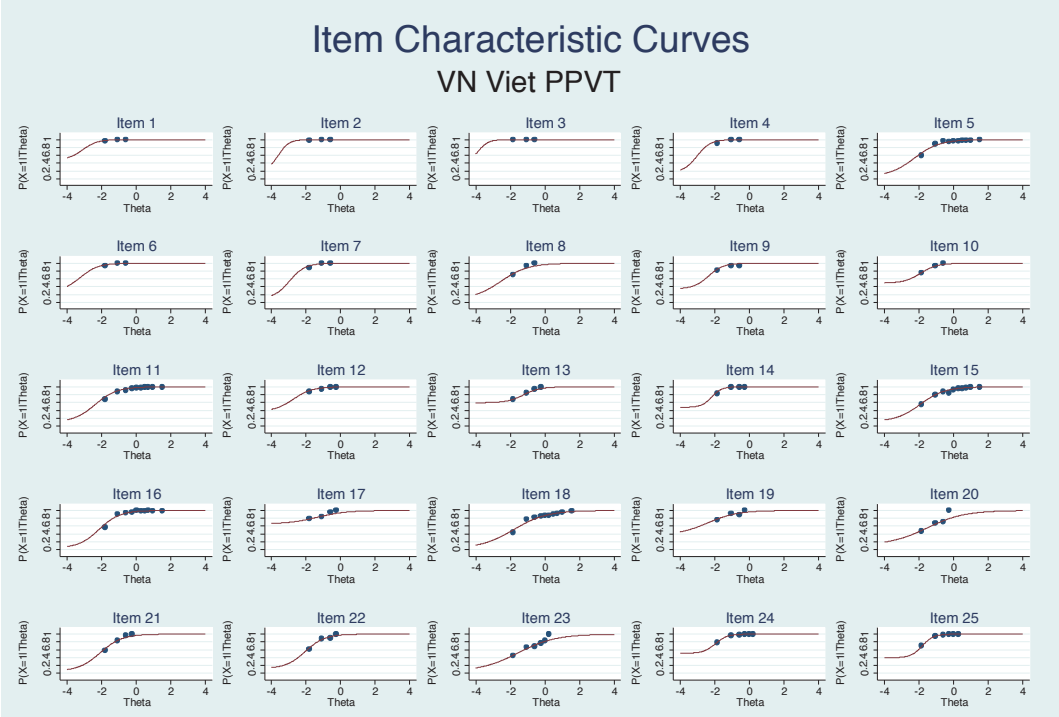
Item Characteristic Curves

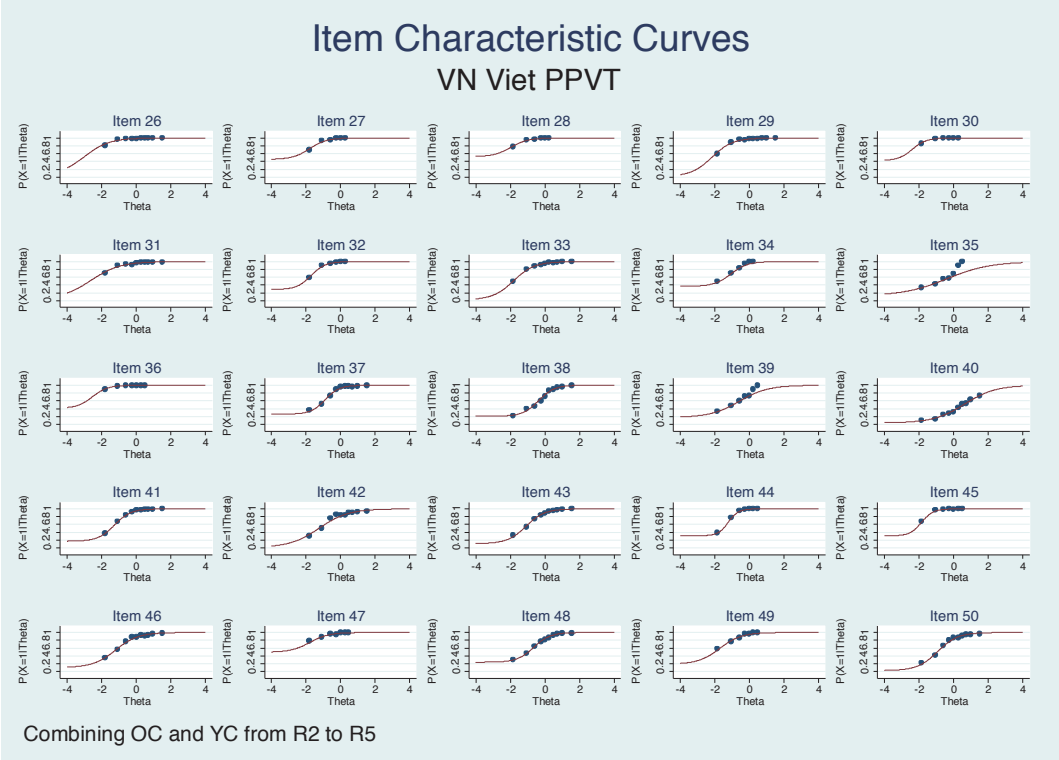
PE Spanish PPVT

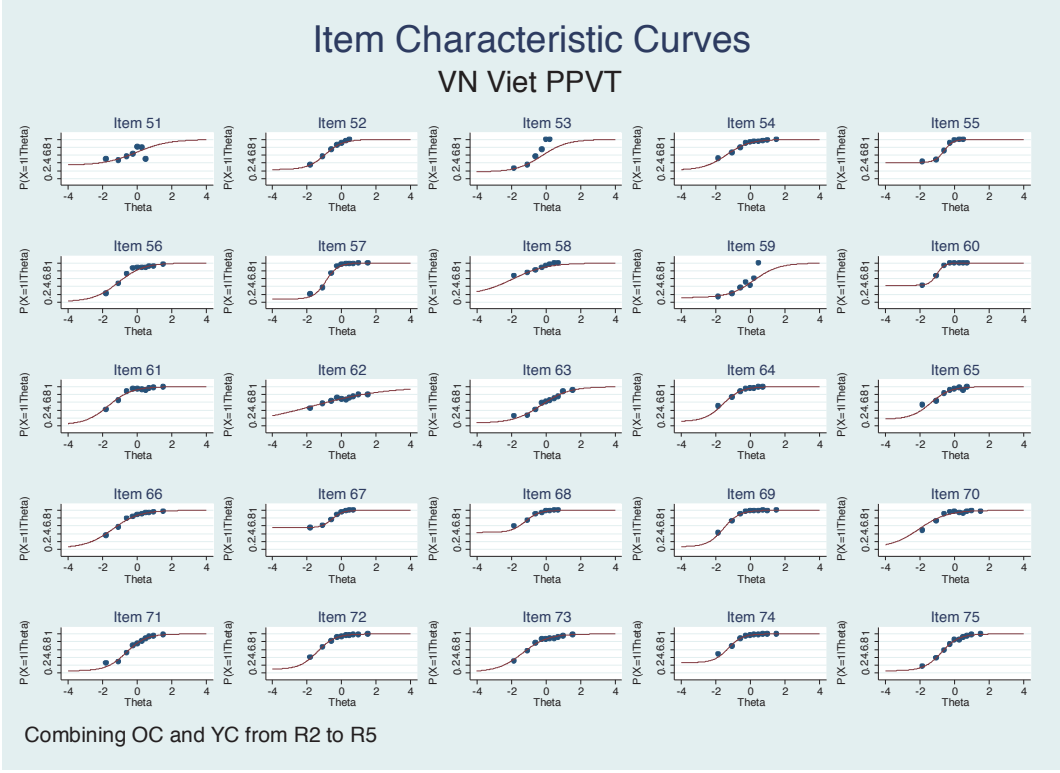


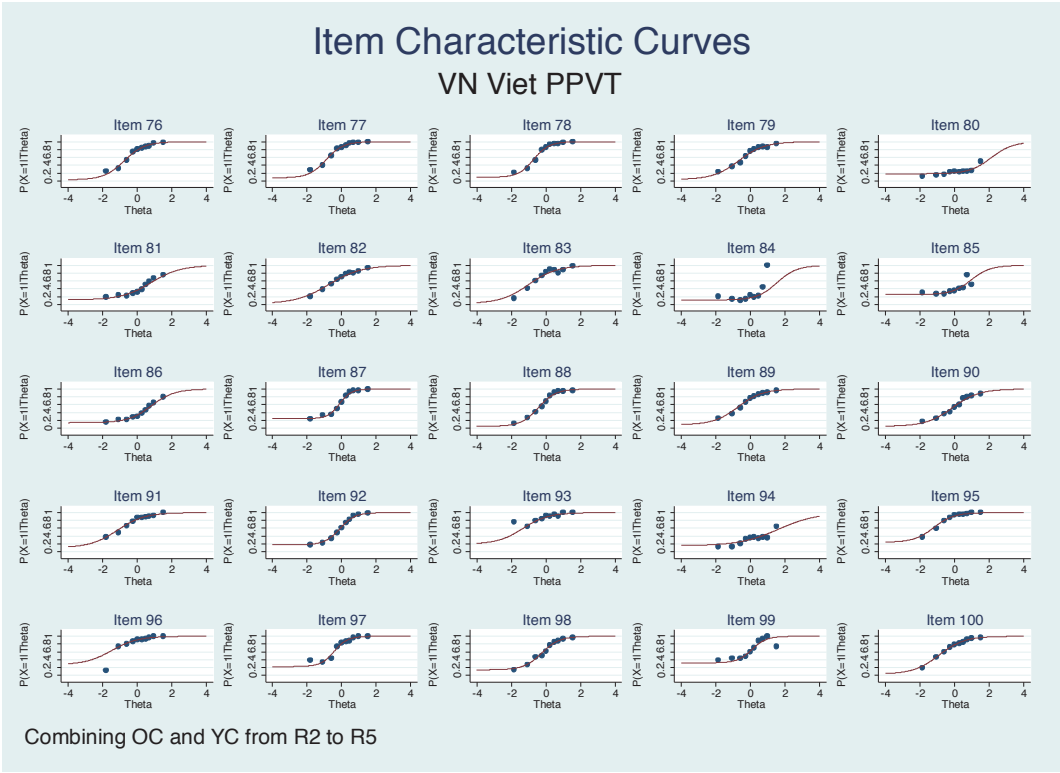
Combining OC and YC from R2 to R5

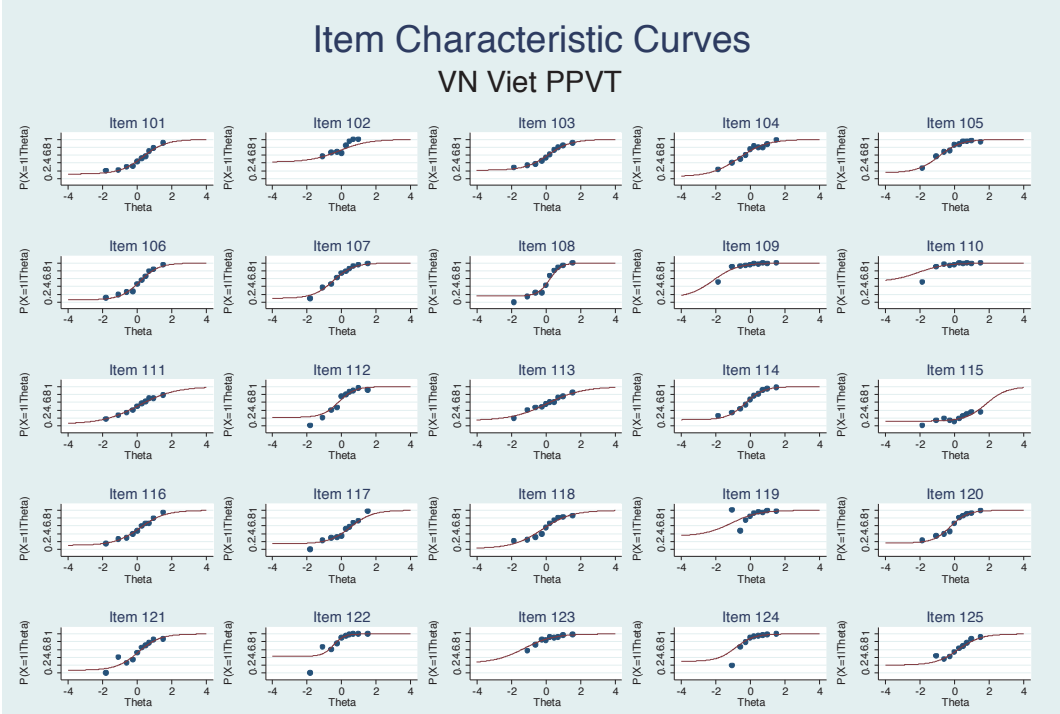
Figure 6. *Item Characteristic Curves for PPVT analysis, Vietnamese*



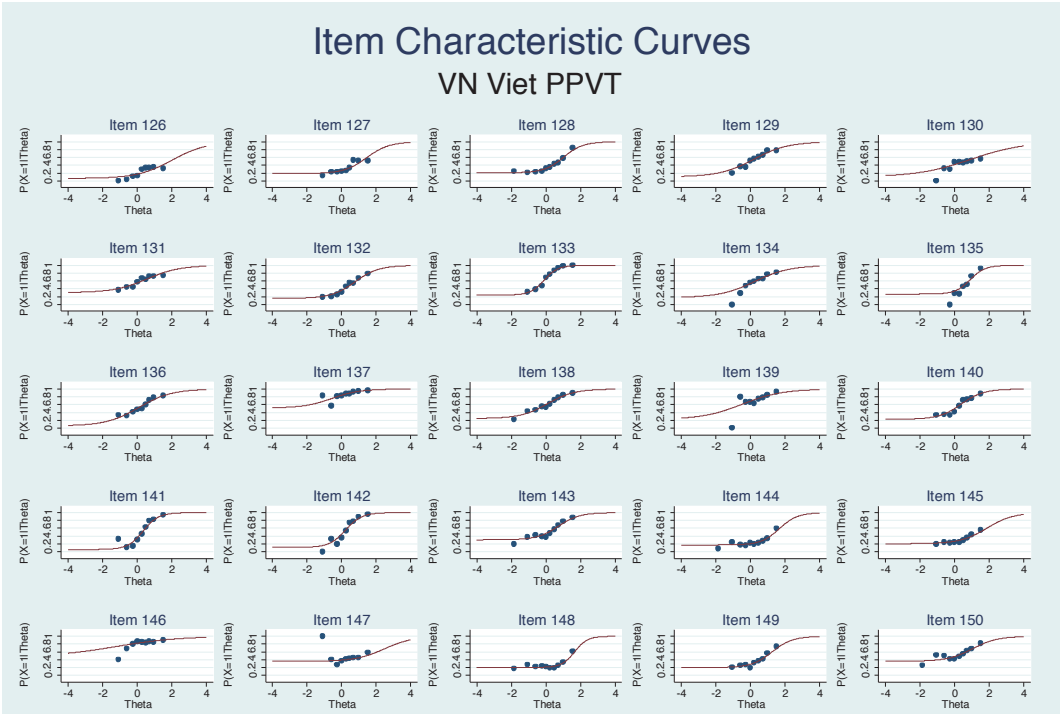



Item Characteristic Curves
VN Viet PPVT
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
Item 65
Item 66
Item 67
Item 68
Item 69
Item 70
Item 71
Item 72
Item 73
Item 74
Item 75
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5


Item Characteristic Curves
VN Viet PPVT
Item 76
Item 77
Item 78
Item 79
Item 80
Item 81
Item 82
Item 83
Item 84
Item 85
Item 86
Item 87
Item 88
Item 89
Item 90
Item 91
Item 92
Item 93
Item 94
Item 95
Item 96
Item 97
Item 98
Item 99
Item 100
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

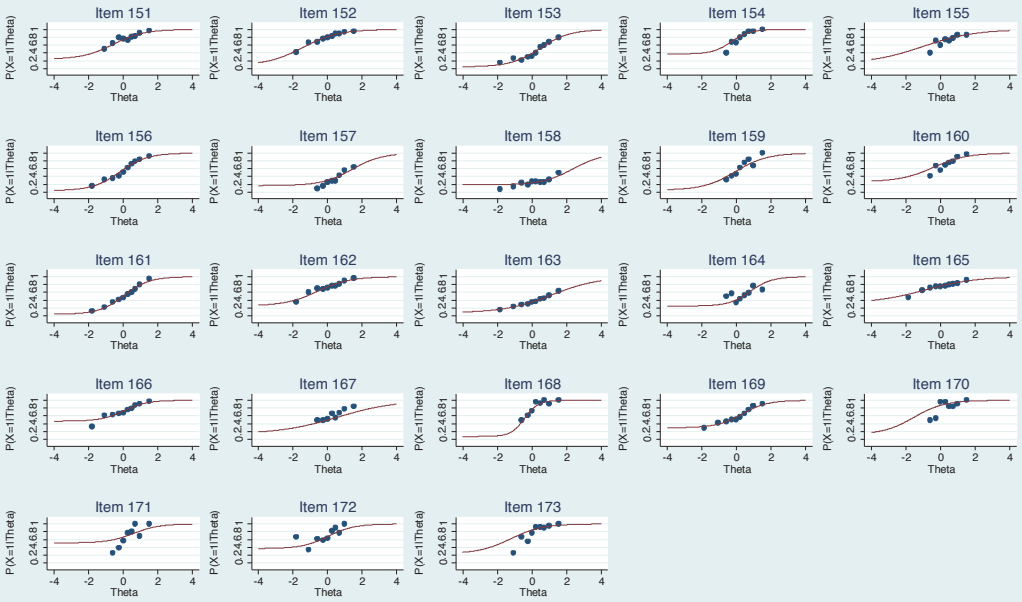
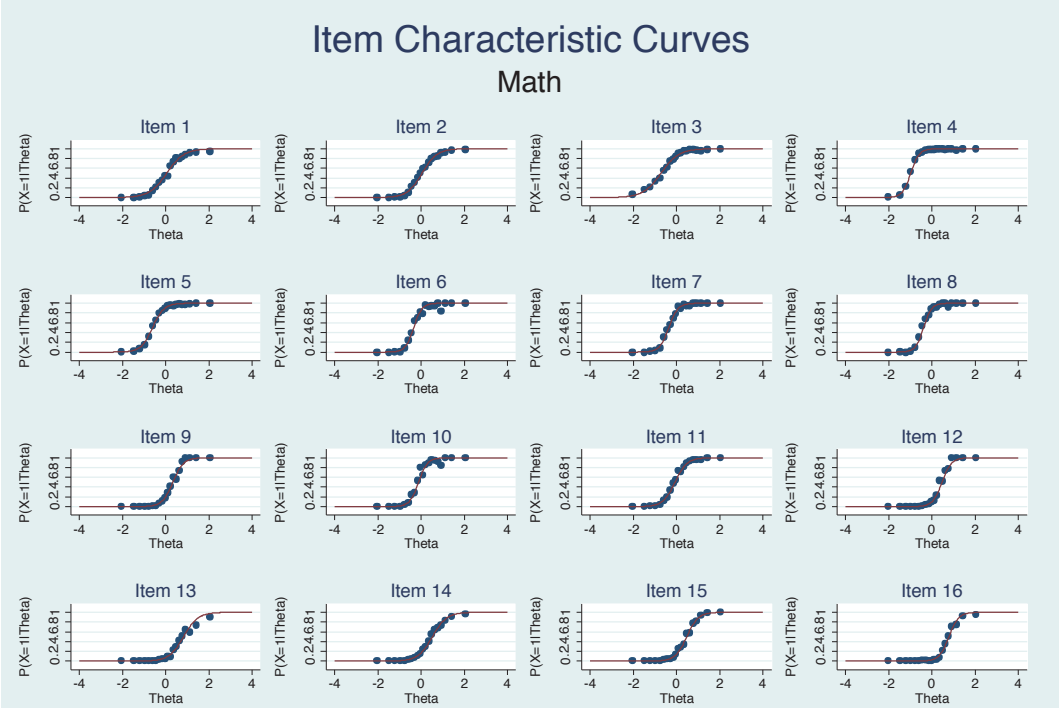
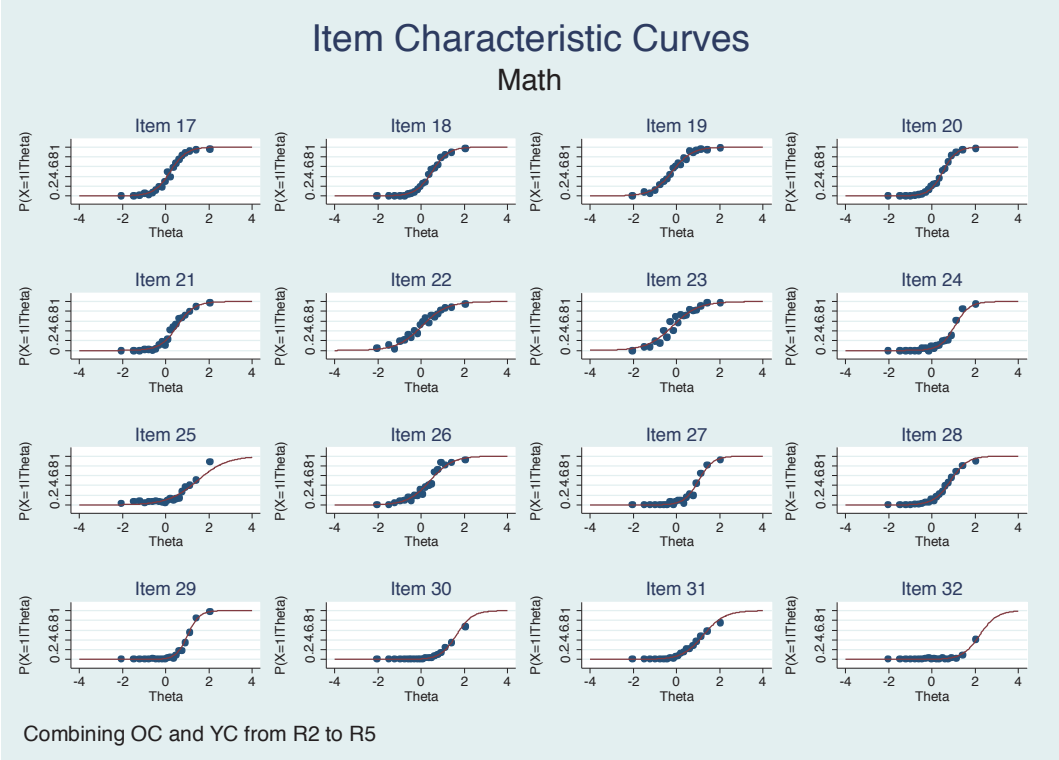
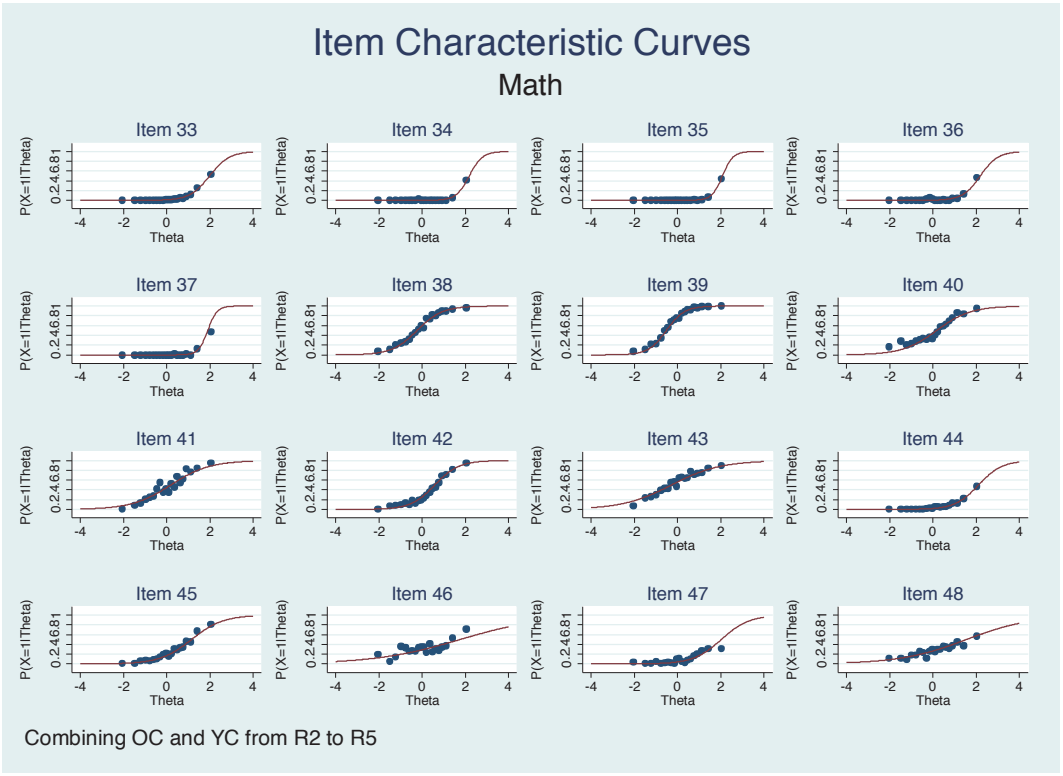

Item Characteristic Curves
VN Viet PPVT
Item 151
Item 152
Item 153
Item 154
Item 155
Item 156
Item 157
Item 158
Item 159
Item 160
Item 161
Item 162
Item 163
Item 164
Item 165
Item 166
Item 167
Item 168
Item 169
Item 170
Item 171
Item 172
Item 173
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

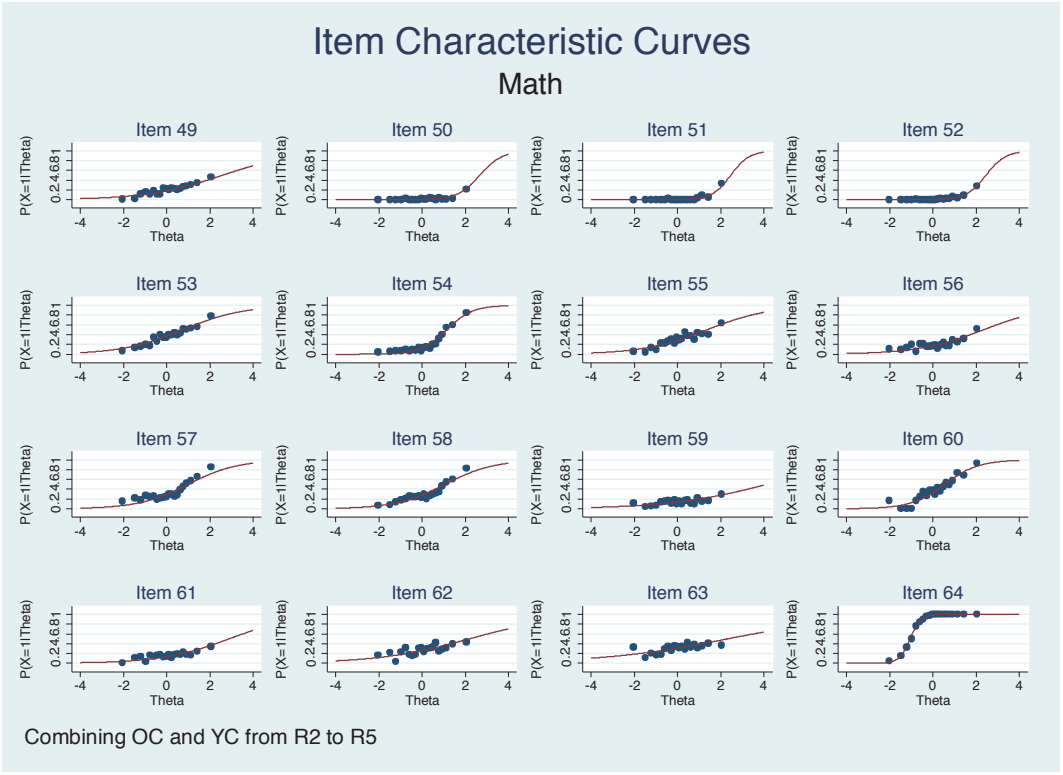
Figure 7. *Item Characteristic Curves for maths analysis, Ethiopia*



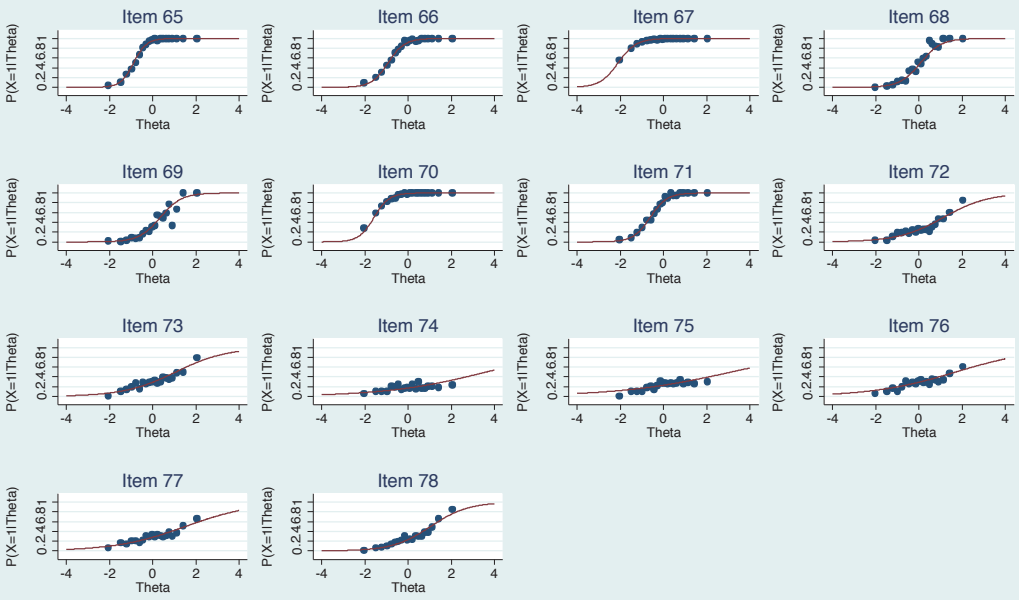




Combining OC and YC from R2 to R5

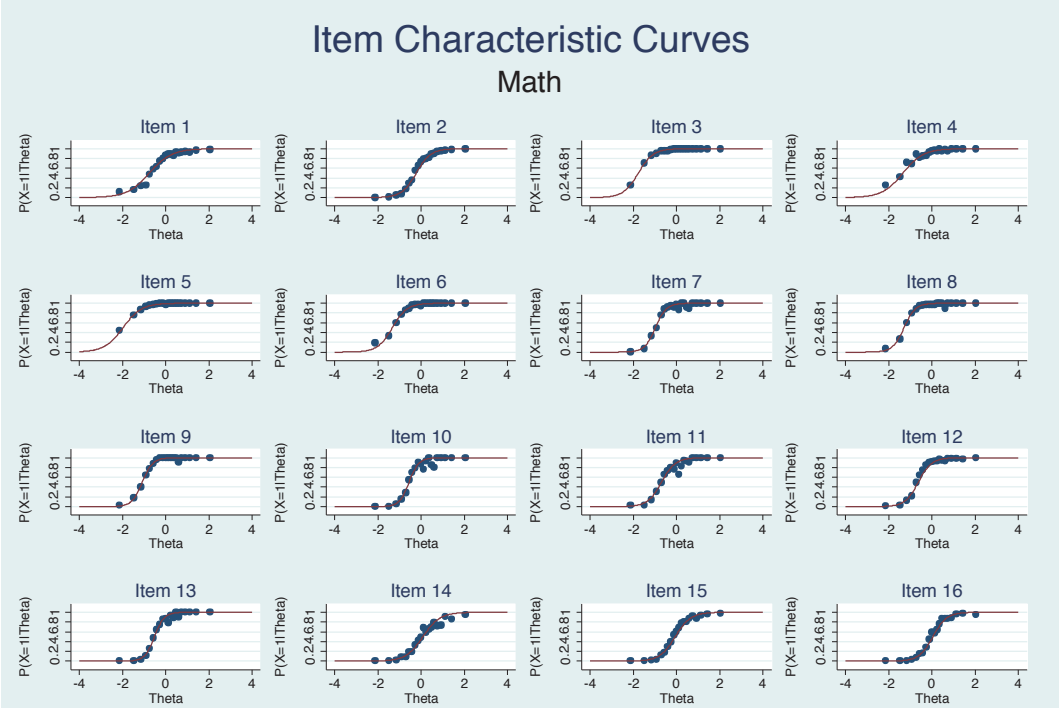


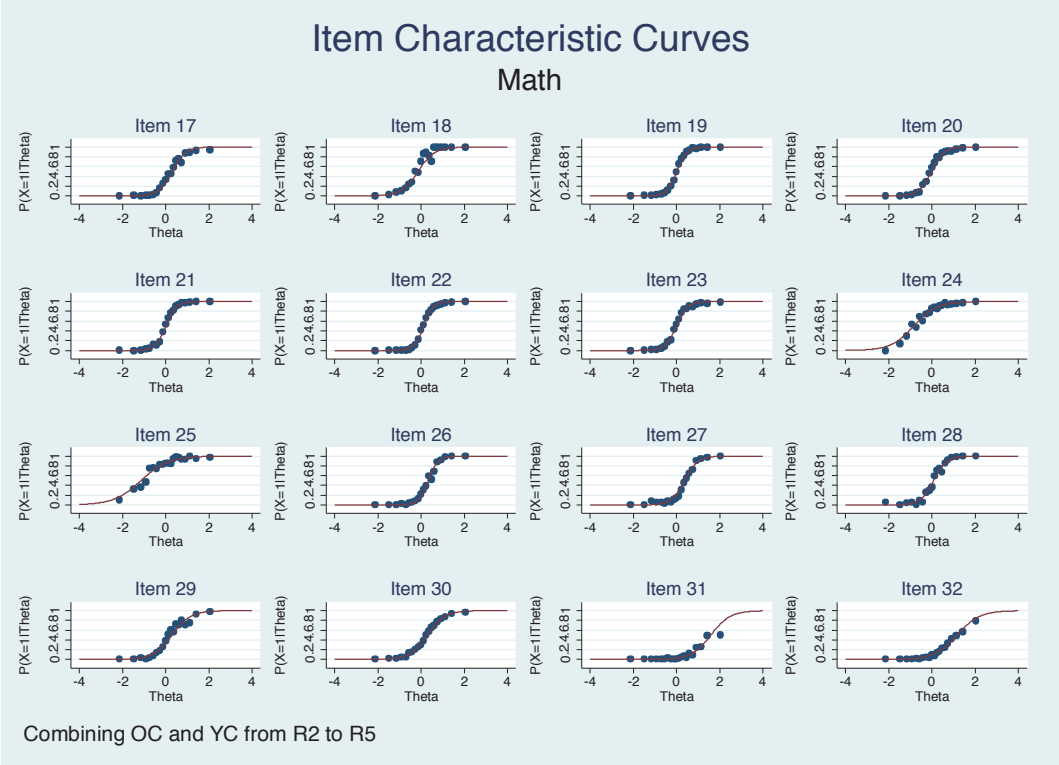
Combining OC and YC from R2 to R5

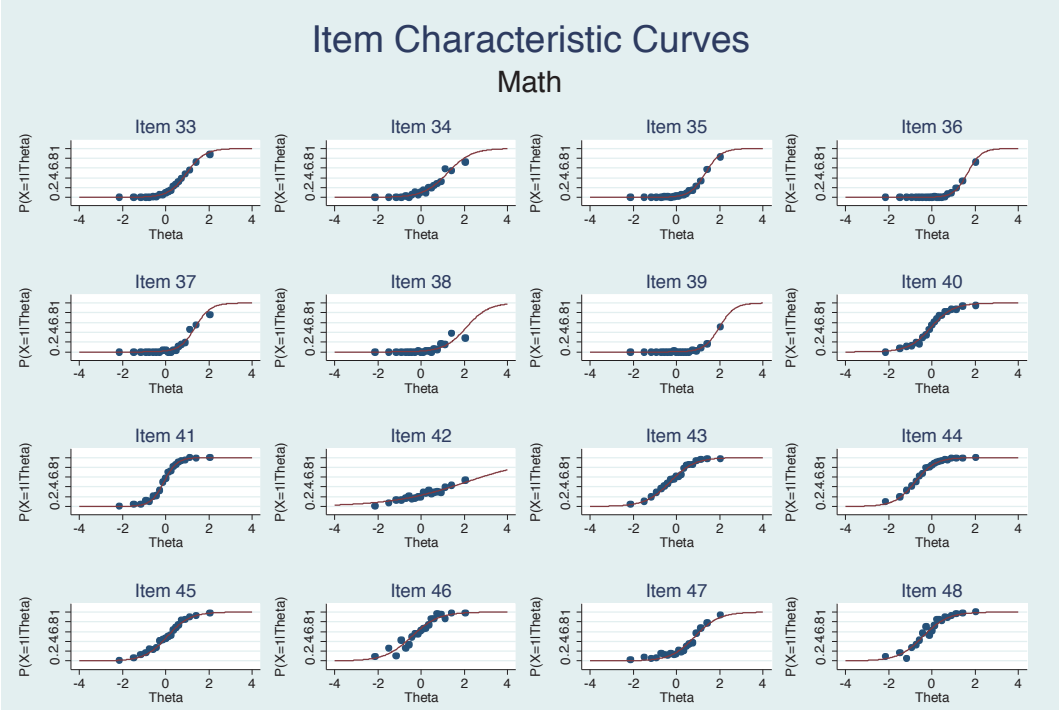


Combining OC and YC from R2 to R5

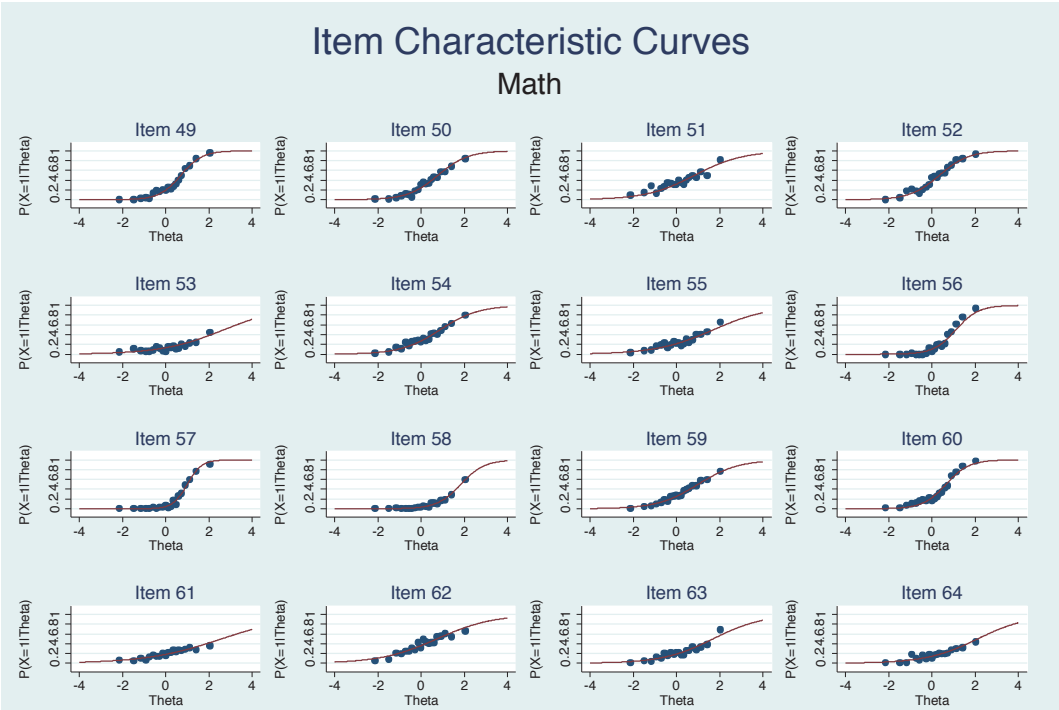
Figure 8. *Item Characteristic Curves for maths analysis, India*



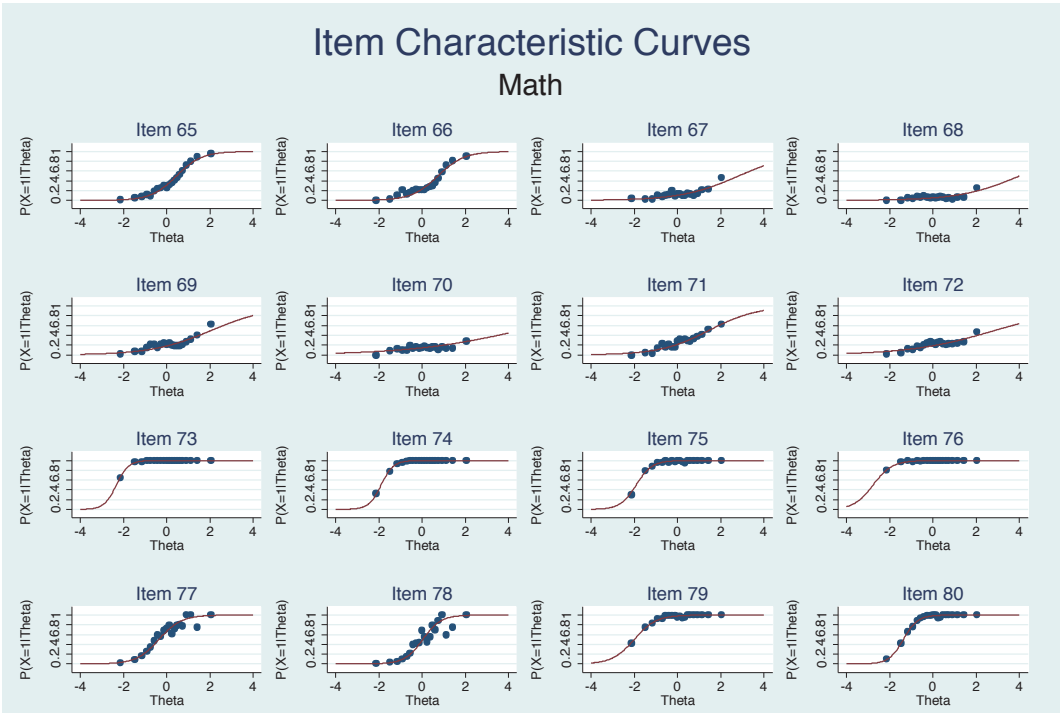




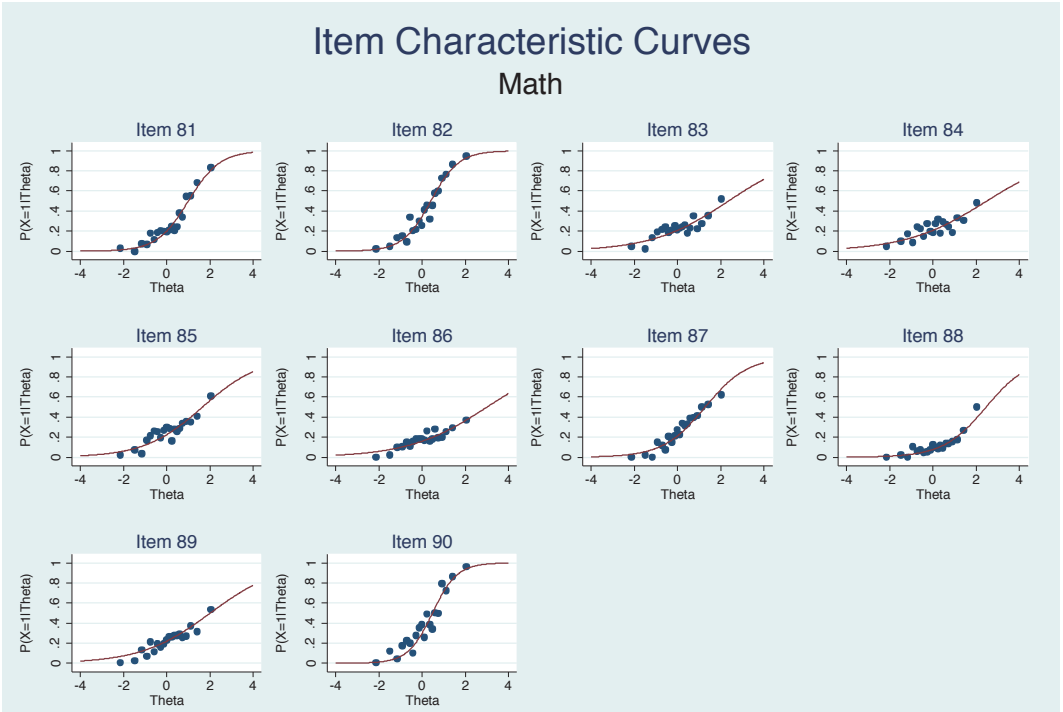
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

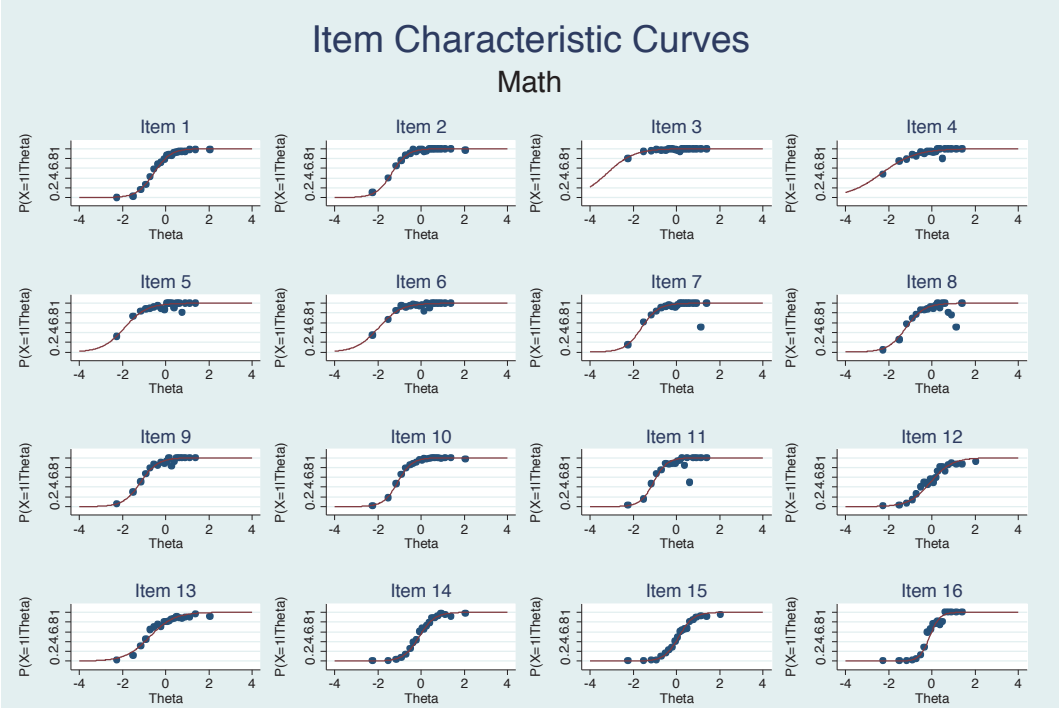


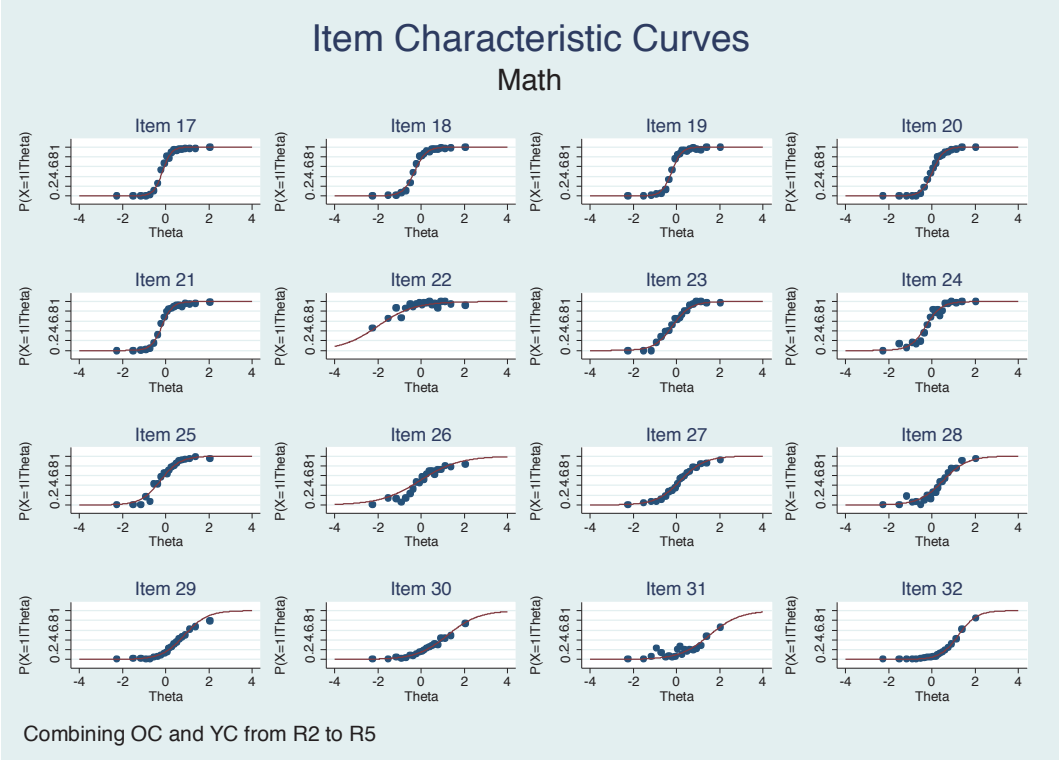
Combining OC and YC from R2 to R5

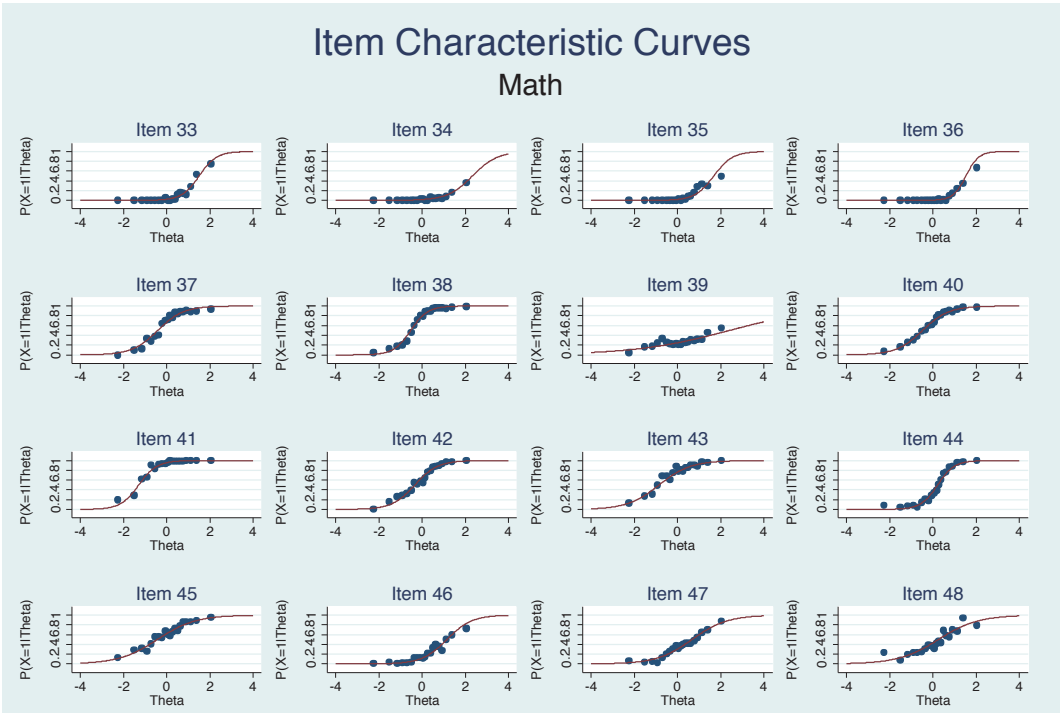


Combining OC and YC from R2 to R5

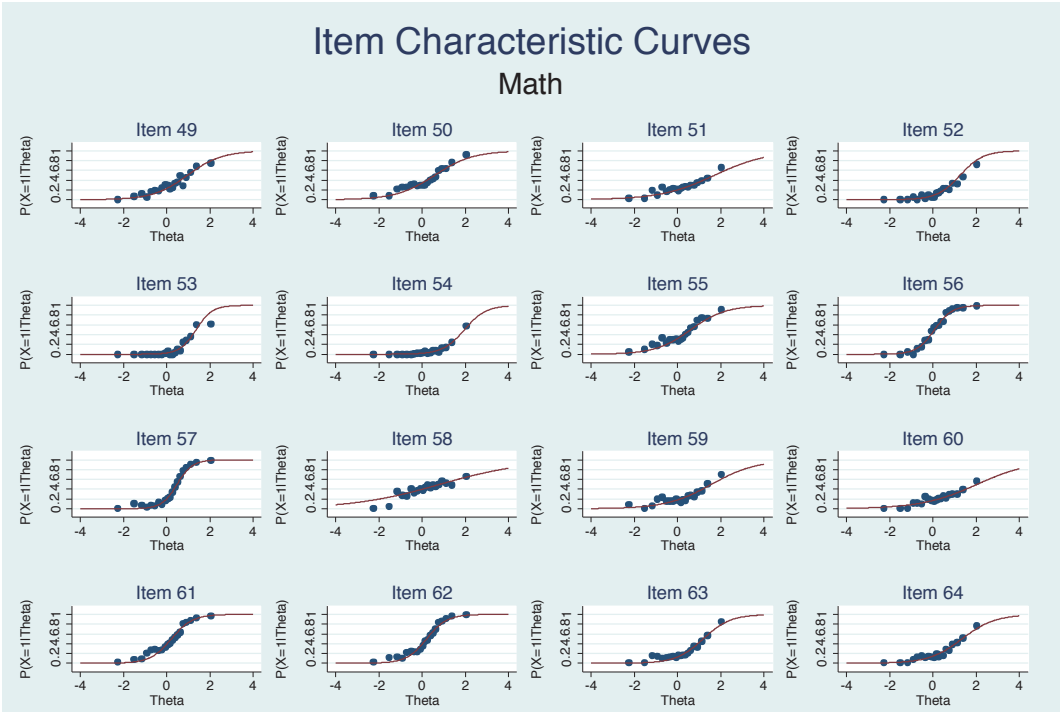
Figure 9. *Item Characteristic Curves for maths analysis, Peru*



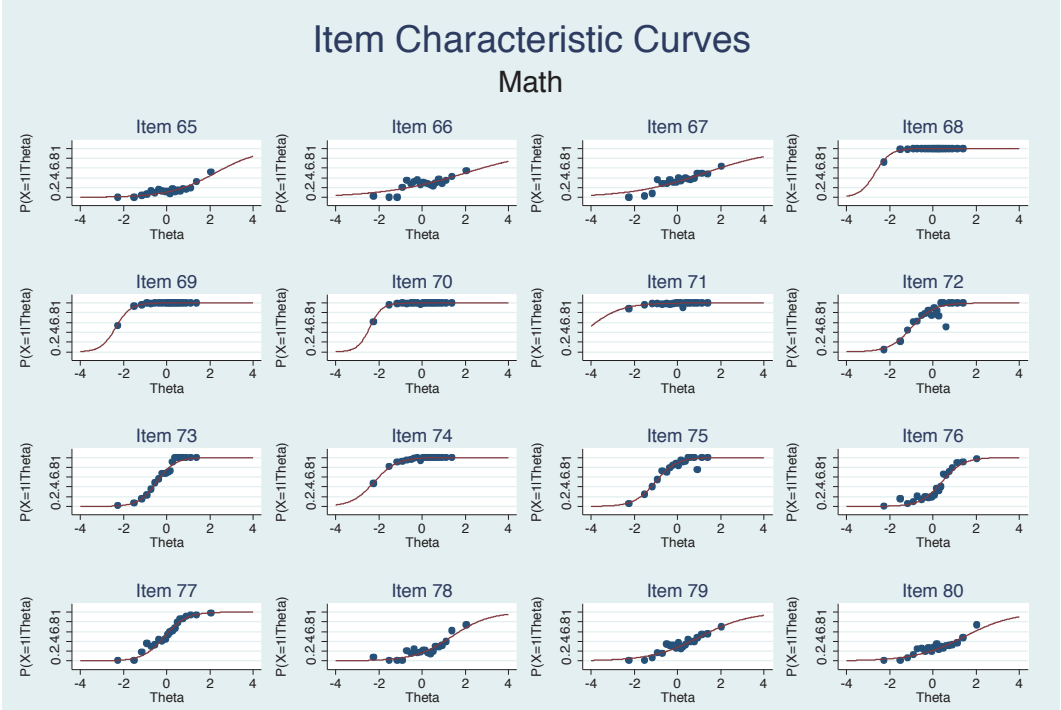




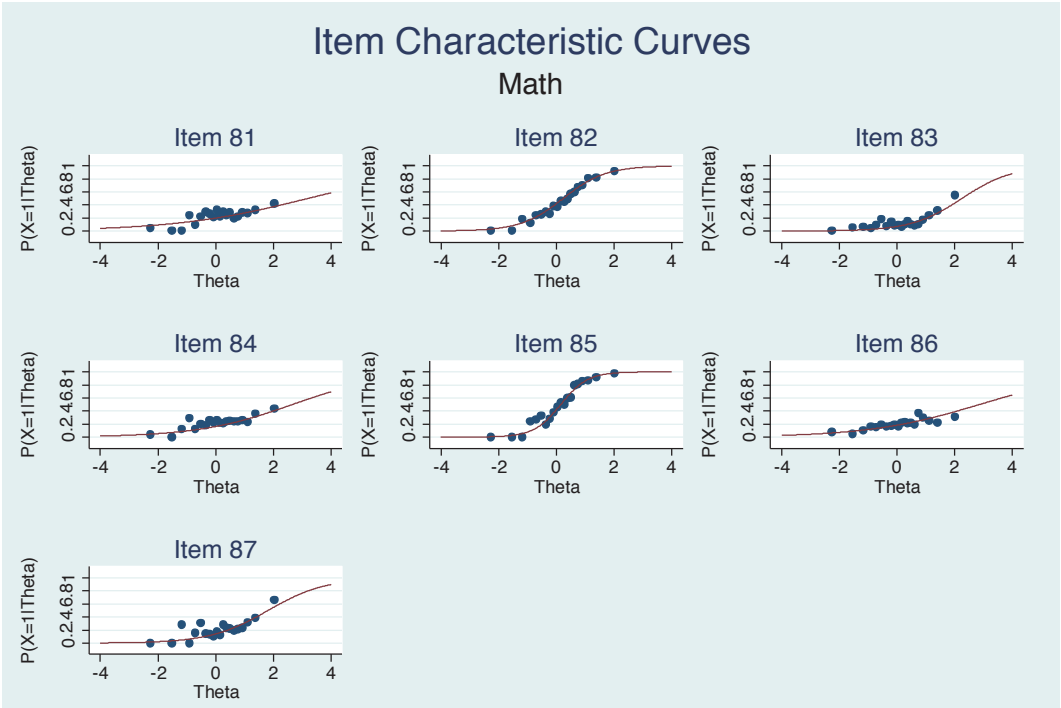
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

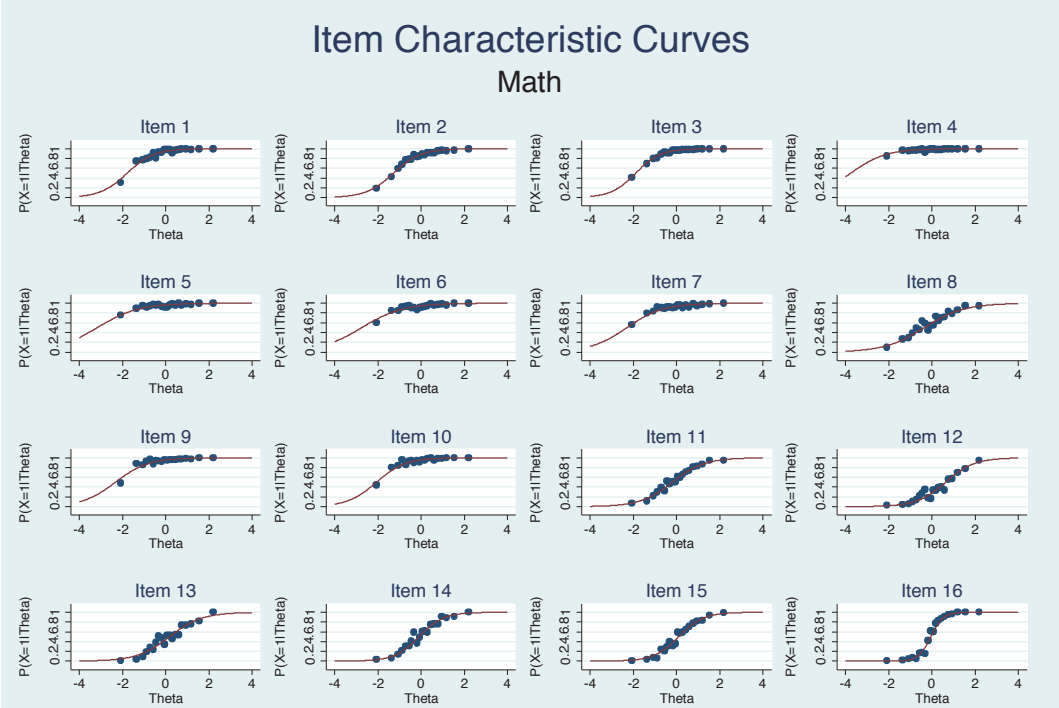


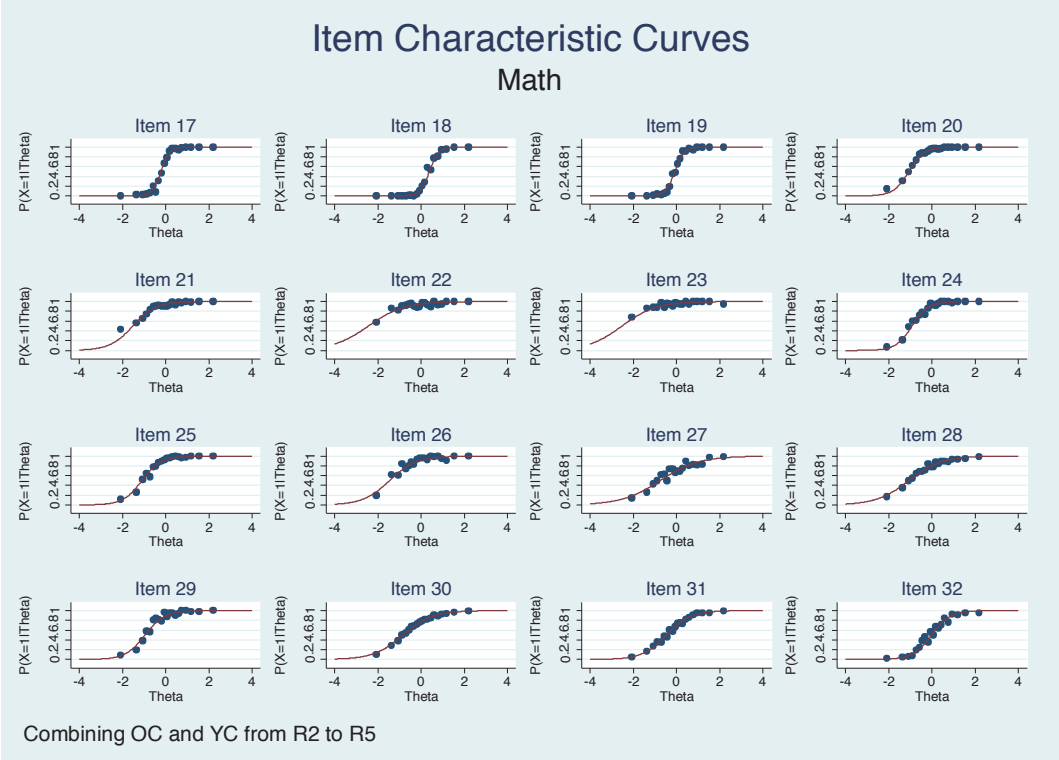
Combining OC and YC from R2 to R5

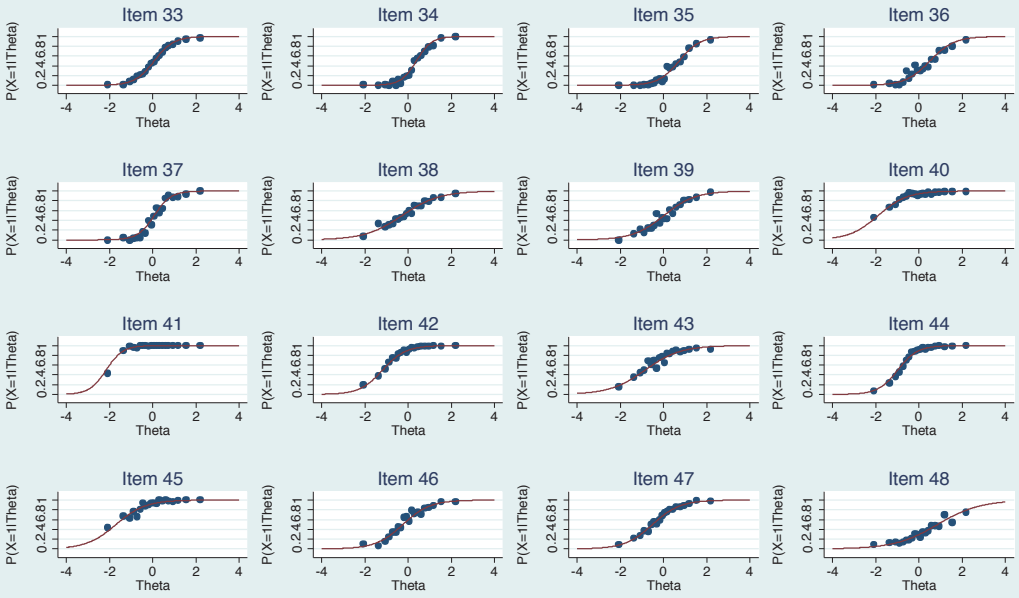


Combining OC and YC from R2 to R5

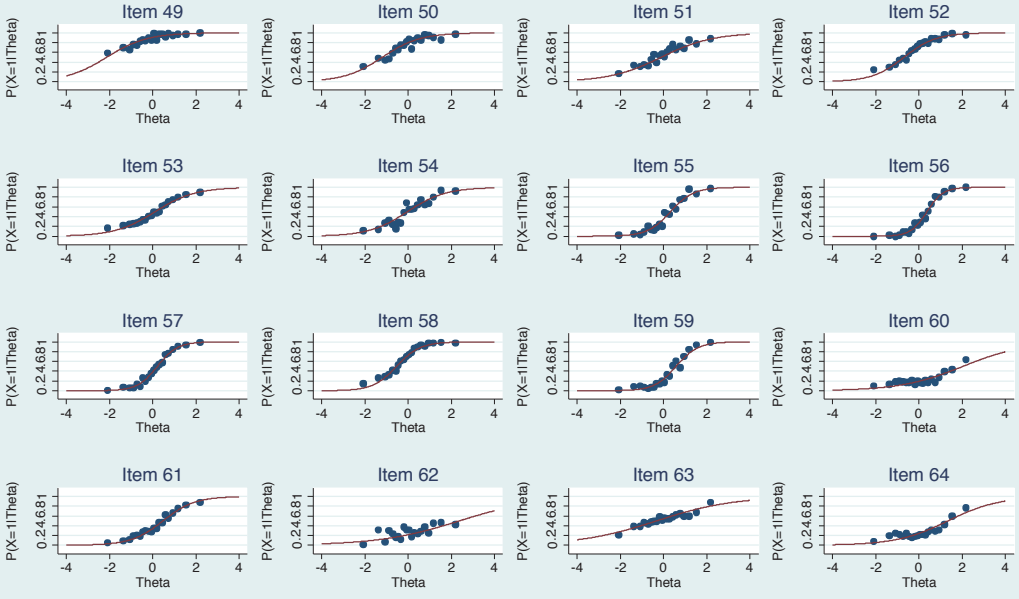
Figure 10. *Item Characteristic Curves for maths analysis, Vietnam*



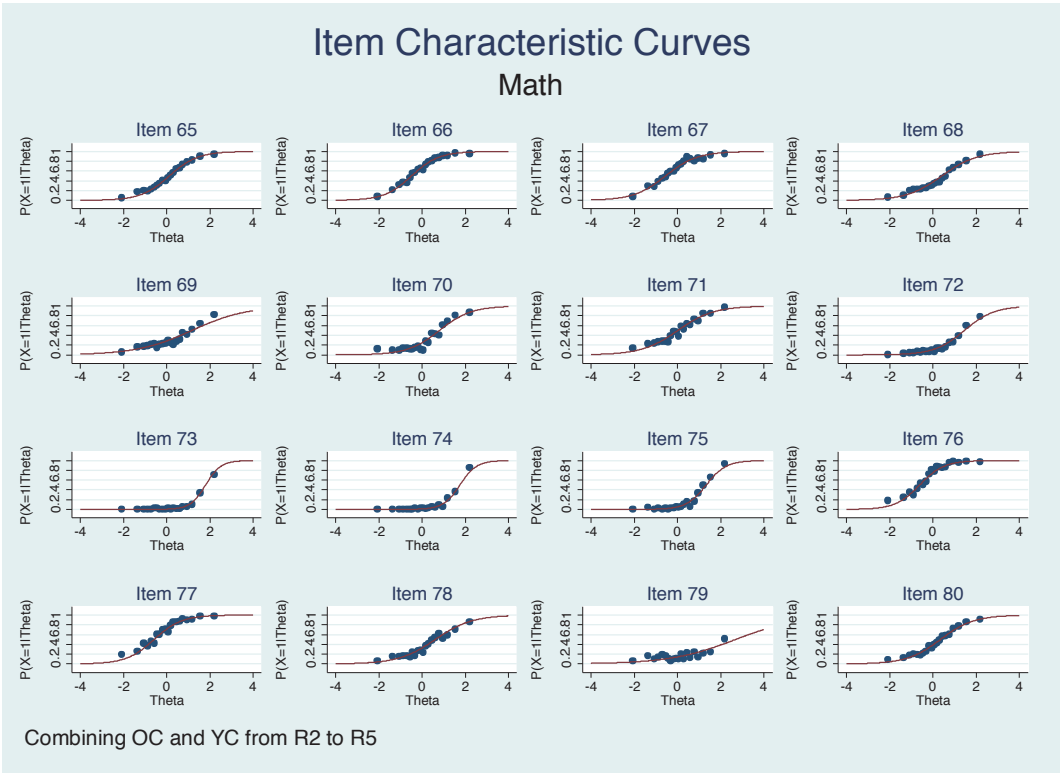




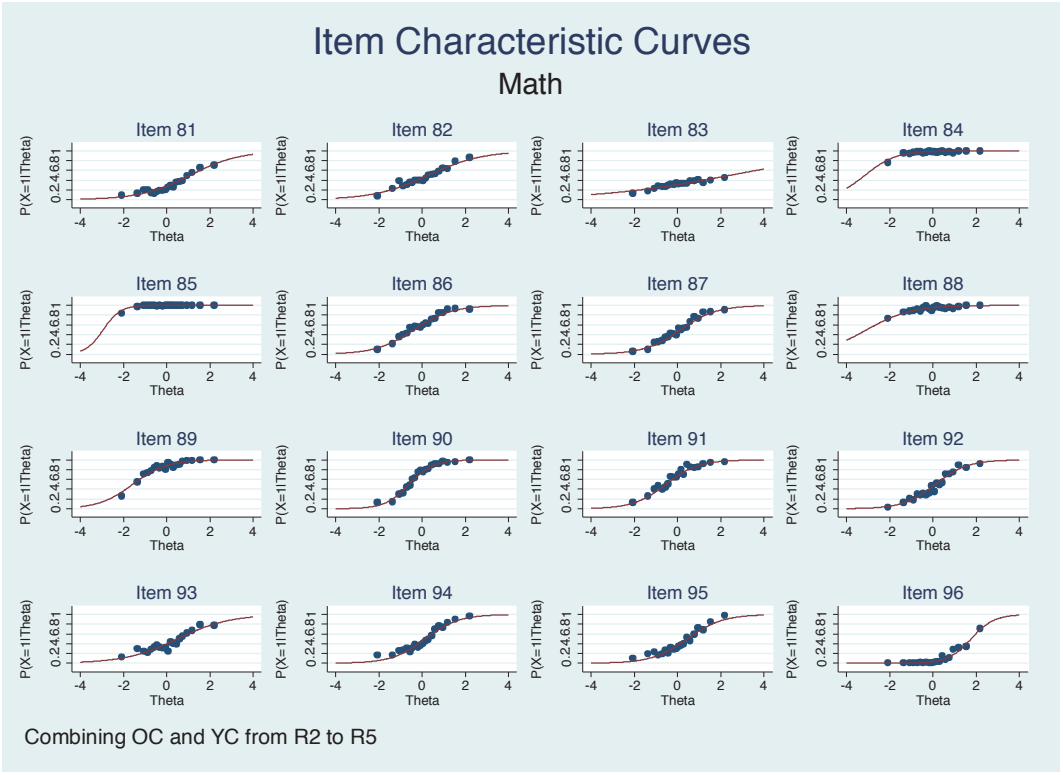
Combining OC and YC from R2 to R5



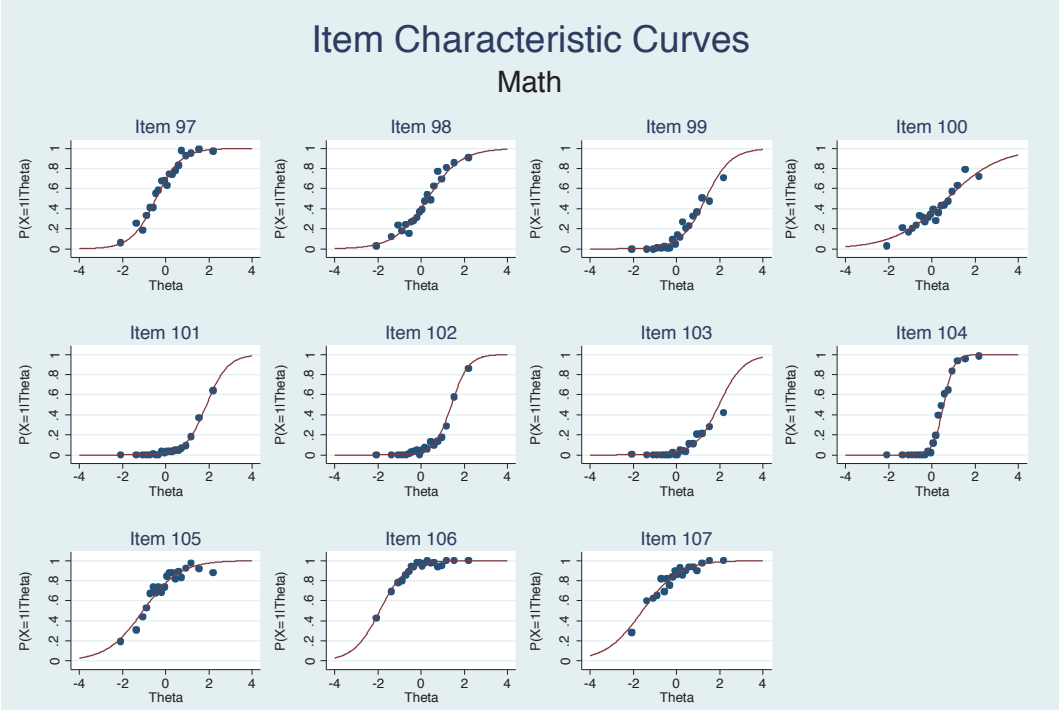
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

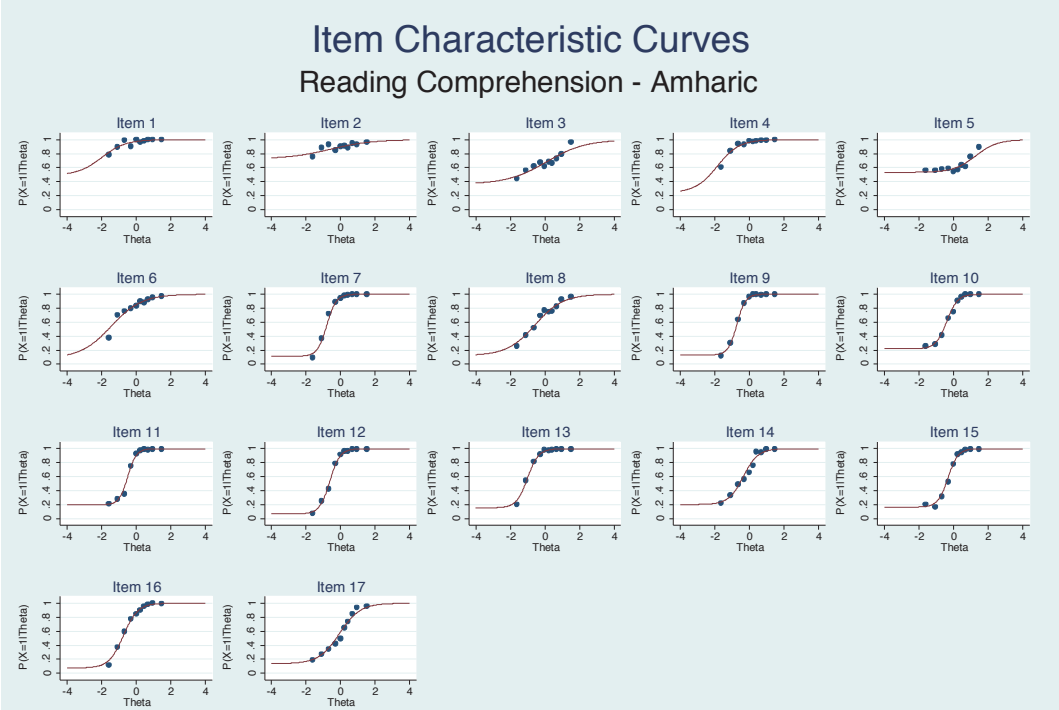


Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

Figure 11. *Item Characteristic Curves for reading comprehension analysis, Amharic*



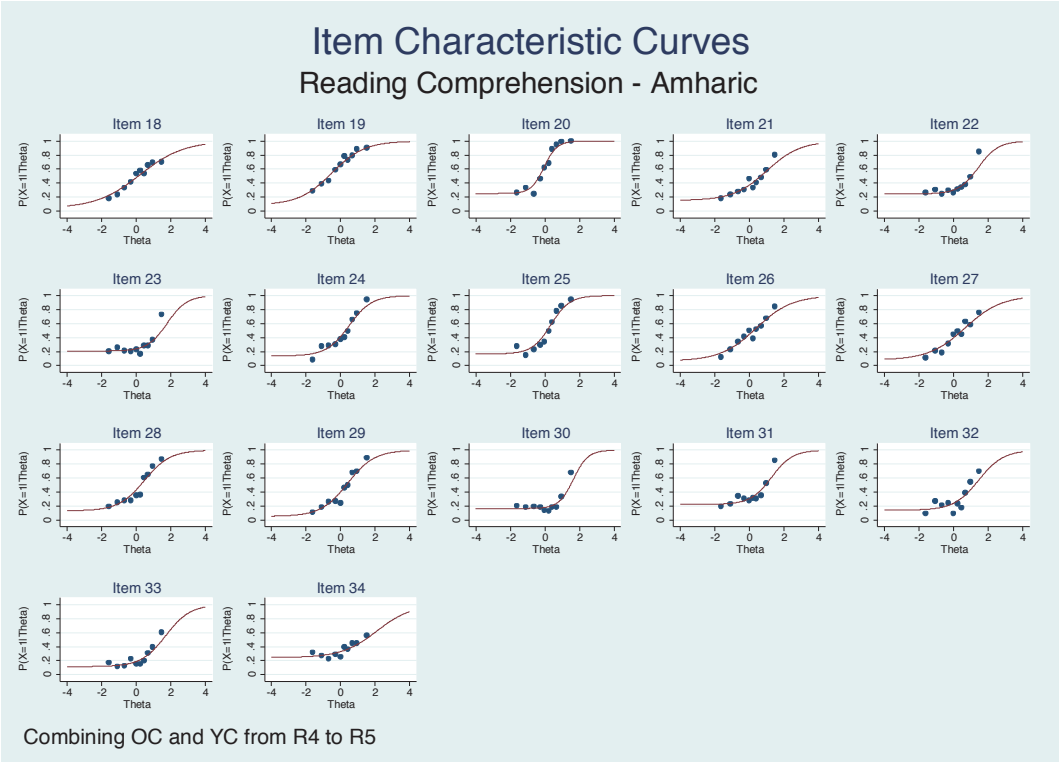
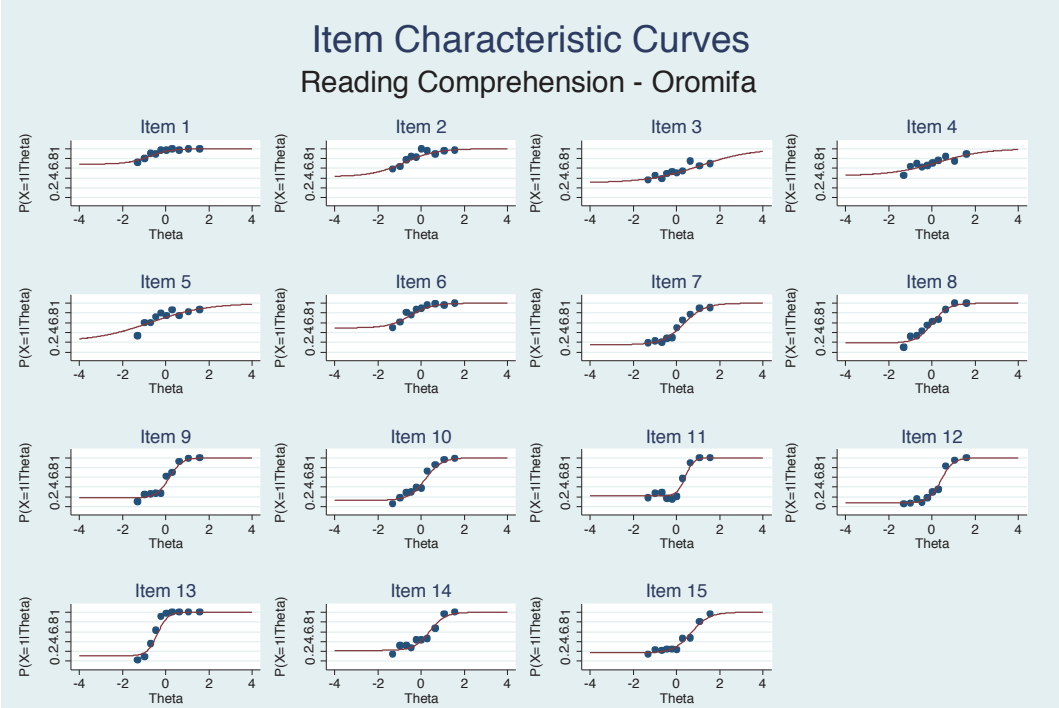


Figure 12. *Item Characteristic Curves for reading comprehension analysis, Oromifa*



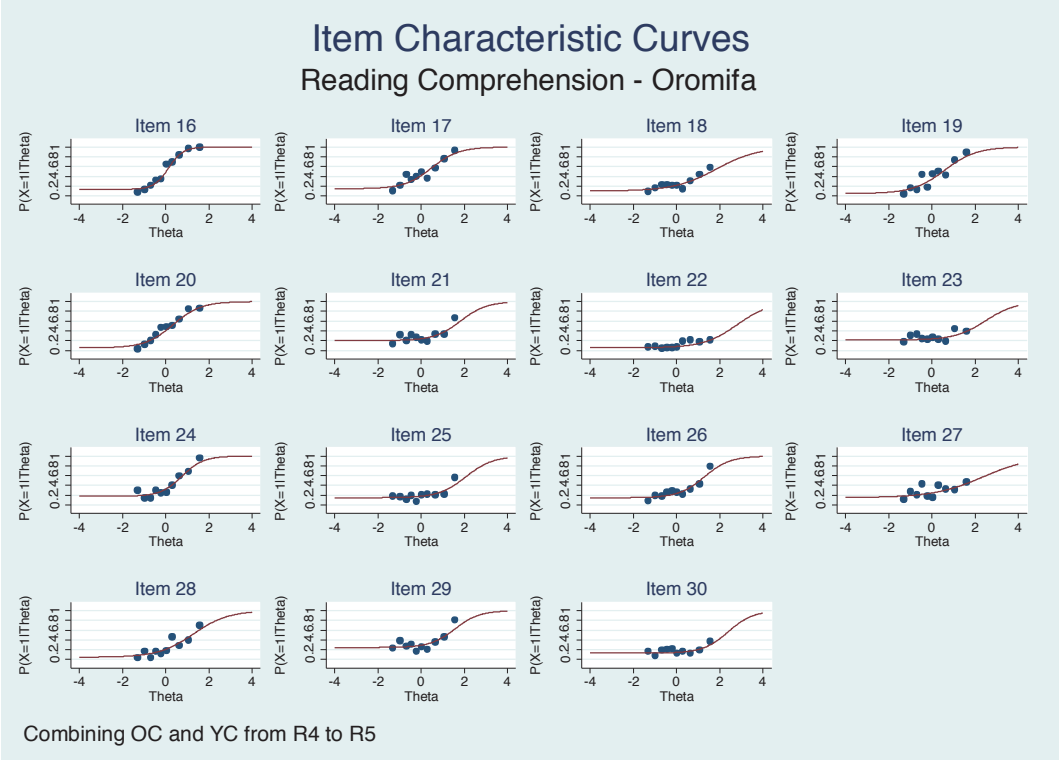
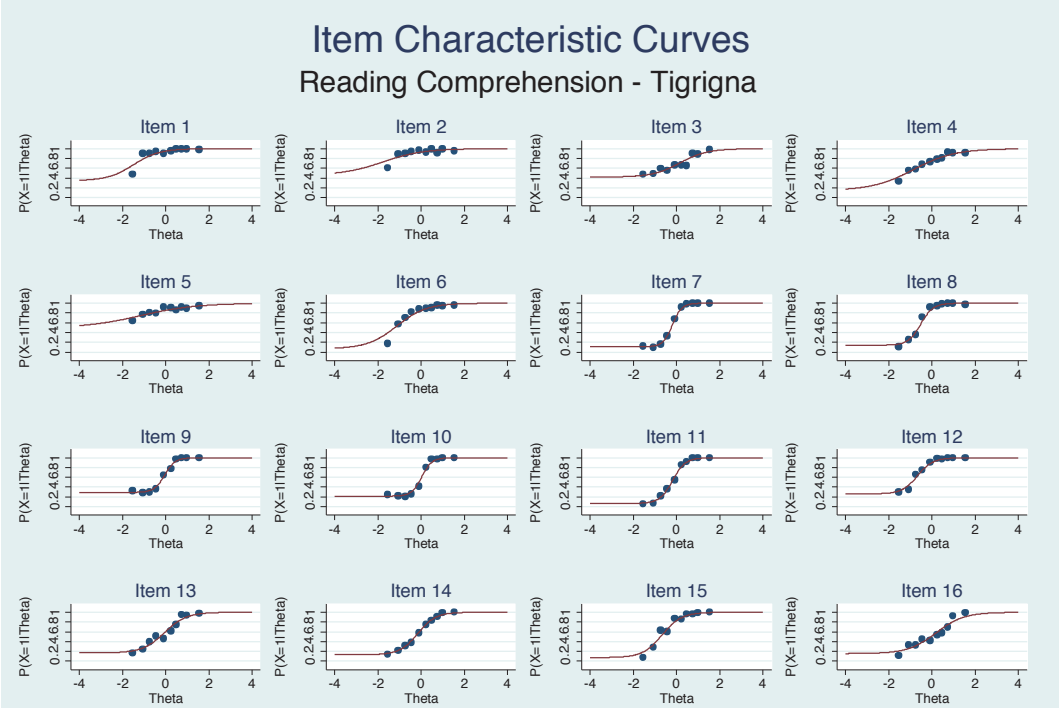


Figure 13. *Item Characteristic Curves for reading comprehension analysis, Tigrigna*



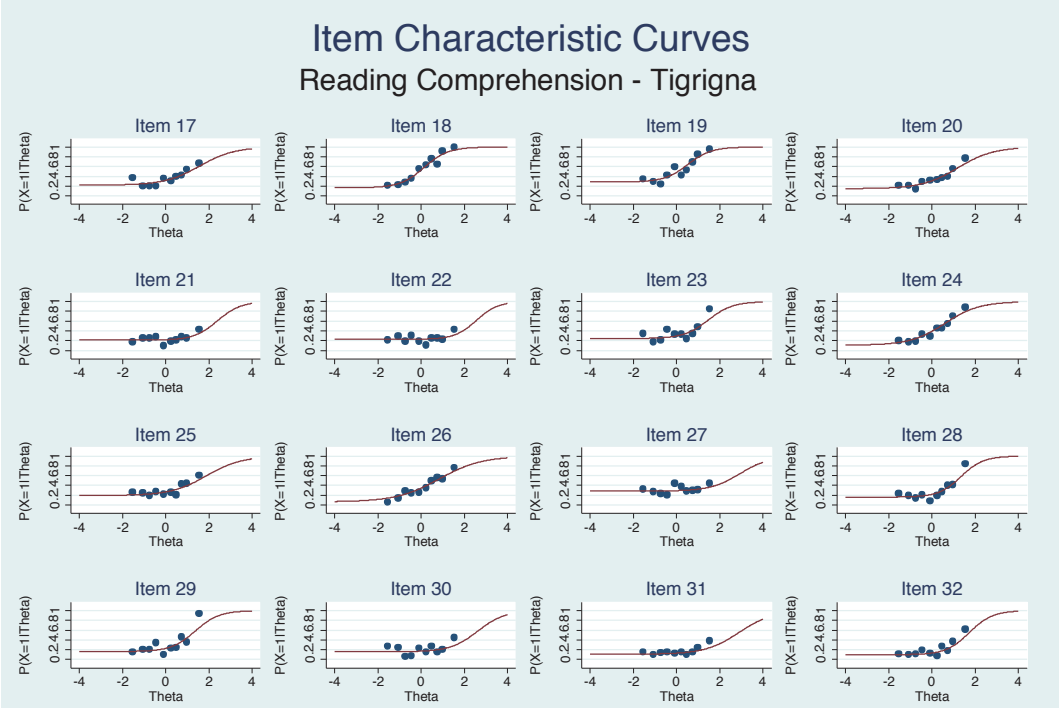
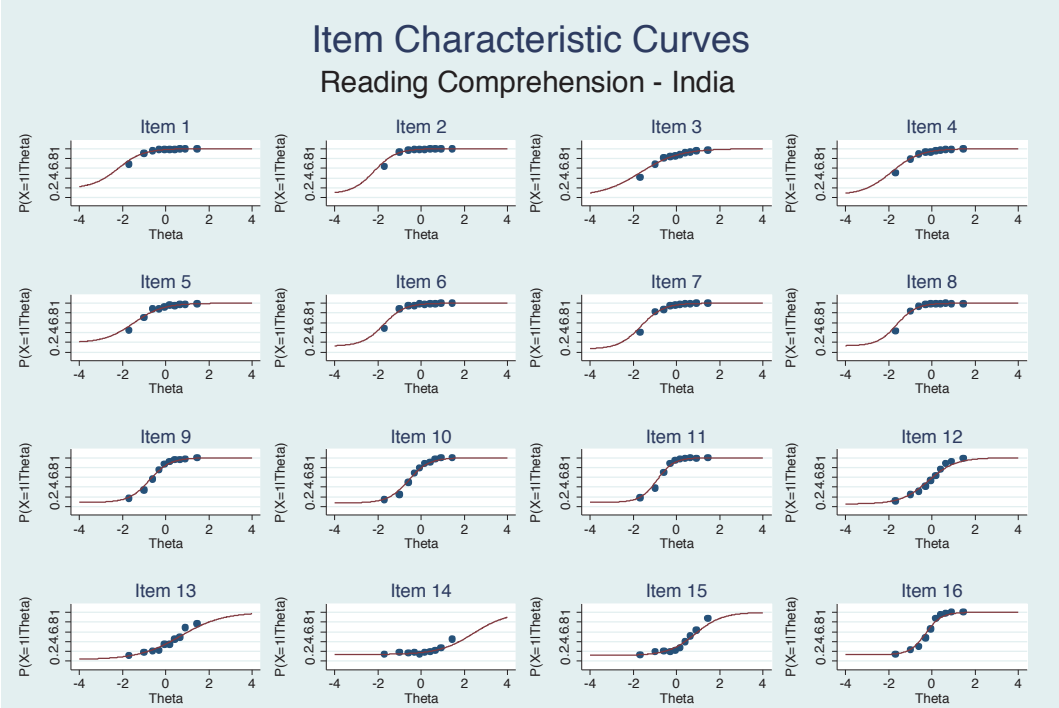


Figure 14. *Item Characteristic Curves for reading comprehension analysis, Telugu*



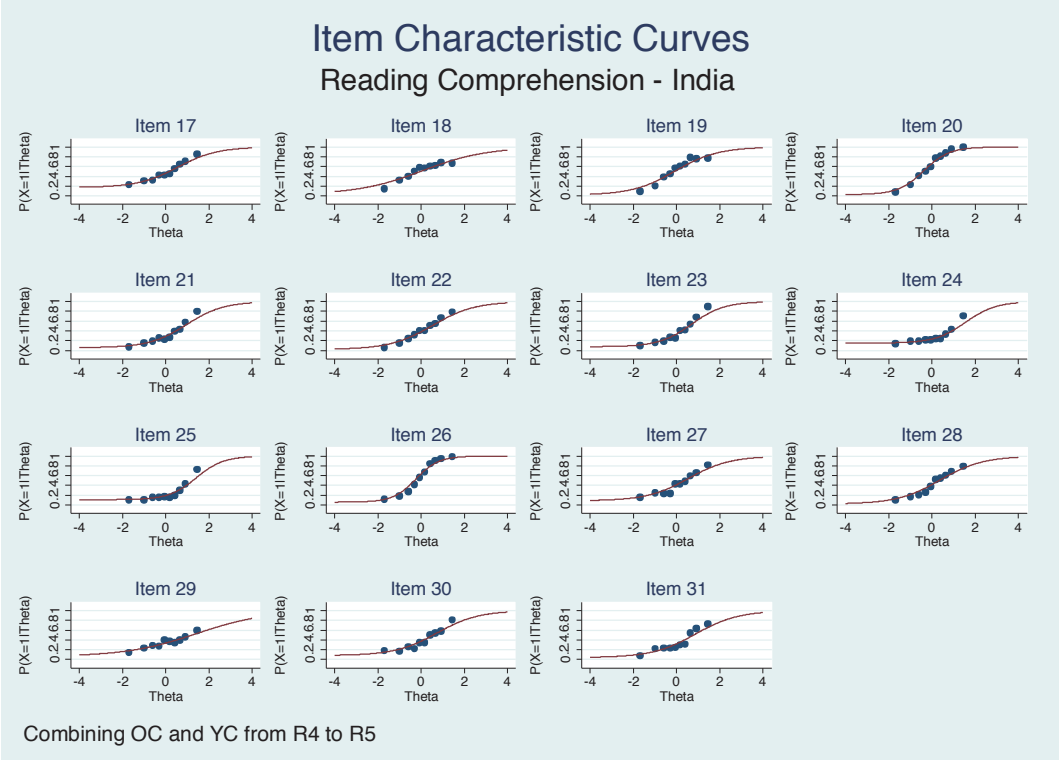
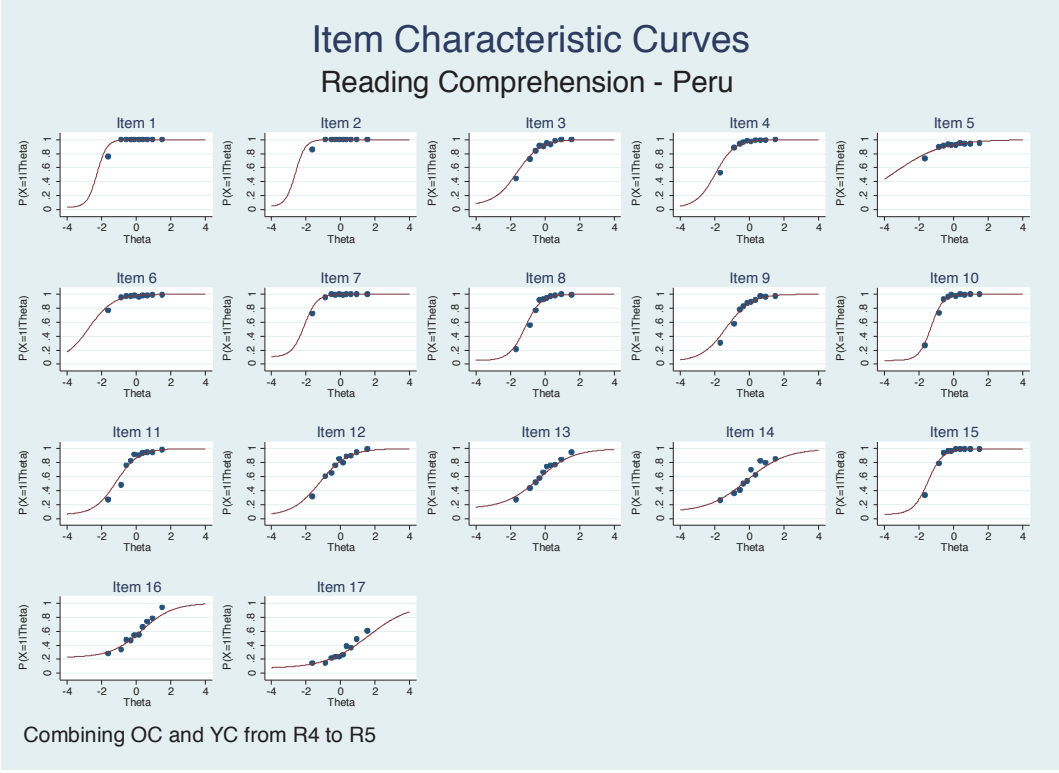


Figure 15. *Item Characteristic Curves for reading comprehension analysis, Spanish*



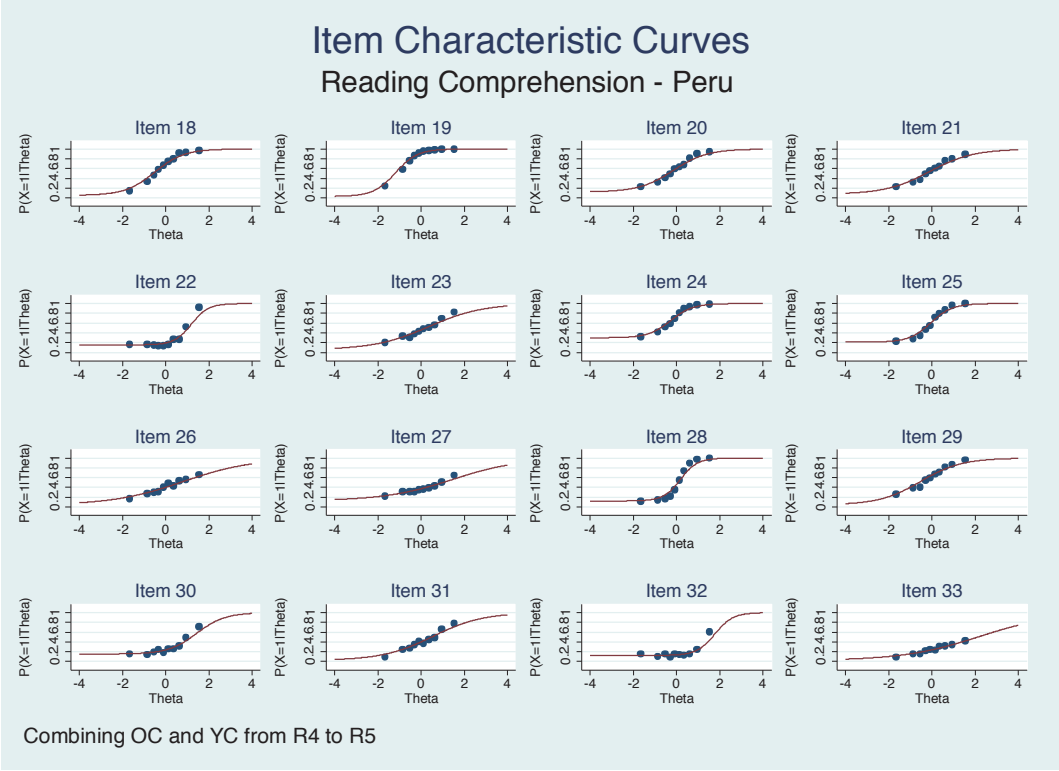
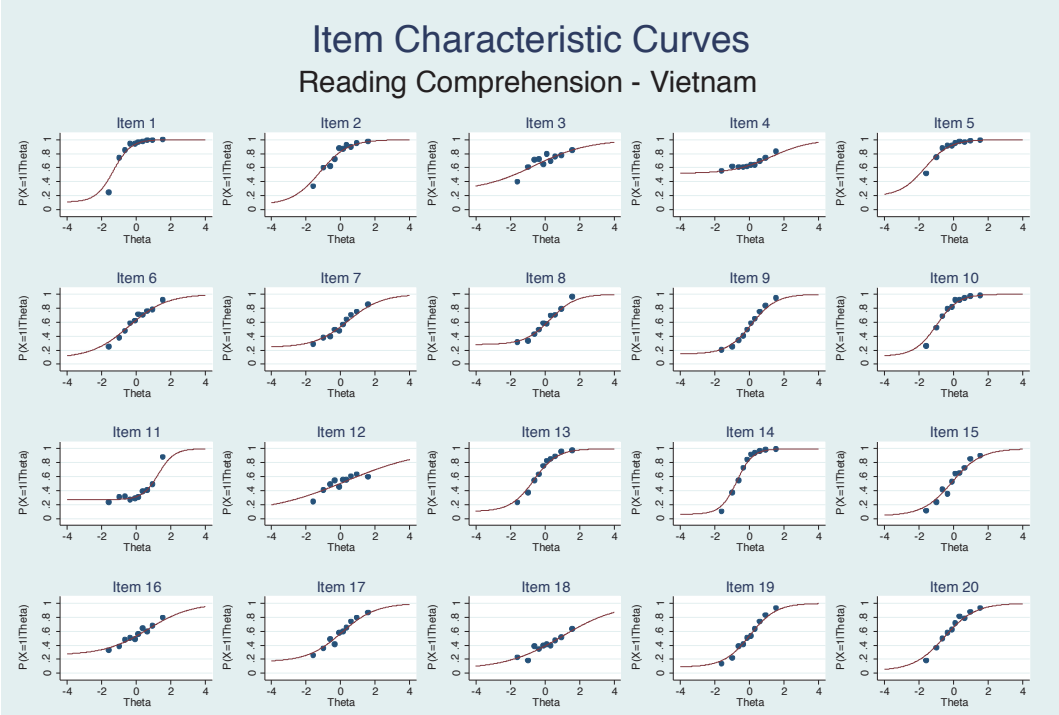
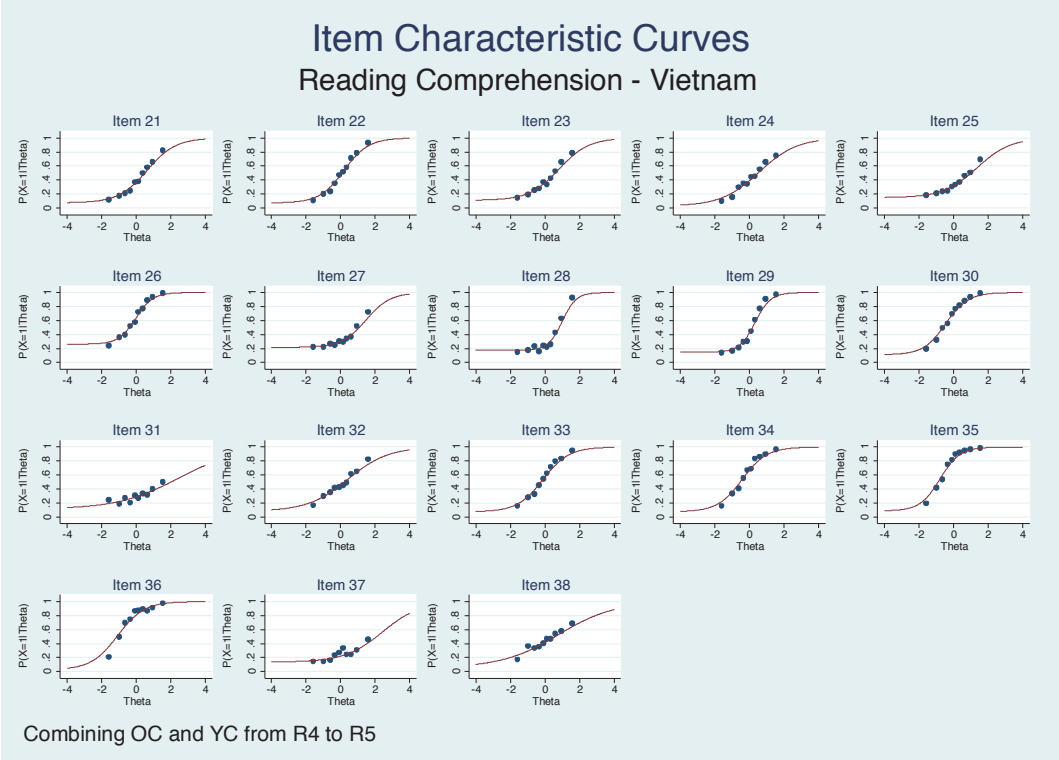


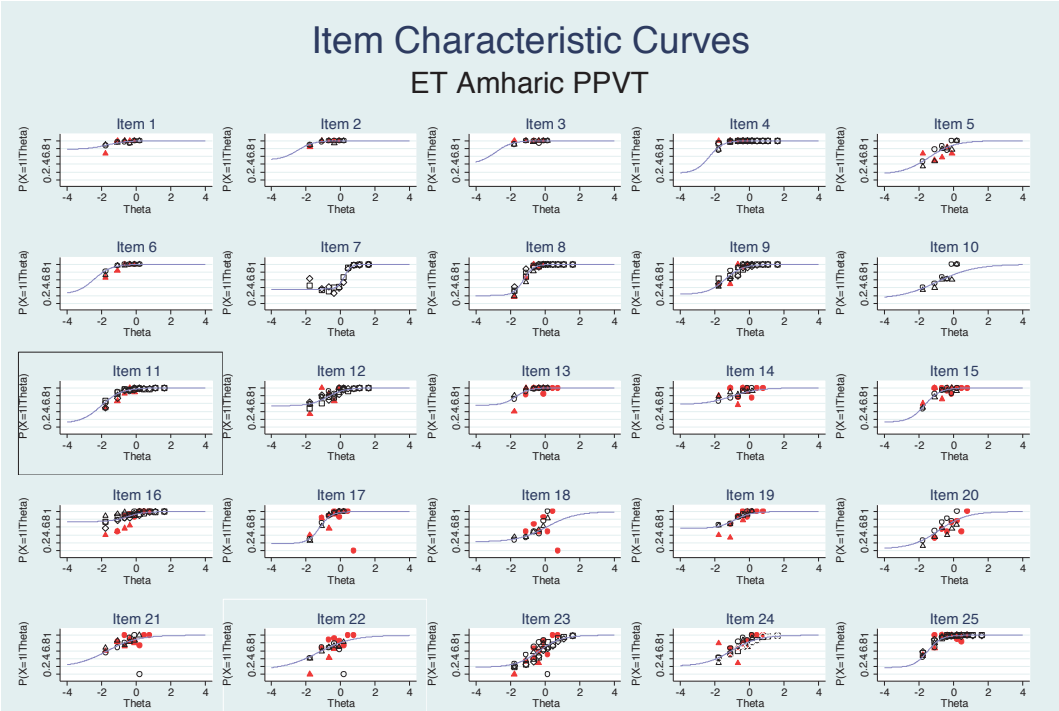
Figure 16. *Item Characteristic Curves for reading comprehension analysis, Vietnamese*

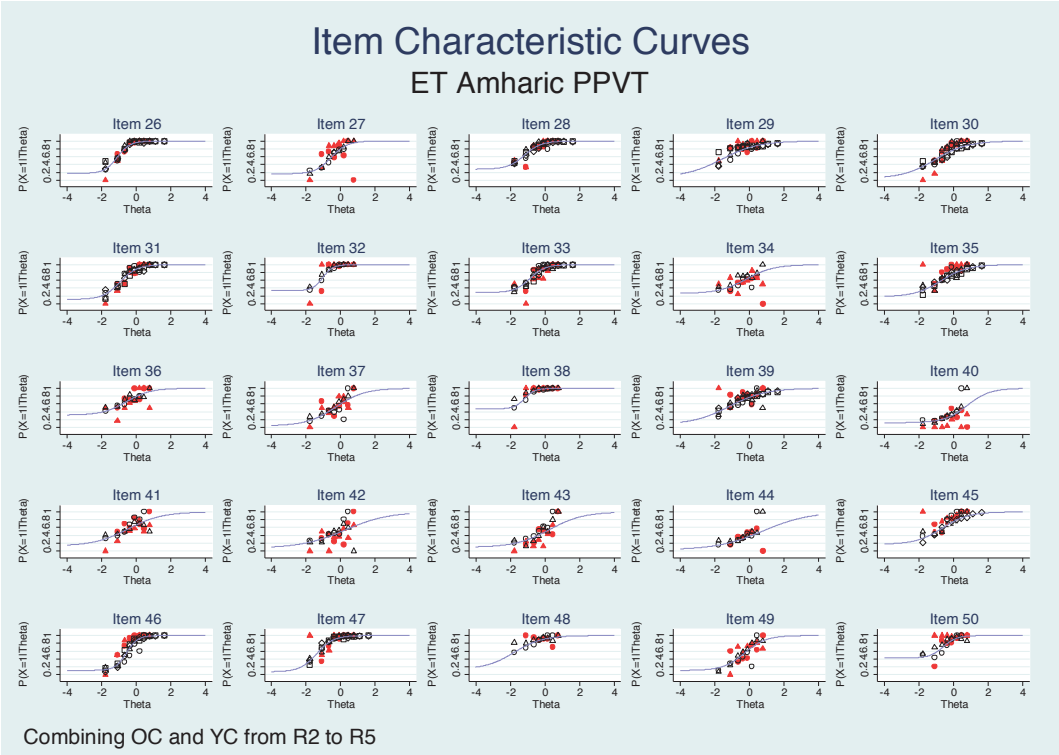


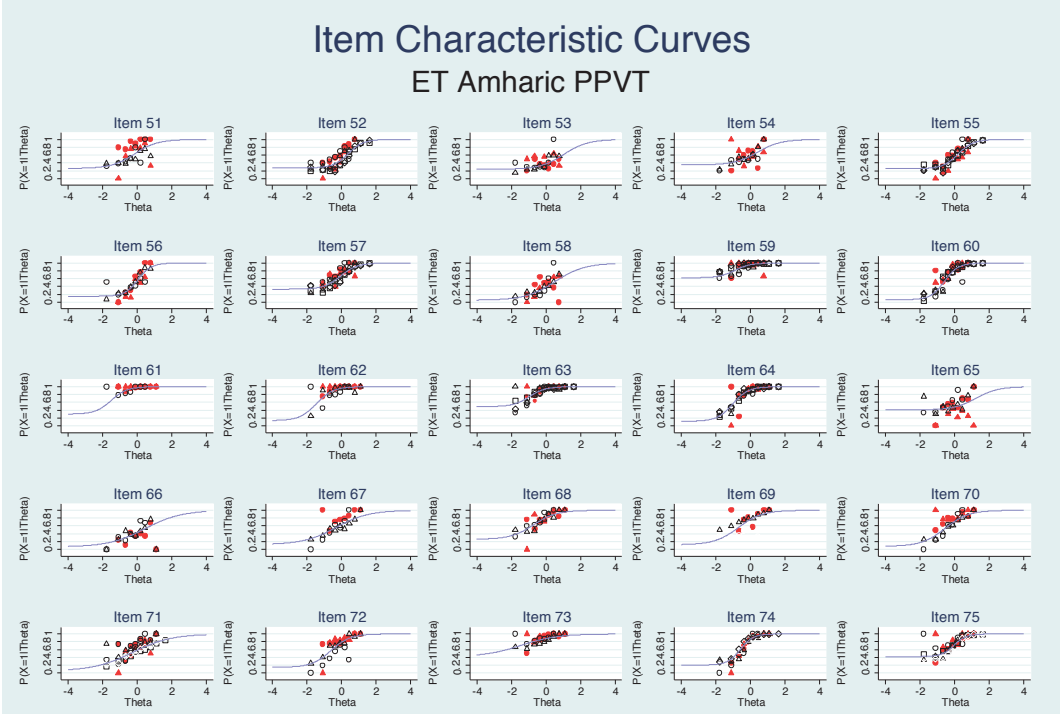


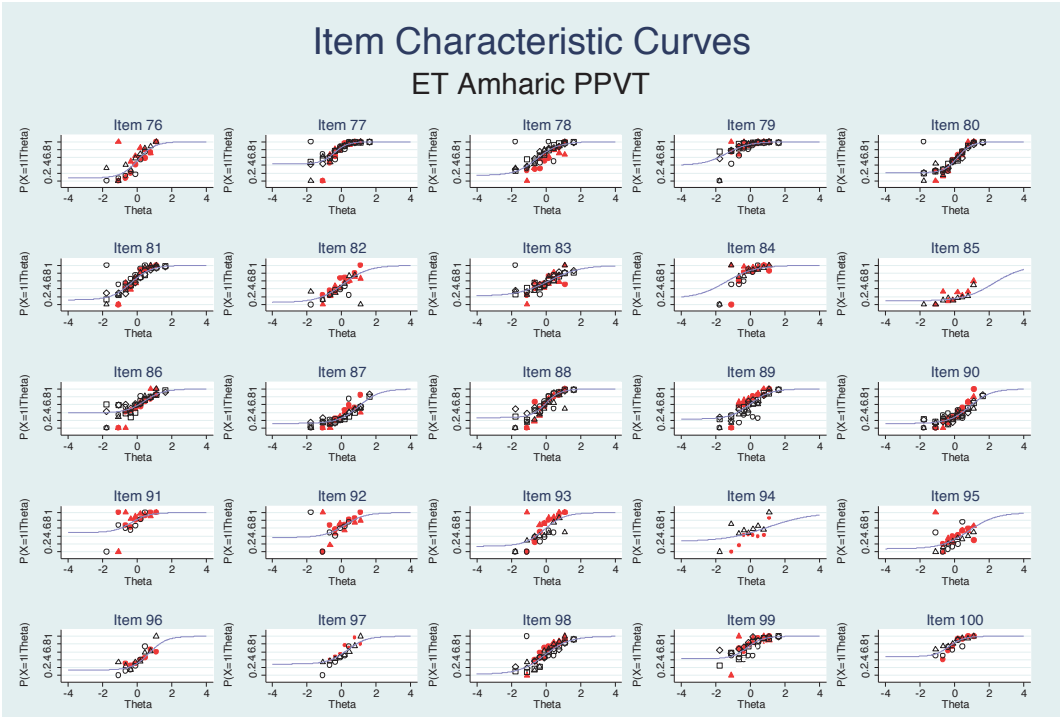
Appendix C.
Differential Item Functioning (DIF)

Figure 1. *Differential Item Functioning for PPVT analysis, Amharic*



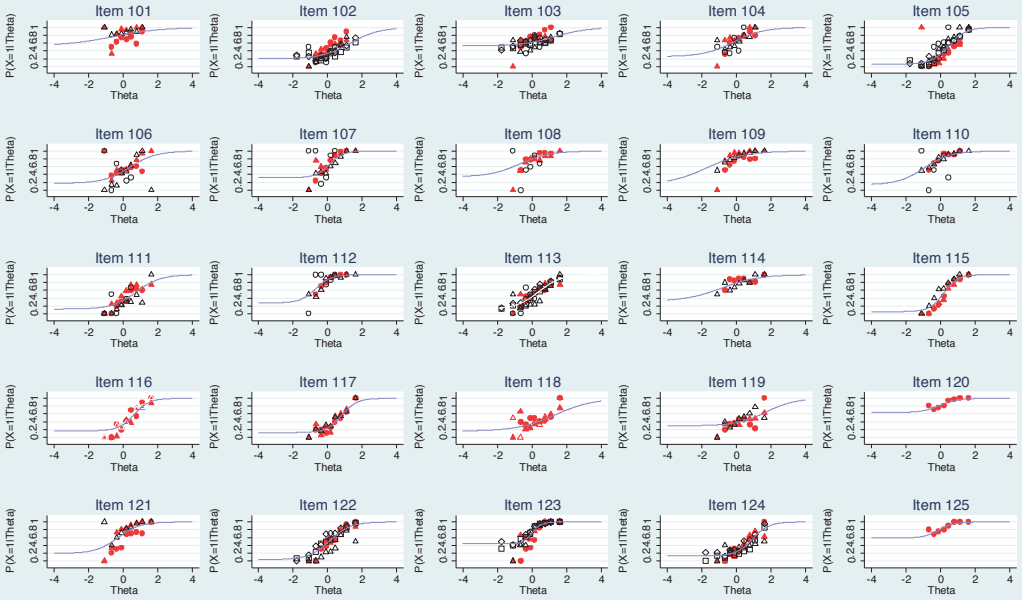



Item Characteristic Curves
ET Amharic PPVT
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
Item 65
Item 66
Item 67
Item 68
Item 69
Item 70
Item 71
Item 72
Item 73
Item 74
Item 75
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5


Item Characteristic Curves
ET Amharic PPVT
Item 76
Item 77
Item 78
Item 79
Item 80
Item 81
Item 82
Item 83
Item 84
Item 85
Item 86
Item 87
Item 88
Item 89
Item 90
Item 91
Item 92
Item 93
Item 94
Item 95
Item 96
Item 97
Item 98
Item 99
Item 100
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

Item Characteristic Curves

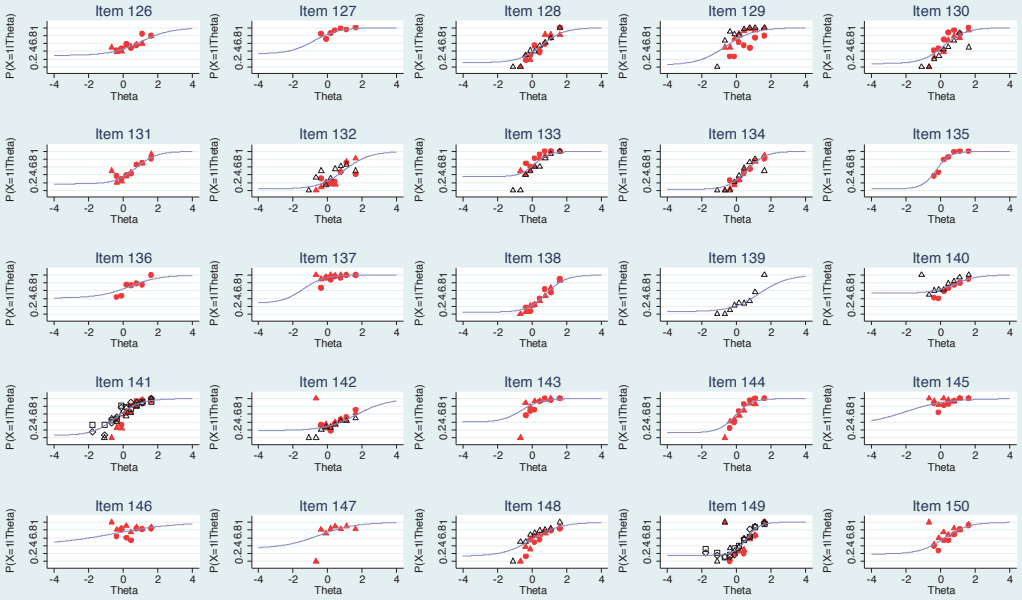
ET Amharic PPVT



Combining OC and YC from R2 to R5

Item Characteristic Curves

ET Amharic PPVT



Combining OC and YC from R2 to R5

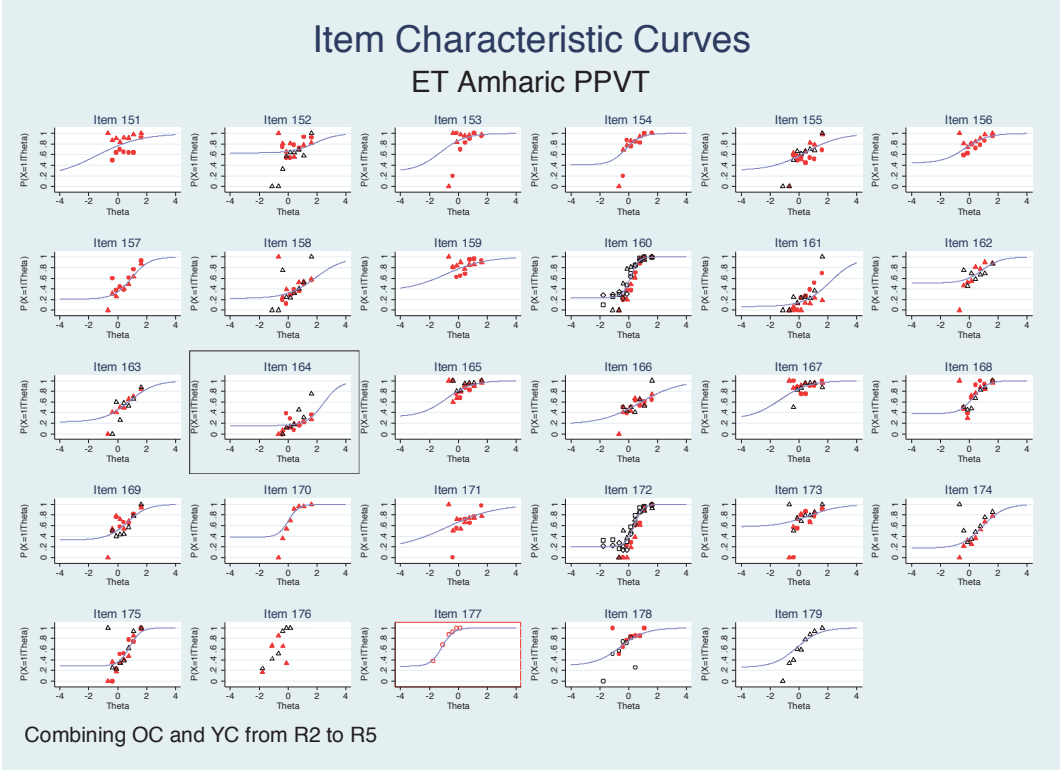
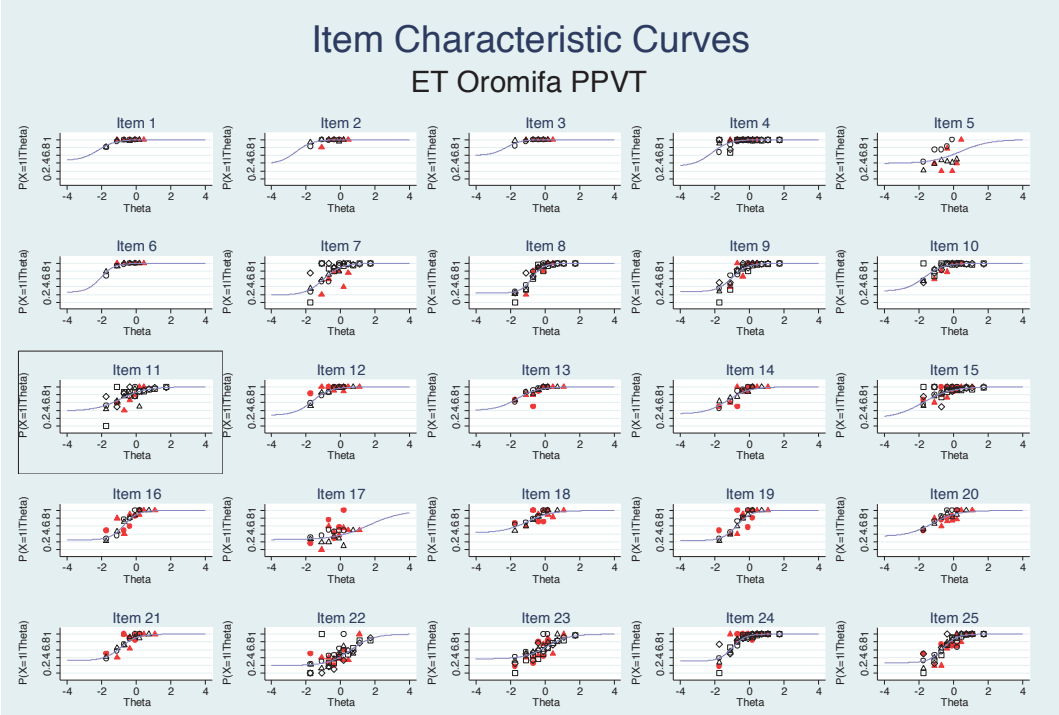
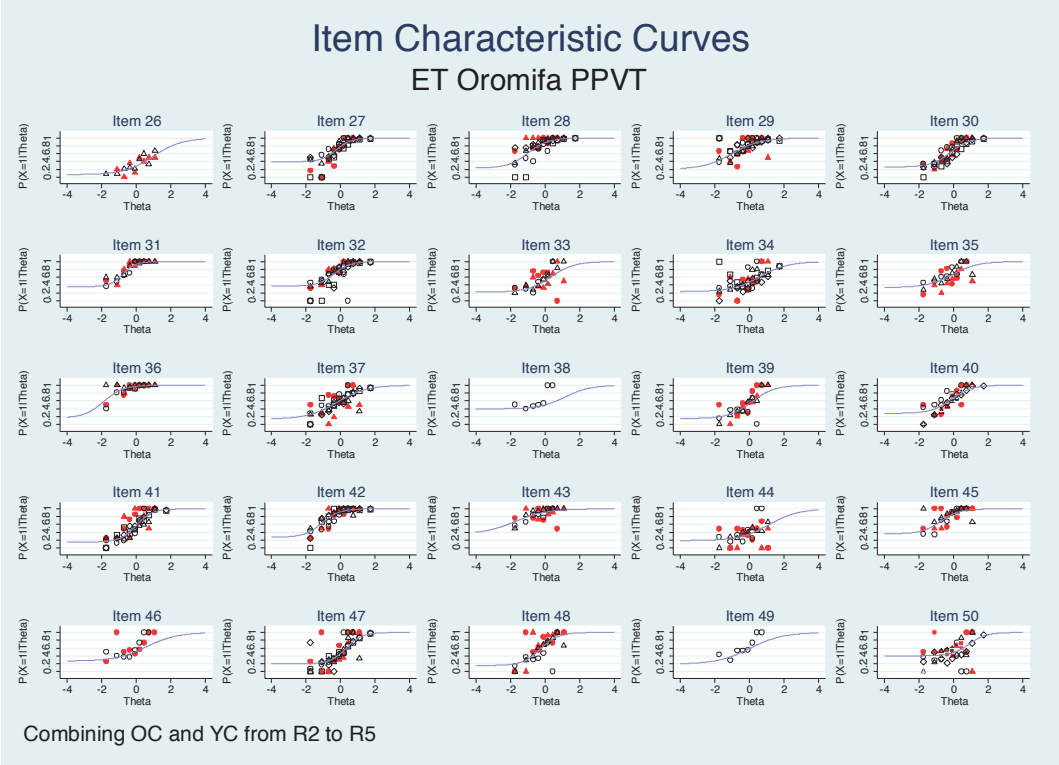
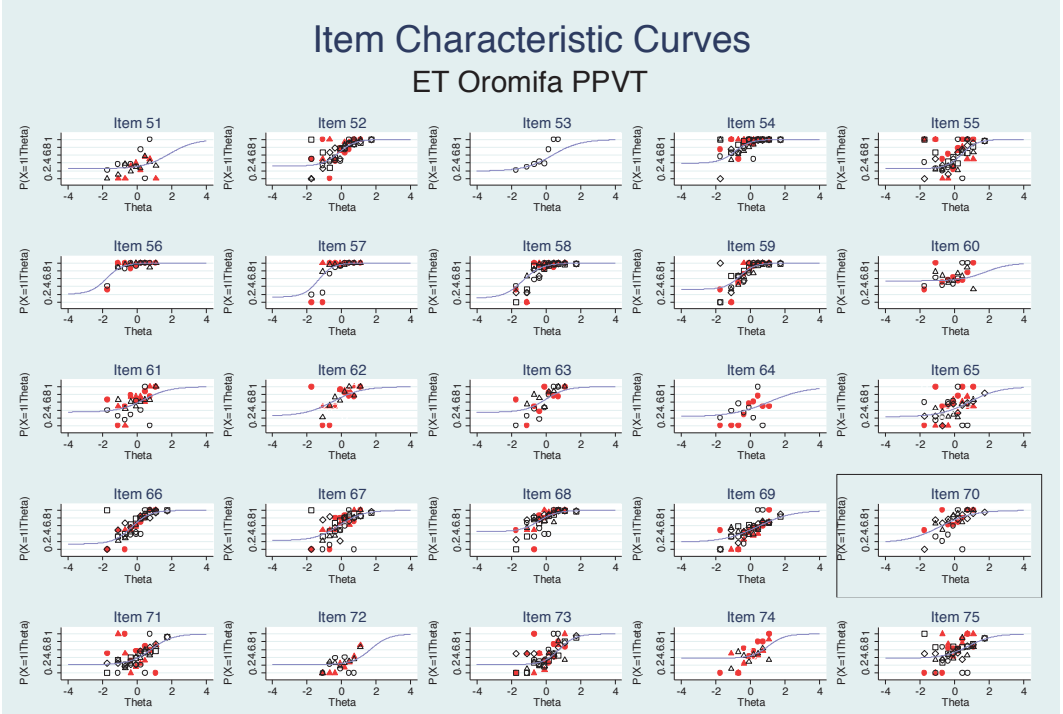
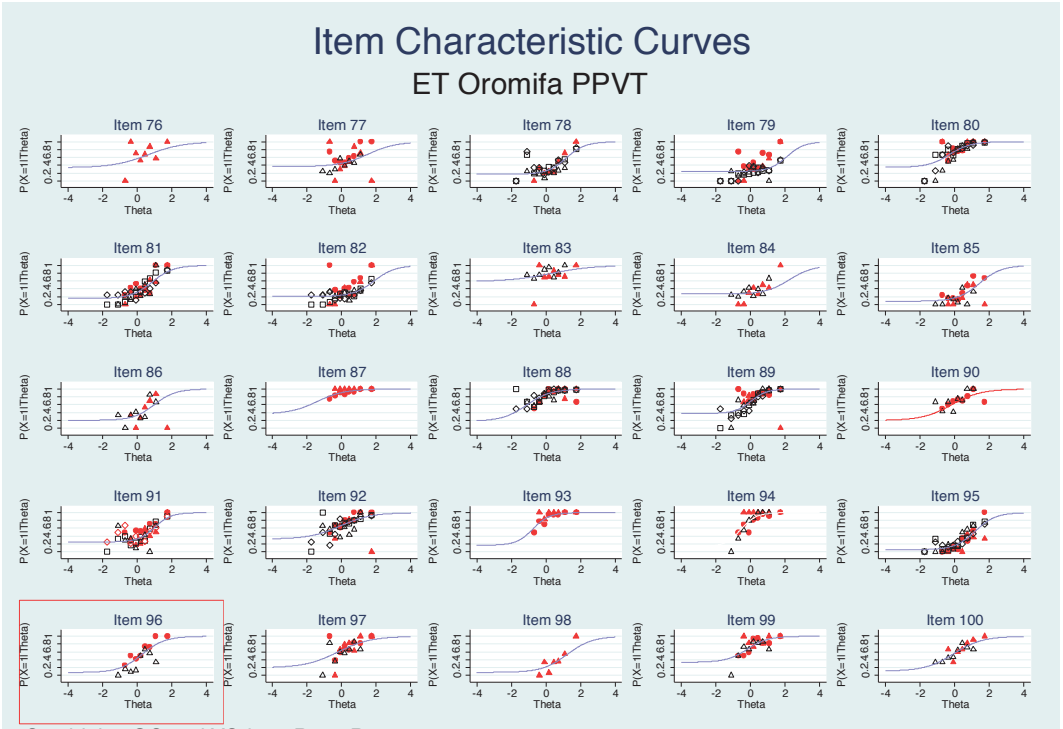

Item Characteristic Curves
ET Amharic PPVT
Item 151
Item 152
Item 153
Item 154
Item 155
Item 156
Item 157
Item 158
Item 159
Item 160
Item 161
Item 162
Item 163
Item 164
Item 165
Item 166
Item 167
Item 168
Item 169
Item 170
Item 171
Item 172
Item 173
Item 174
Item 175
Item 176
Item 177
Item 178
Item 179
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

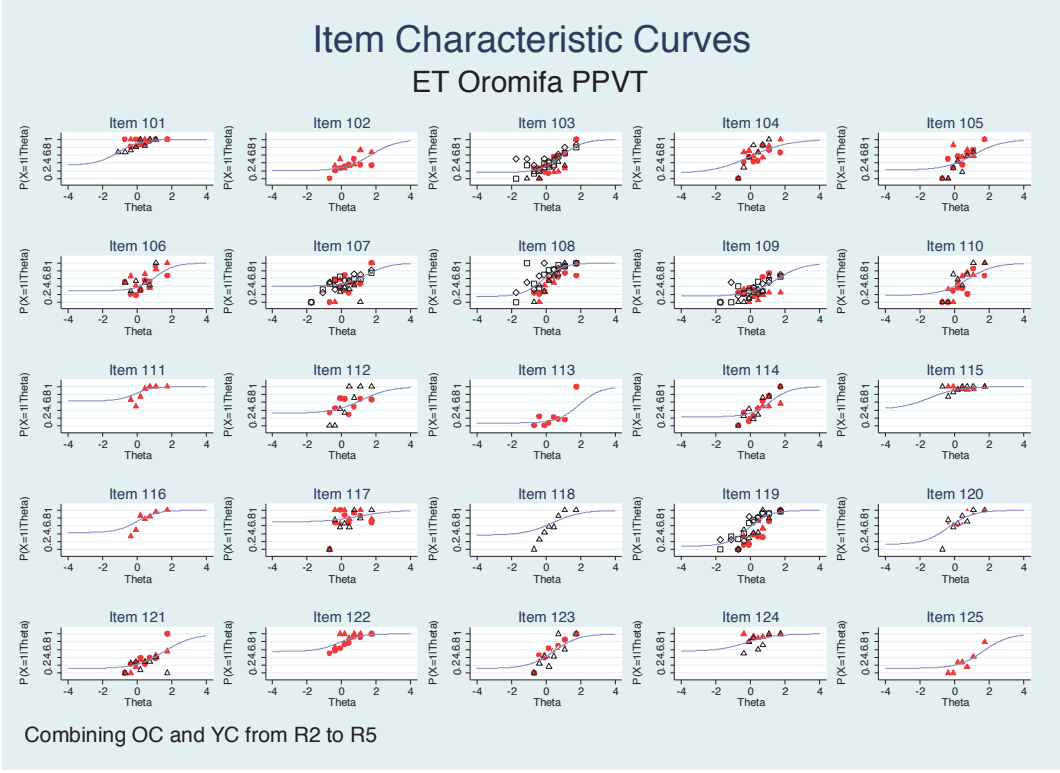
Figure 2. *Differential Item Functioning for PPVT analysis, Oromifa*






Item Characteristic Curves
ET Oromifa PPVT
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
Item 65
Item 66
Item 67
Item 68
Item 69
Item 70
Item 71
Item 72
Item 73
Item 74
Item 75
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5


Item Characteristic Curves
ET Oromifa PPVT
Item 76
Item 77
Item 78
Item 79
Item 80
Item 81
Item 82
Item 83
Item 84
Item 85
Item 86
Item 87
Item 88
Item 89
Item 90
Item 91
Item 92
Item 93
Item 94
Item 95
Item 96
Item 97
Item 98
Item 99
Item 100
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5


Item Characteristic Curves
ET Oromifa PPVT
Item 101
Item 102
Item 103
Item 104
Item 105
Item 106
Item 107
Item 108
Item 109
Item 110
Item 111
Item 112
Item 113
Item 114
Item 115
Item 116
Item 117
Item 118
Item 119
Item 120
Item 121
Item 122
Item 123
Item 124
Item 125
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

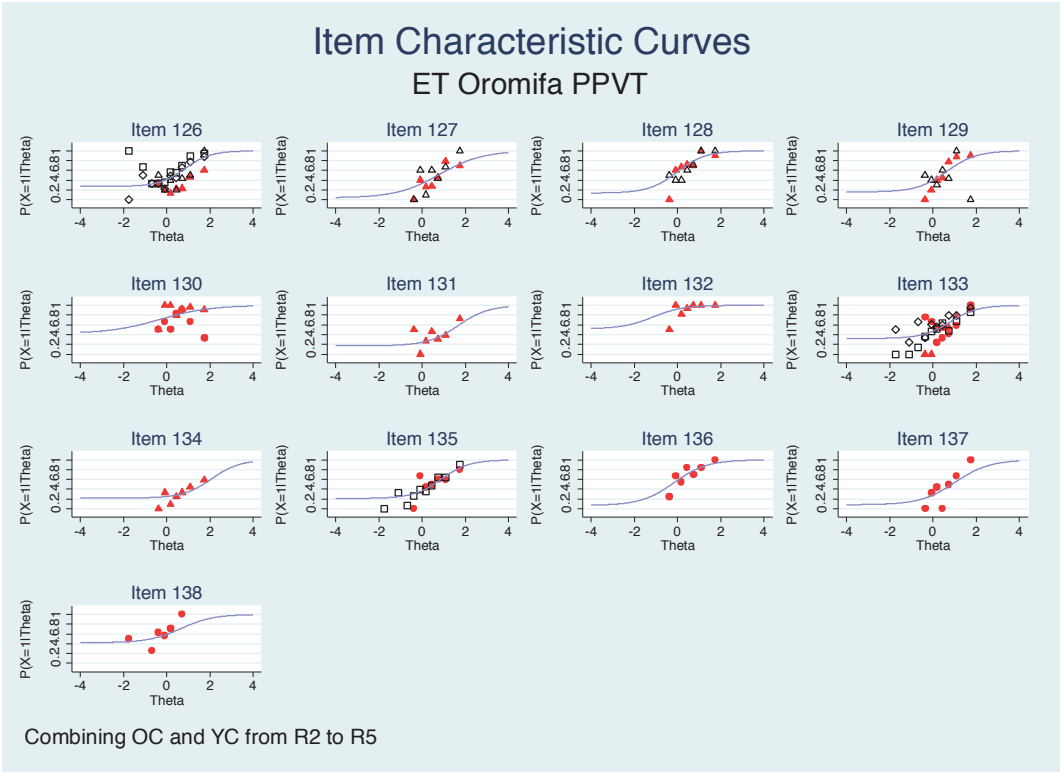
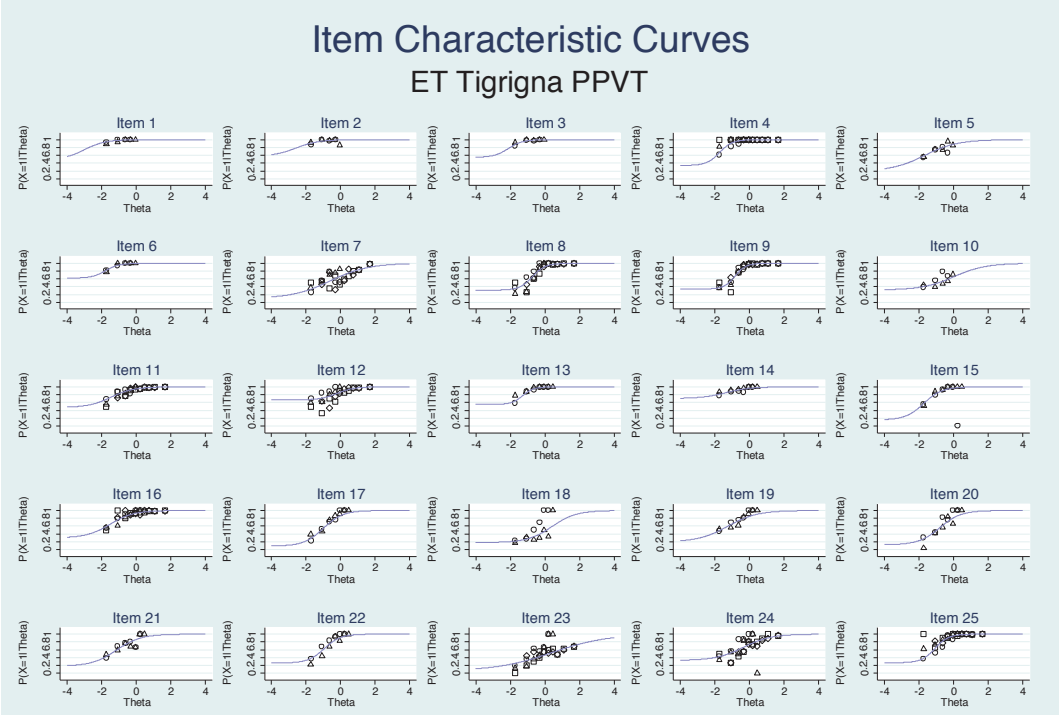
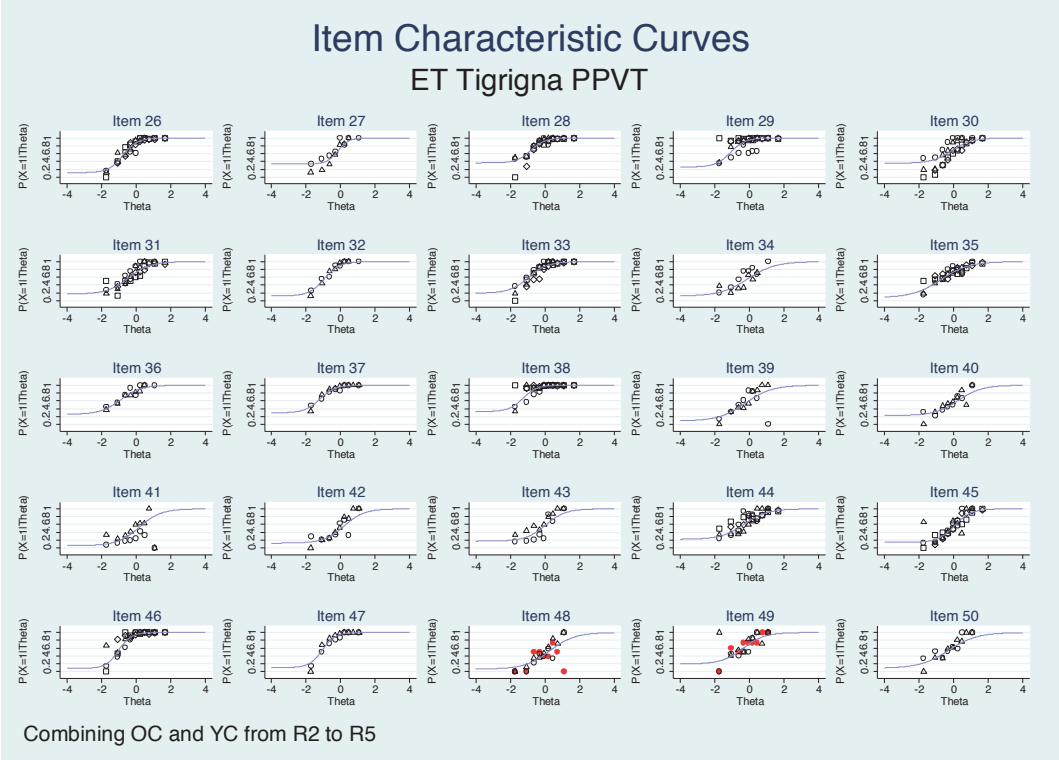

Item Characteristic Curves
ET Oromifa PPVT
Item 126
Item 127
Item 128
Item 129
Item 130
Item 131
Item 132
Item 133
Item 134
Item 135
Item 136
Item 137
Item 138
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

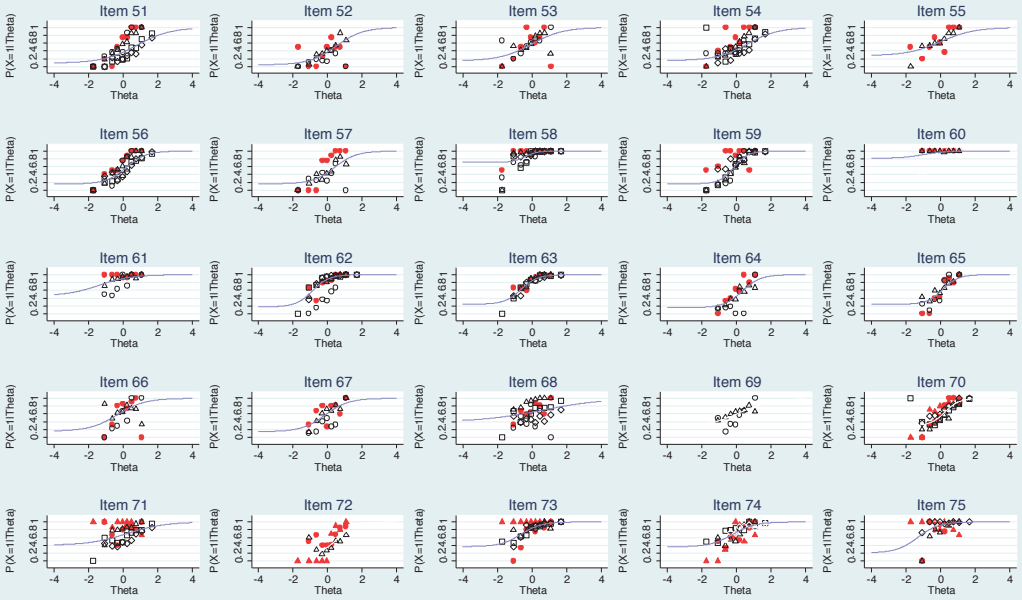
Figure 3. *Differential Item Functioning for PPVT analysis, Tigrigna*





Item Characteristic Curves

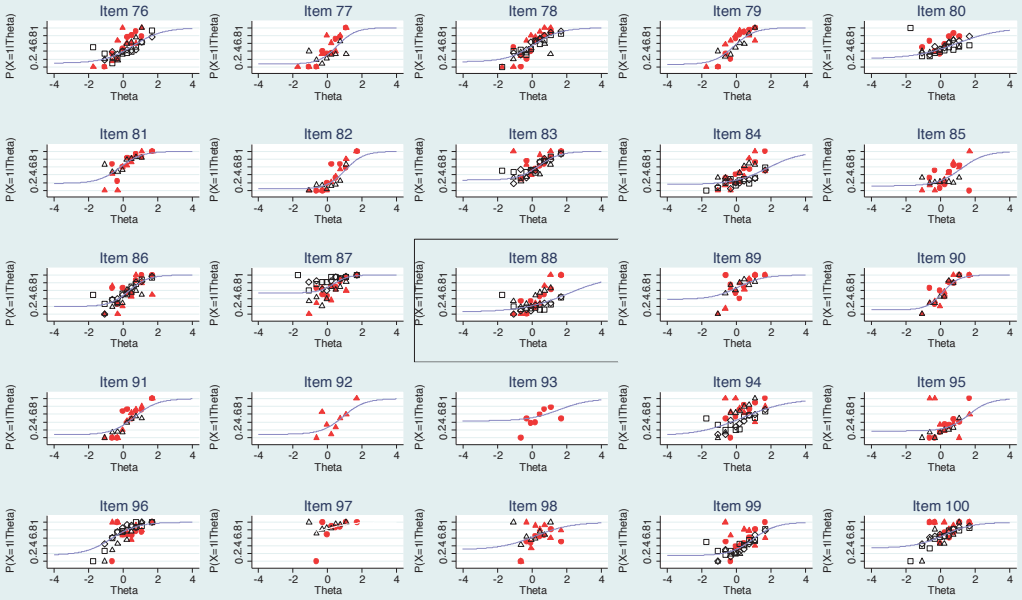
ET Tigrigna PPVT



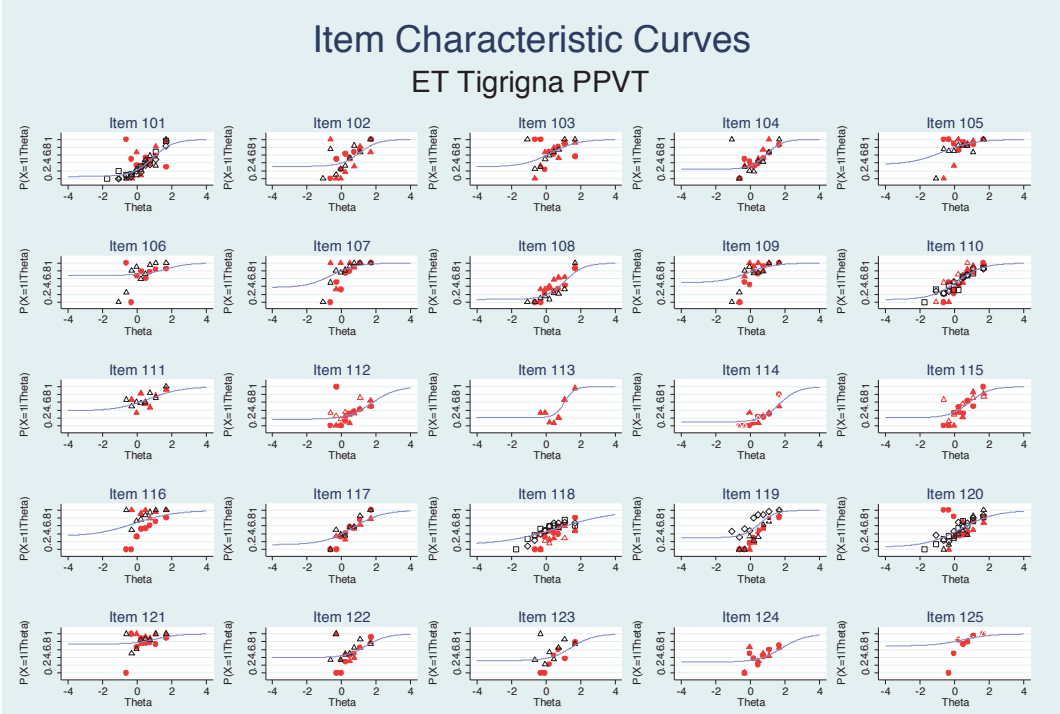
Combining OC and YC from R2 to R5

Item Characteristic Curves

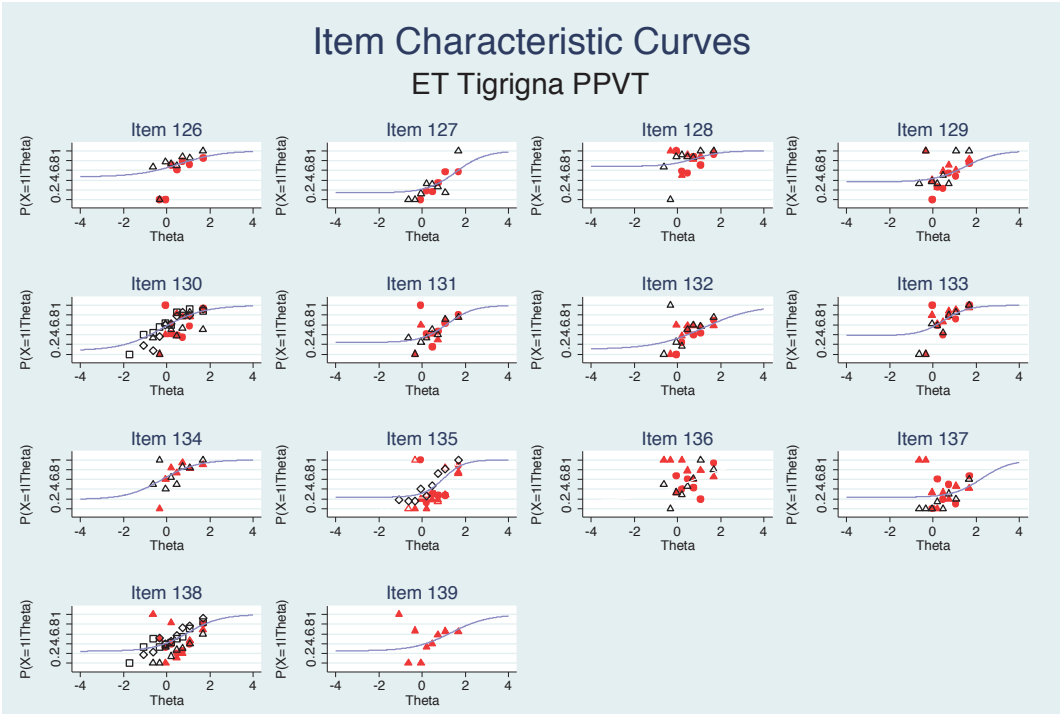
ET Tigrigna PPVT



Combining OC and YC from R2 to R5

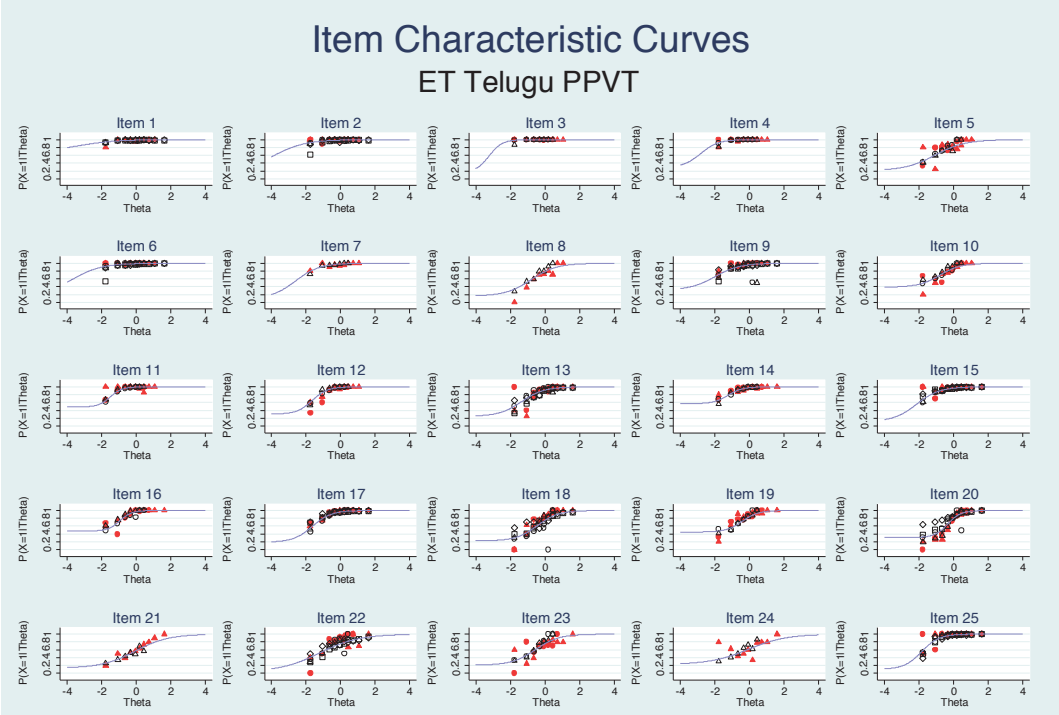


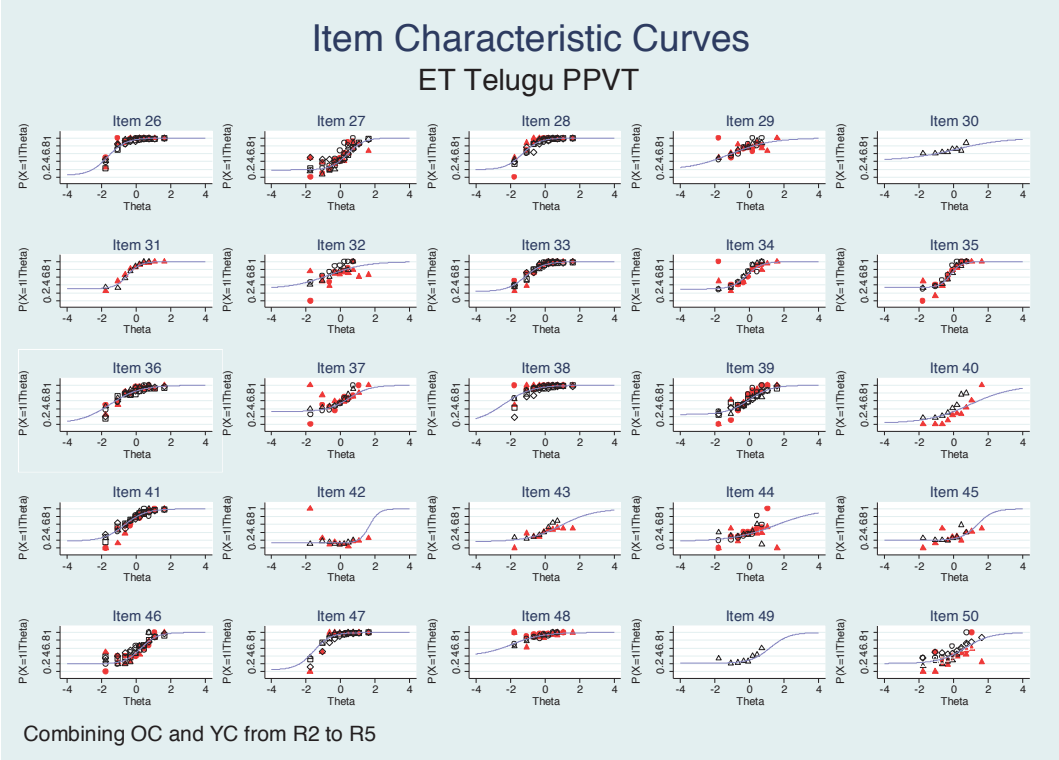
Combining OC and YC from R2 to R5

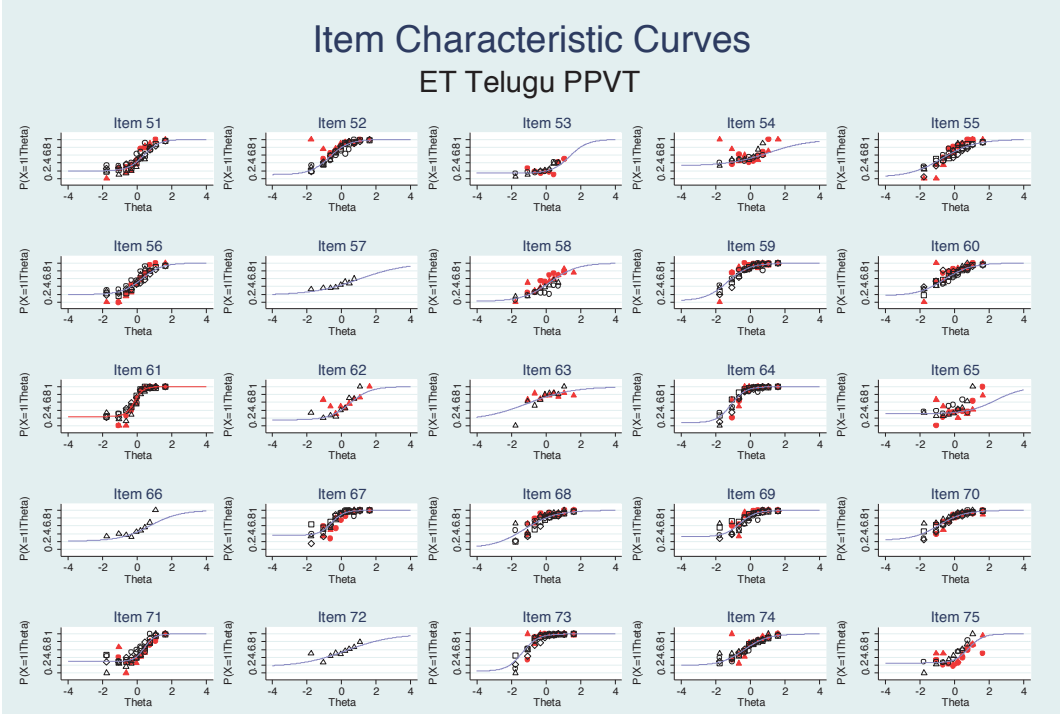


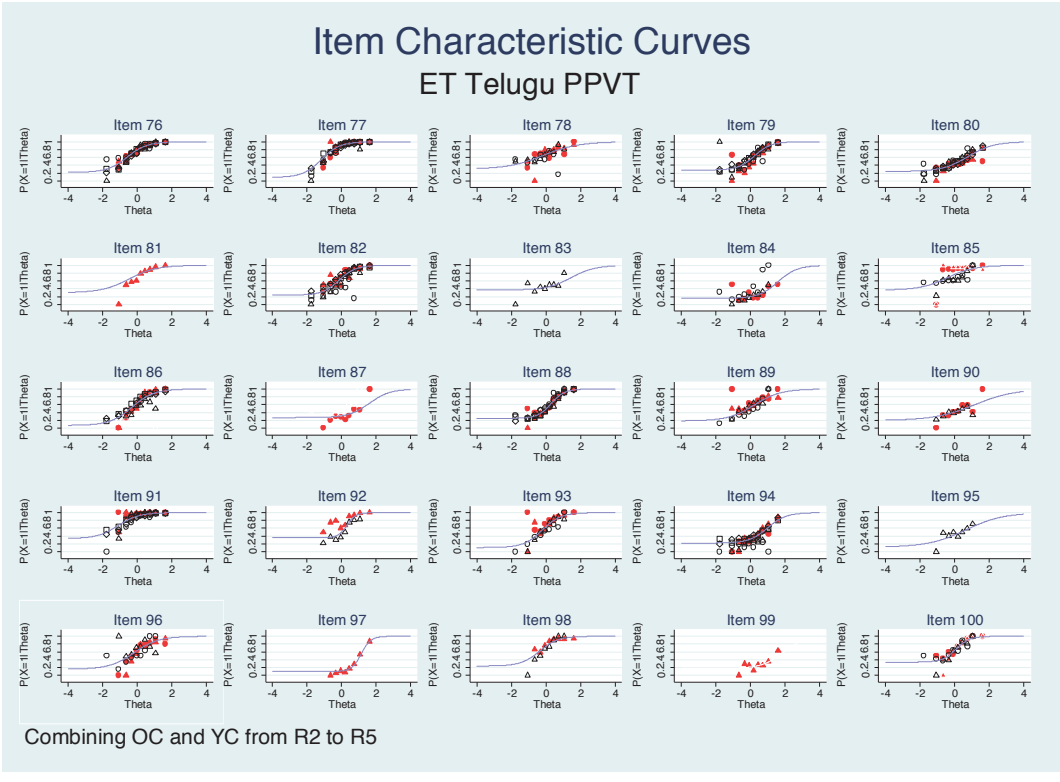
Combining OC and YC from R2 to R5

Figure 4. *Differential Item Functioning for PPVT analysis, Telugu*



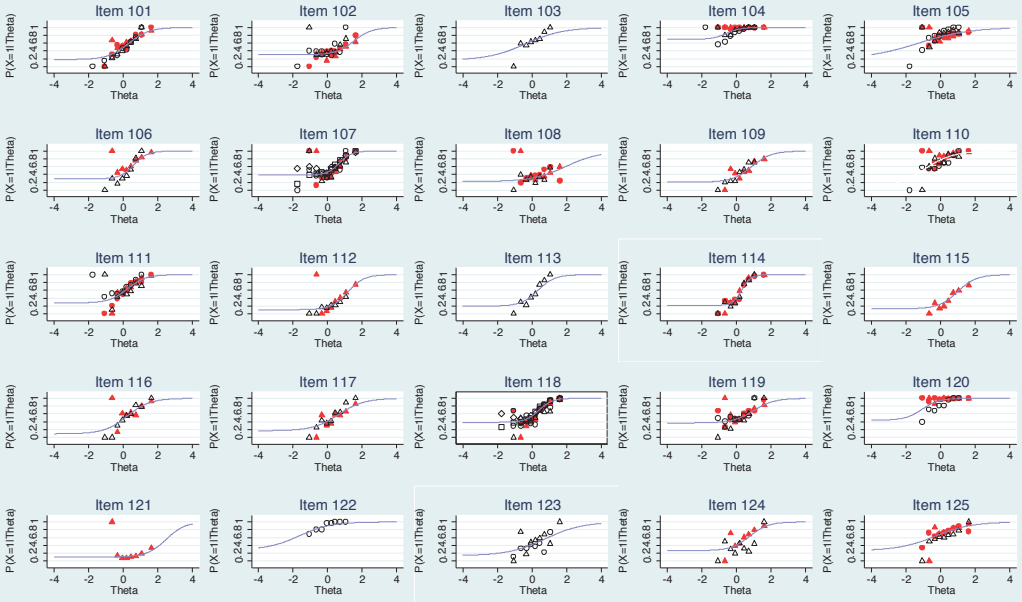



Item Characteristic Curves
ET Telugu PPVT
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
Item 65
Item 66
Item 67
Item 68
Item 69
Item 70
Item 71
Item 72
Item 73
Item 74
Item 75
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5


Item Characteristic Curves
ET Telugu PPVT
Item 76
Item 77
Item 78
Item 79
Item 80
Item 81
Item 82
Item 83
Item 84
Item 85
Item 86
Item 87
Item 88
Item 89
Item 90
Item 91
Item 92
Item 93
Item 94
Item 95
Item 96
Item 97
Item 98
Item 99
Item 100
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5

Item Characteristic Curves

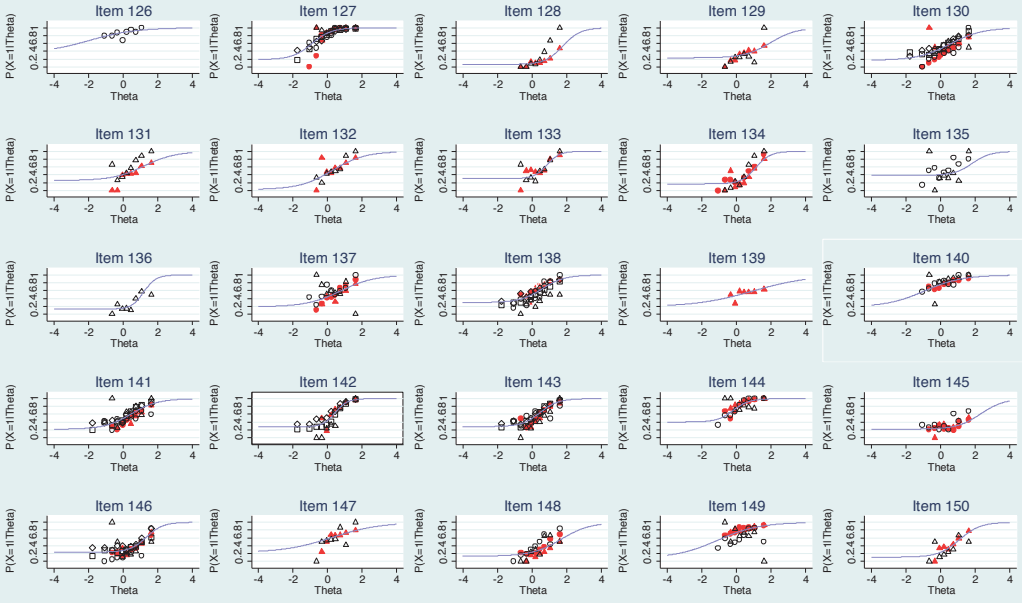
ET Telugu PPVT



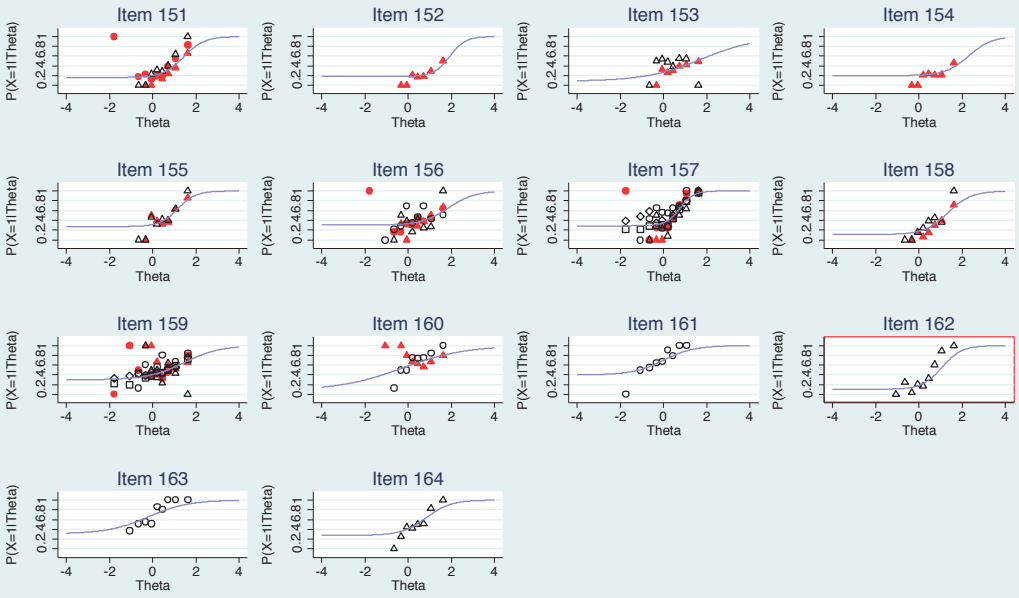
Combining OC and YC from R2 to R5

Item Characteristic Curves

ET Telugu PPVT

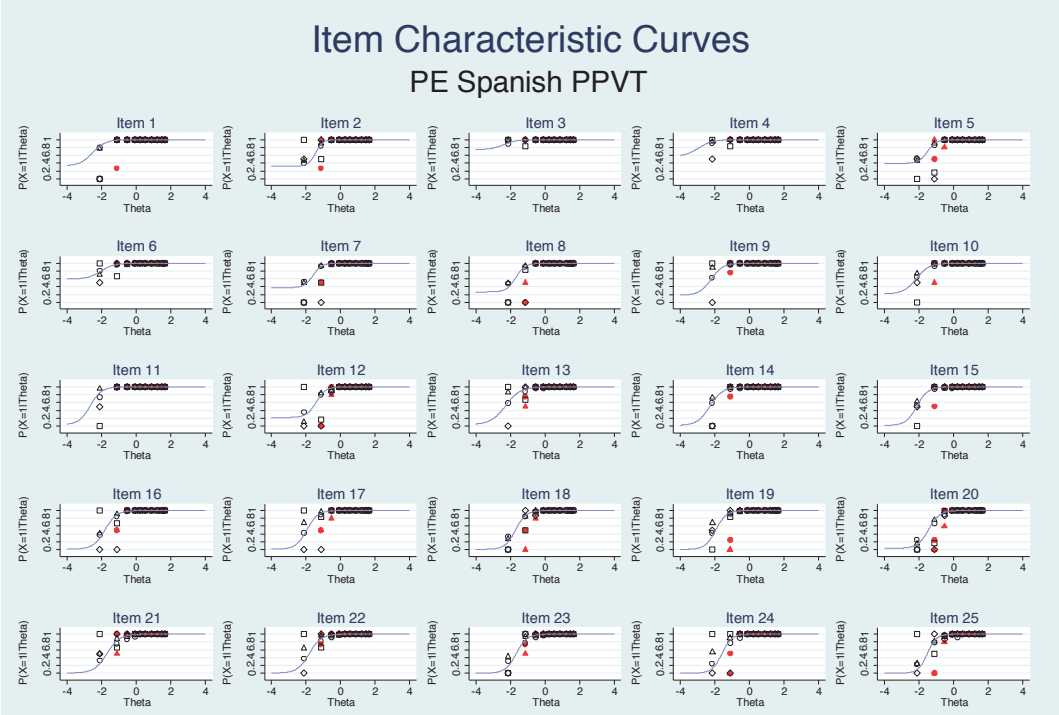


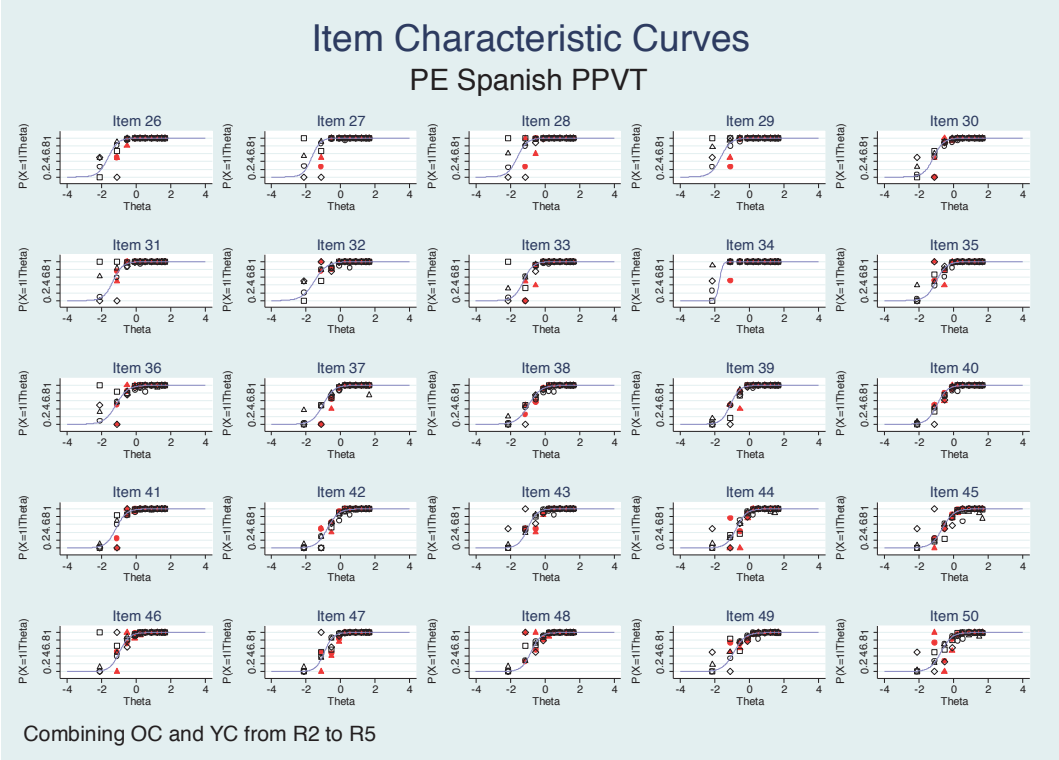
Combining OC and YC from R2 to R5

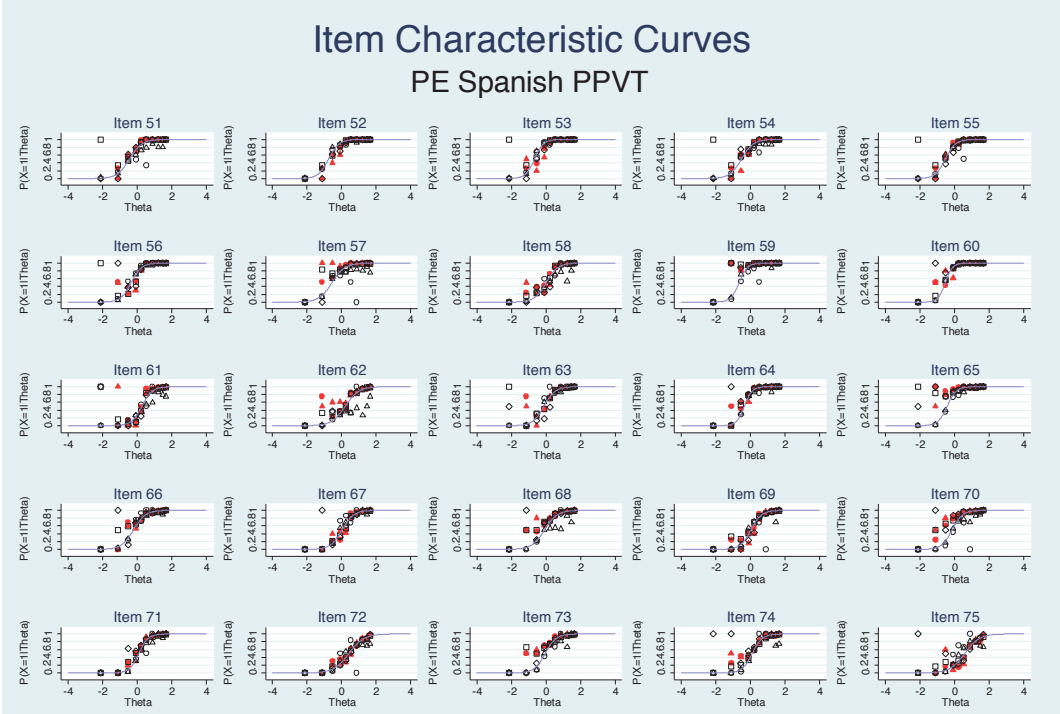


Combining OC and YC from R2 to R5

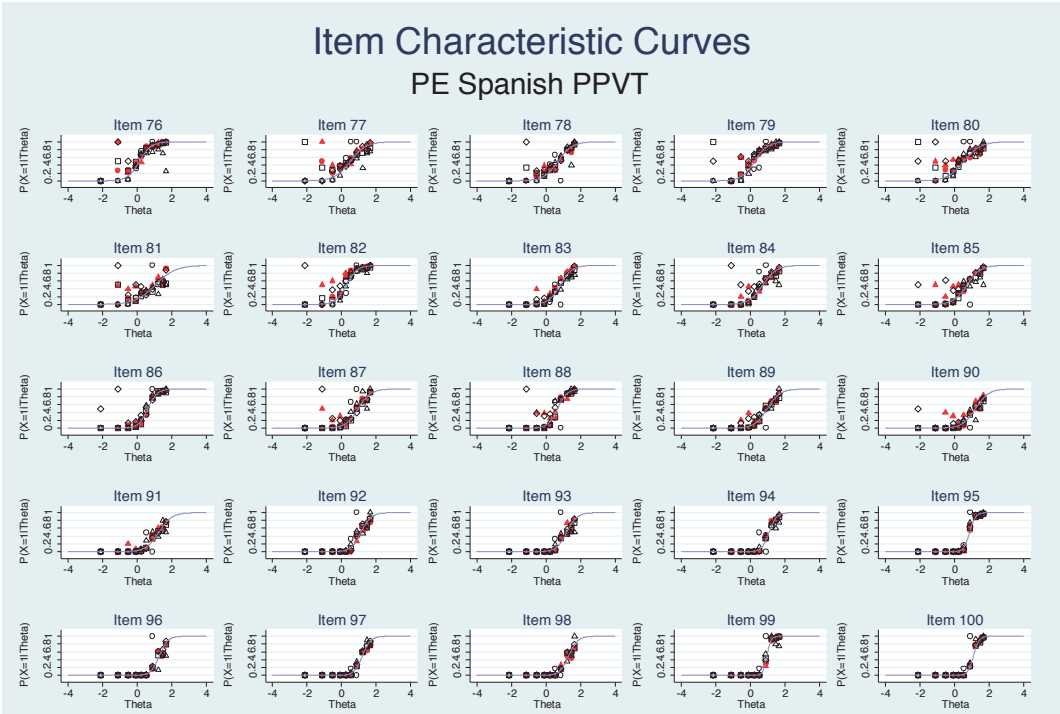
Figure 5. *Differential Item Functioning for PPVT analysis, Spanish*







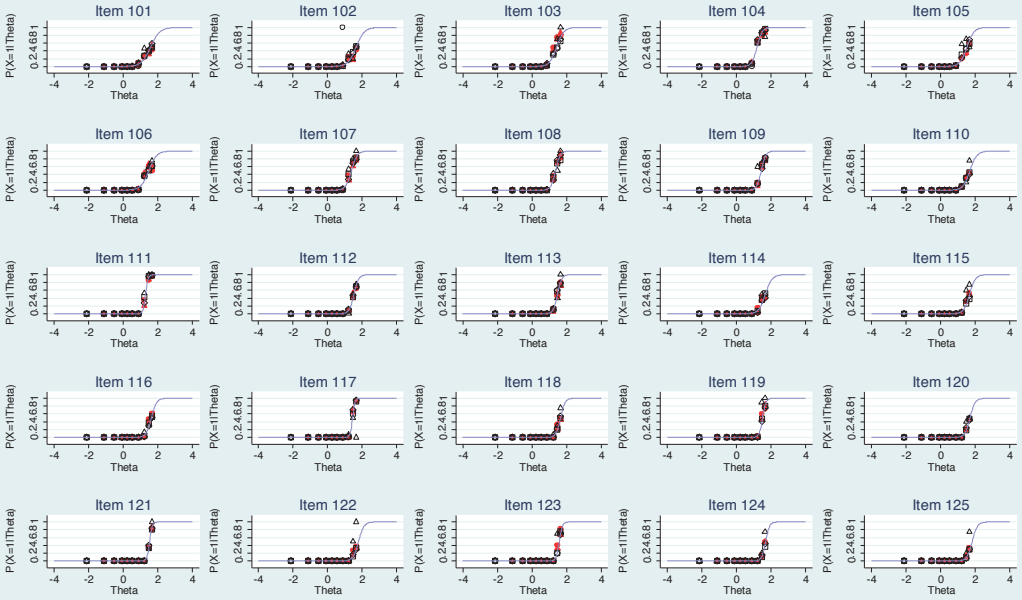
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

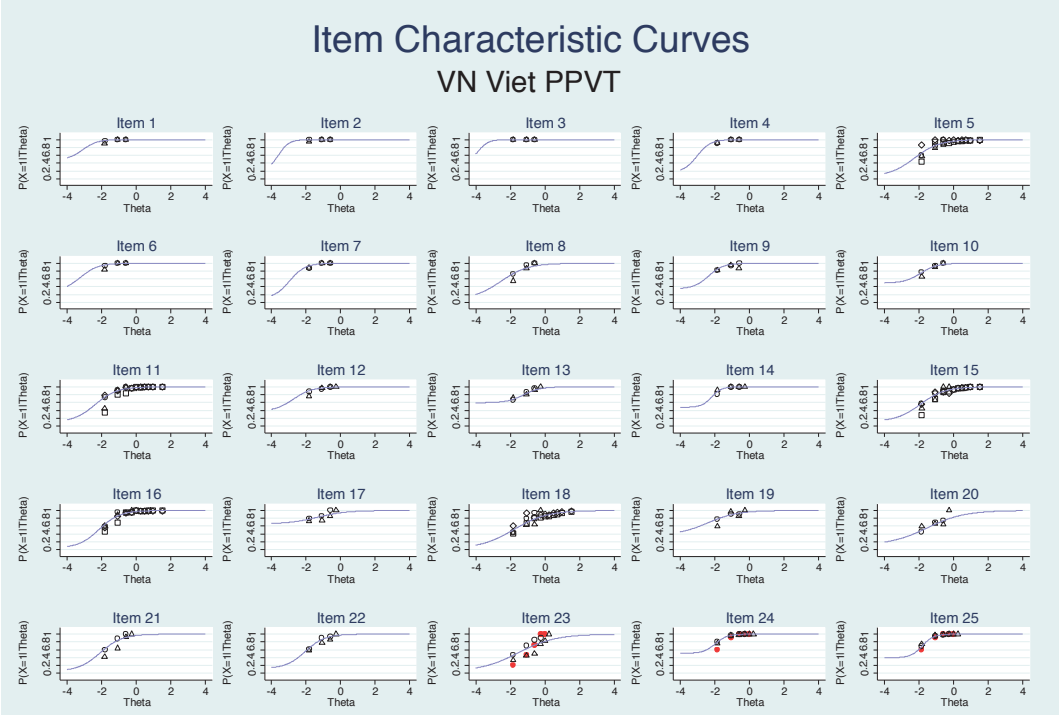
Item Characteristic Curves

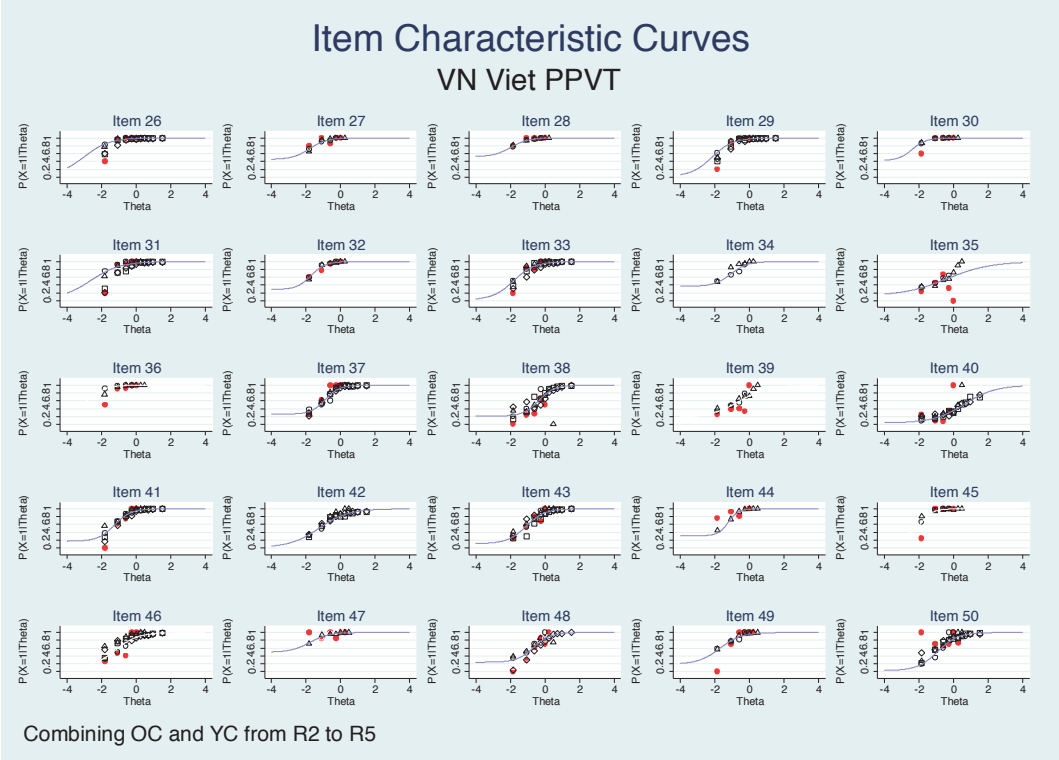
PE Spanish PPVT

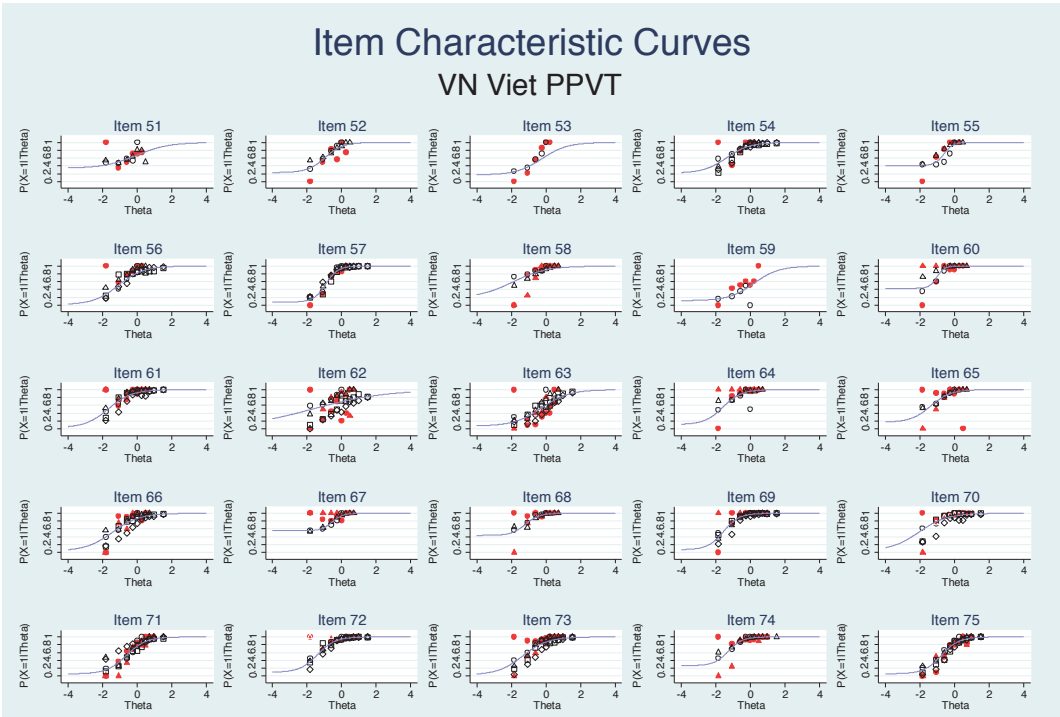


Combining OC and YC from R2 to R5

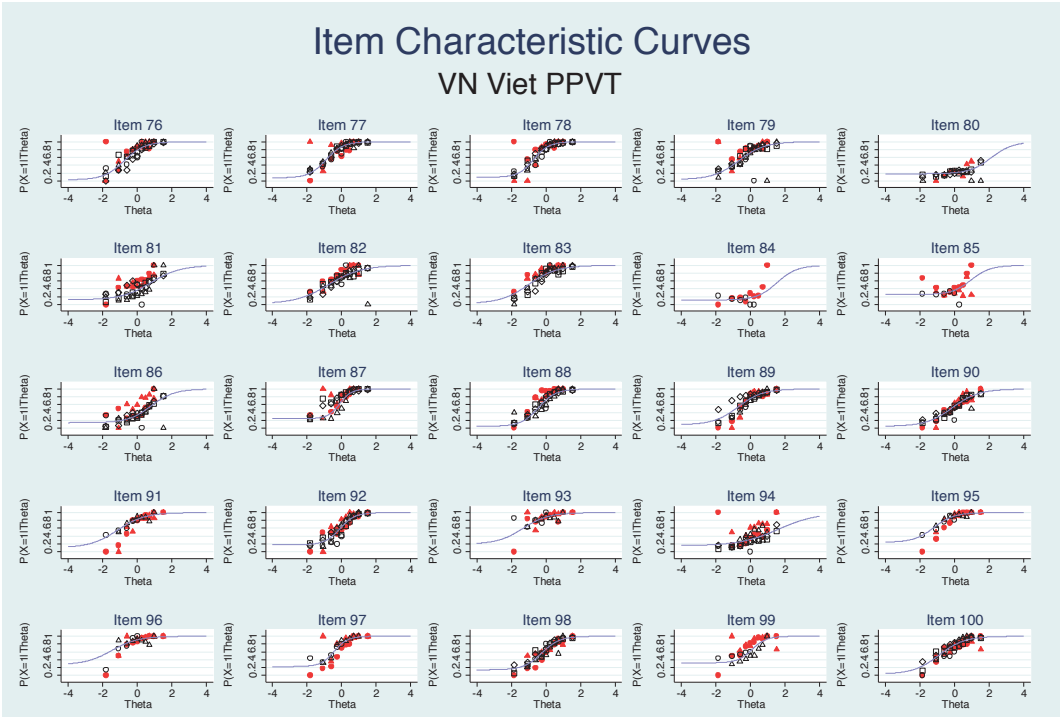
Figure 6. *Differential Item Functioning for PPVT analysis, Vietnamese*



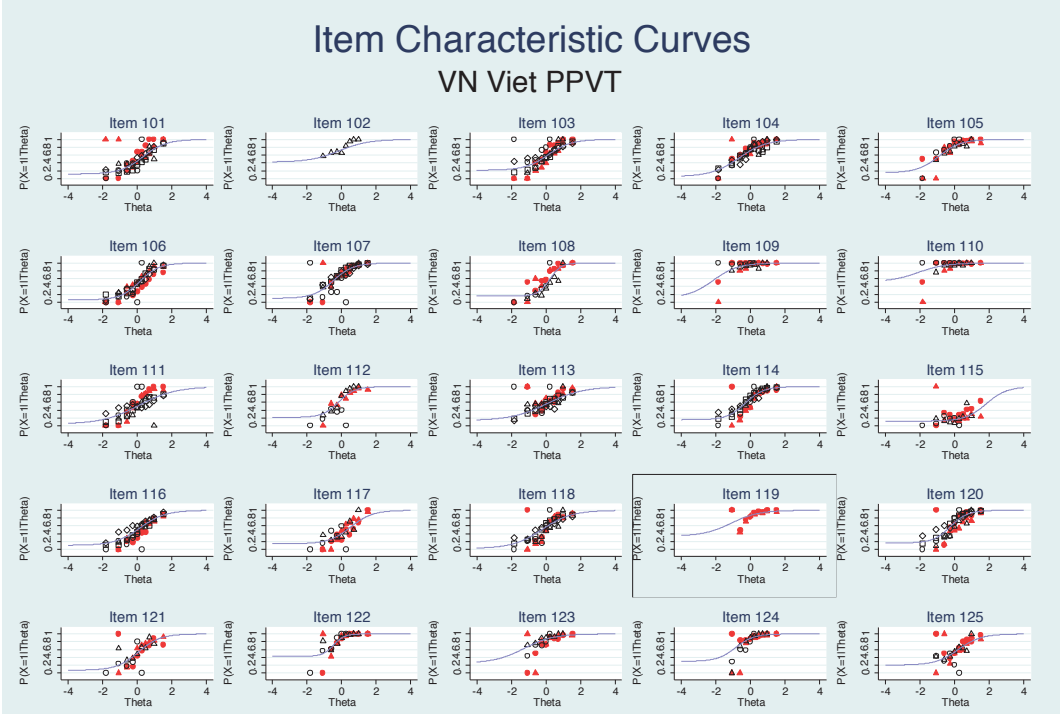




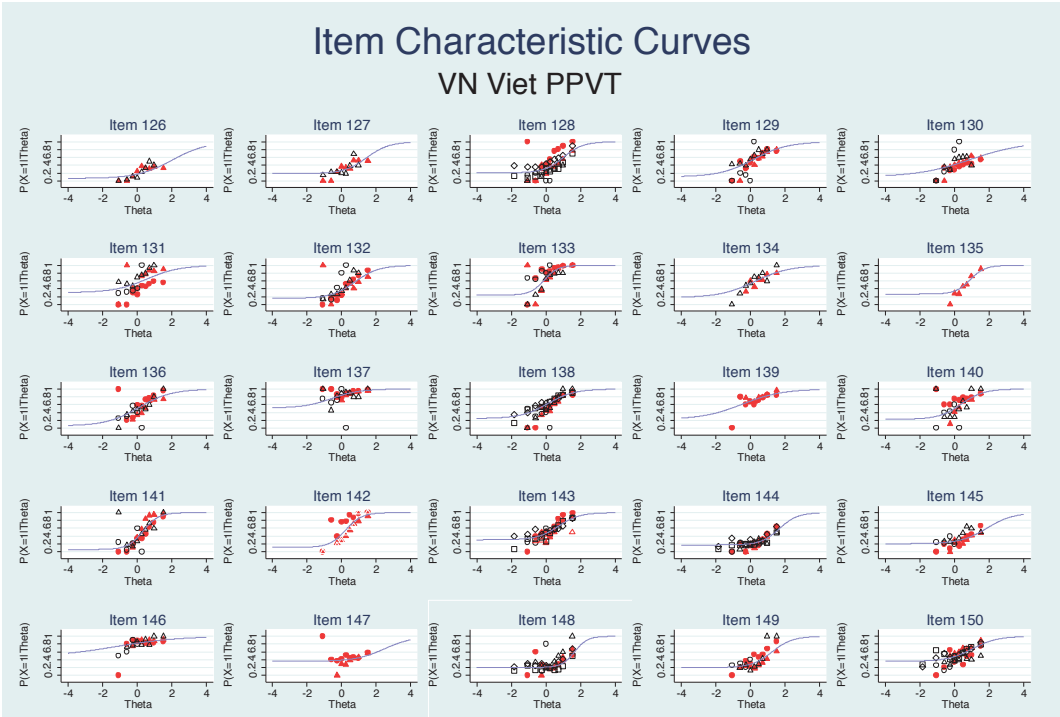
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5



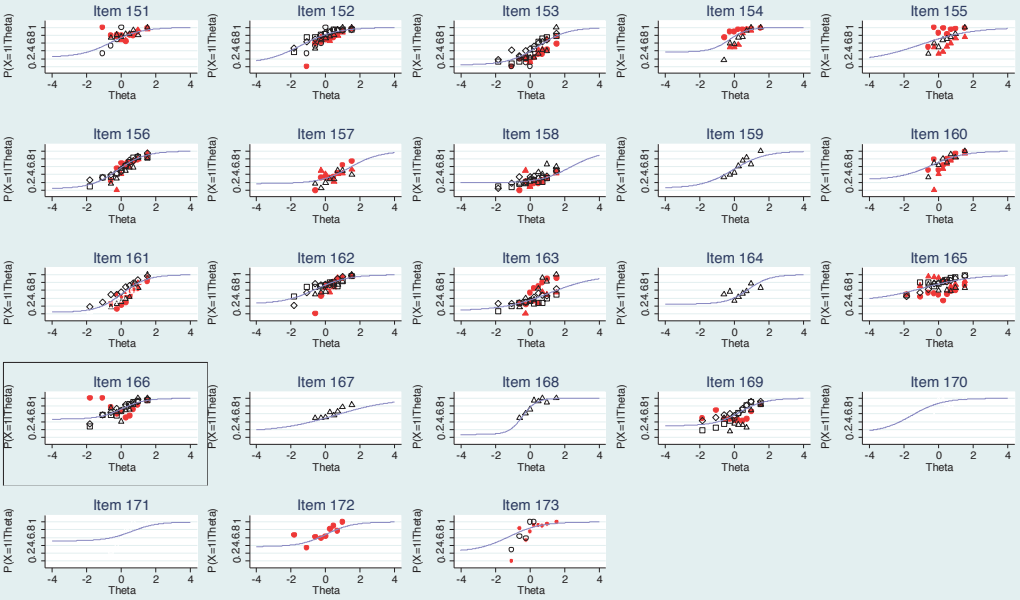
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

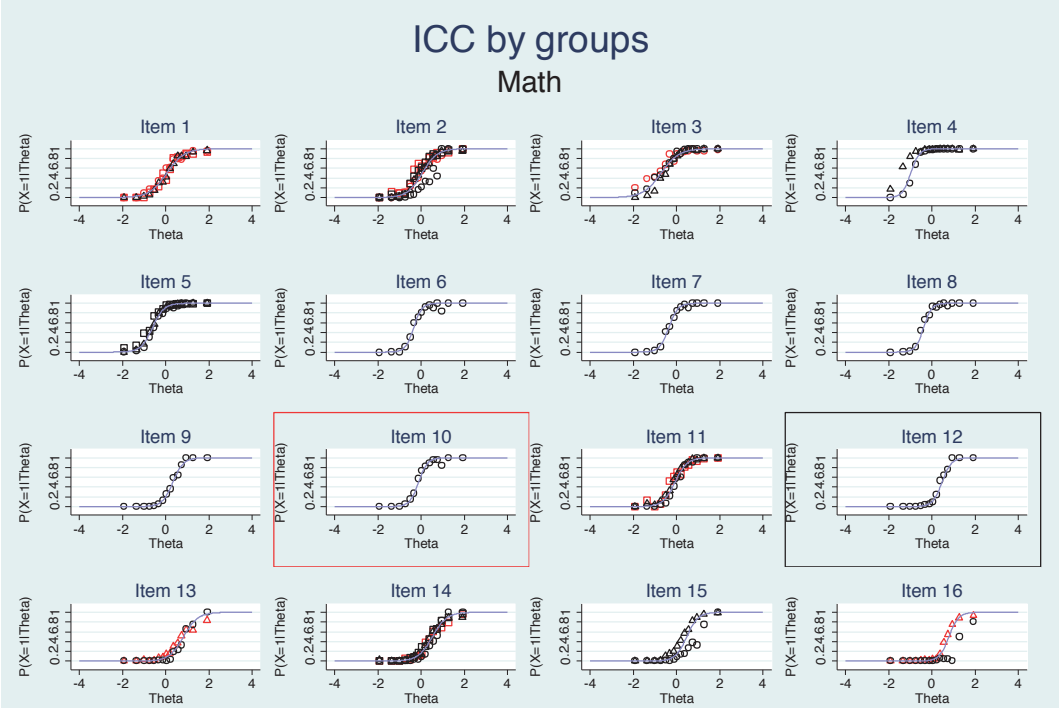
Item Characteristic Curves

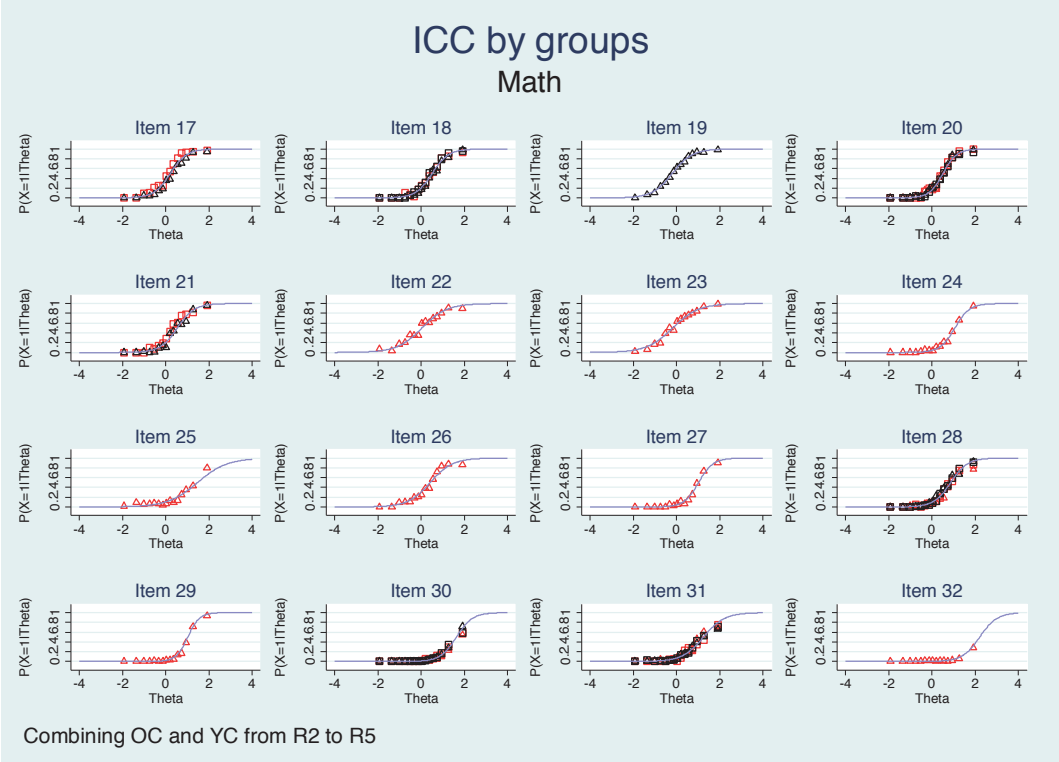
VN Viet PPVT

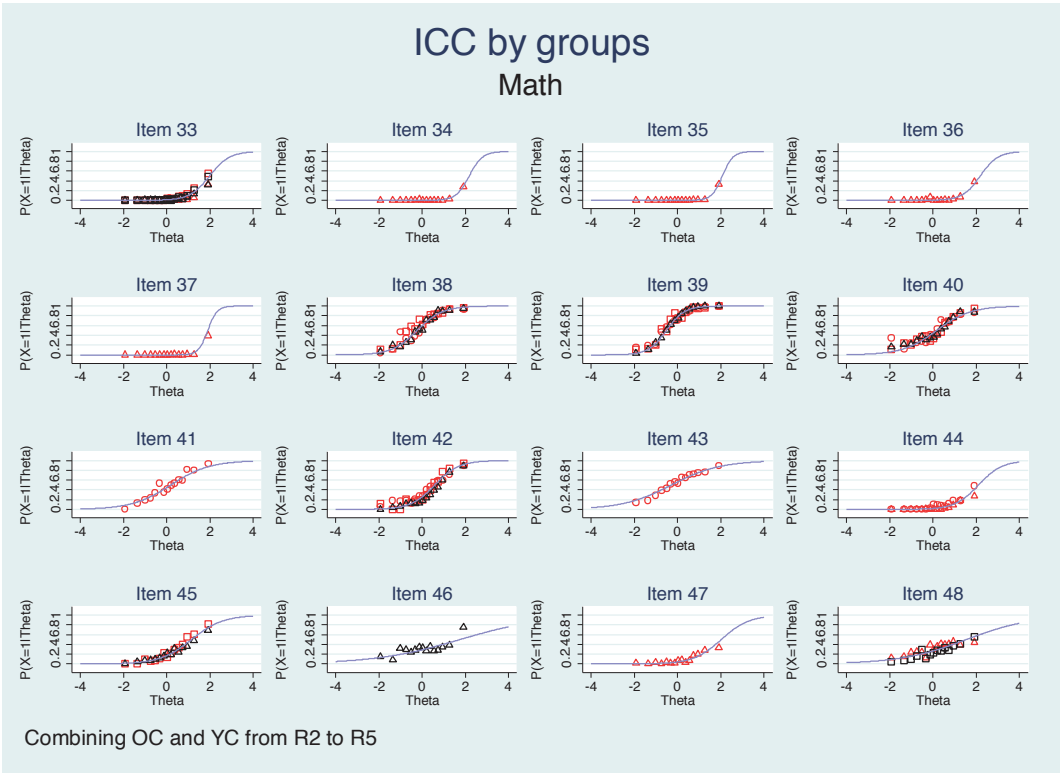


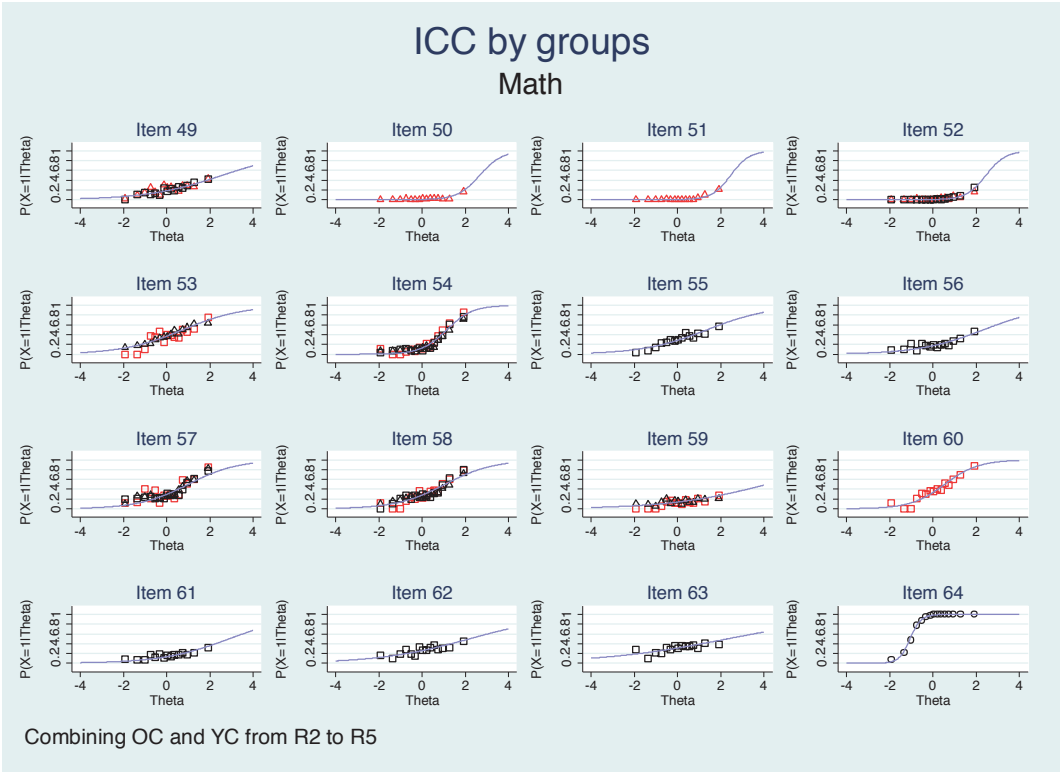
Combining OC and YC from R2 to R5

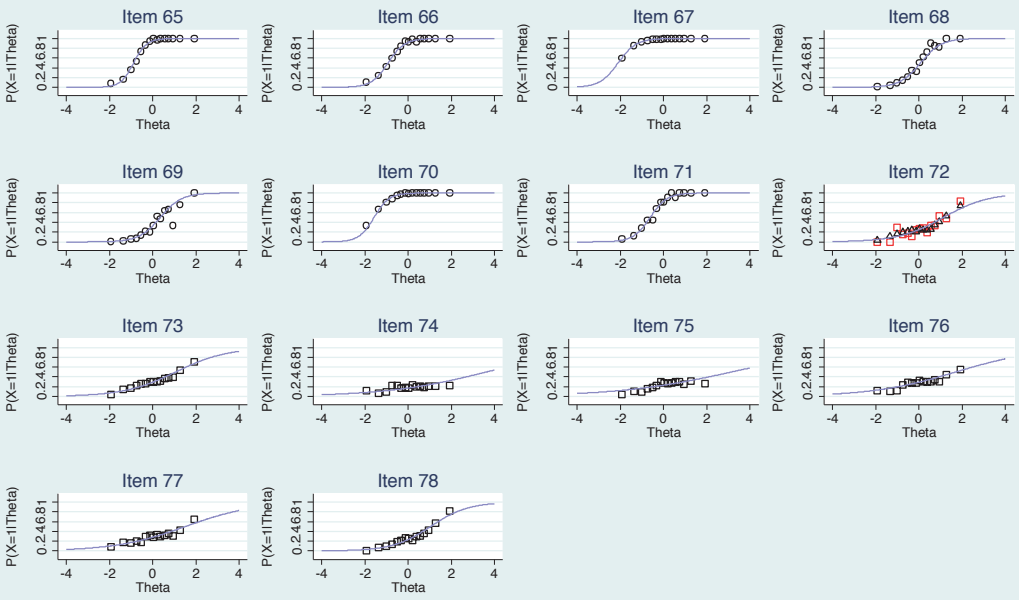
Figure 7. *Differential Item Functioning for maths analysis, Ethiopia*





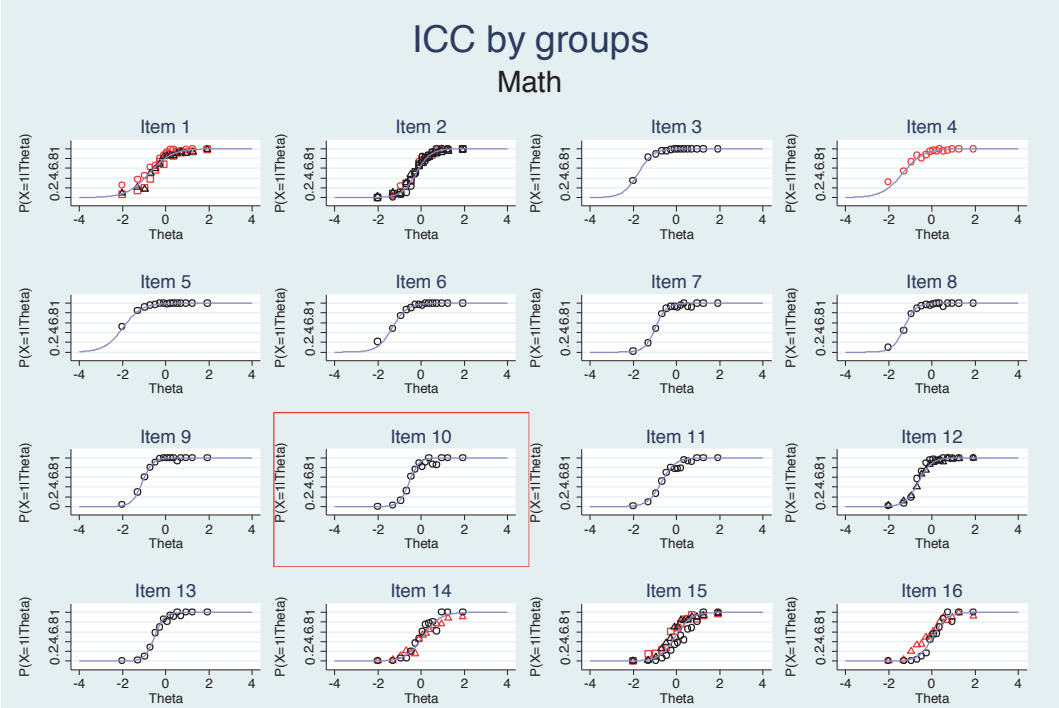

ICC by groups
Math
Item 33
Item 34
Item 35
Item 36
Item 37
Item 38
Item 39
Item 40
Item 41
Item 42
Item 43
Item 44
Item 45
Item 46
Item 47
Item 48
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5

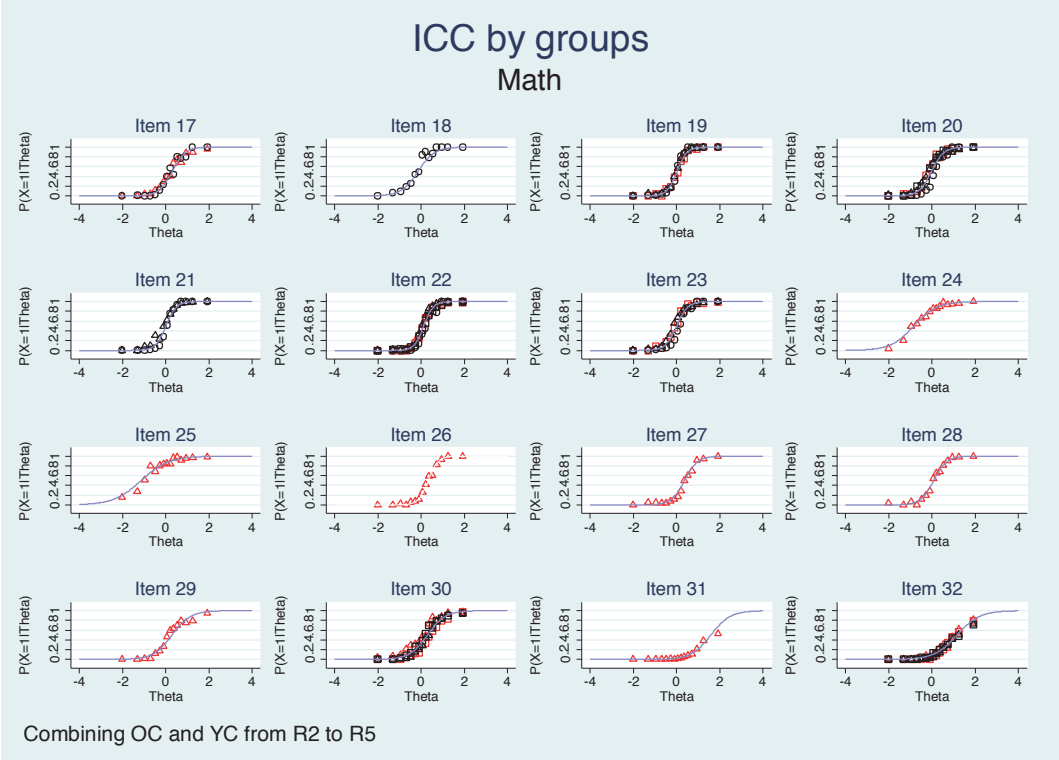

ICC by groups
Math
Item 49
Item 50
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5

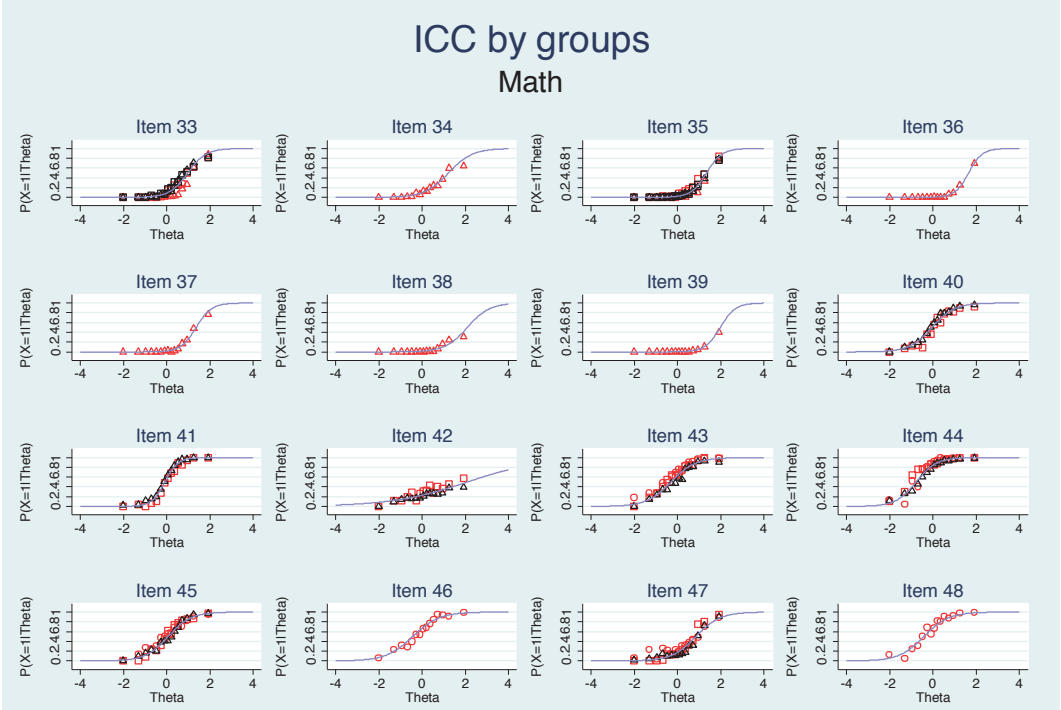


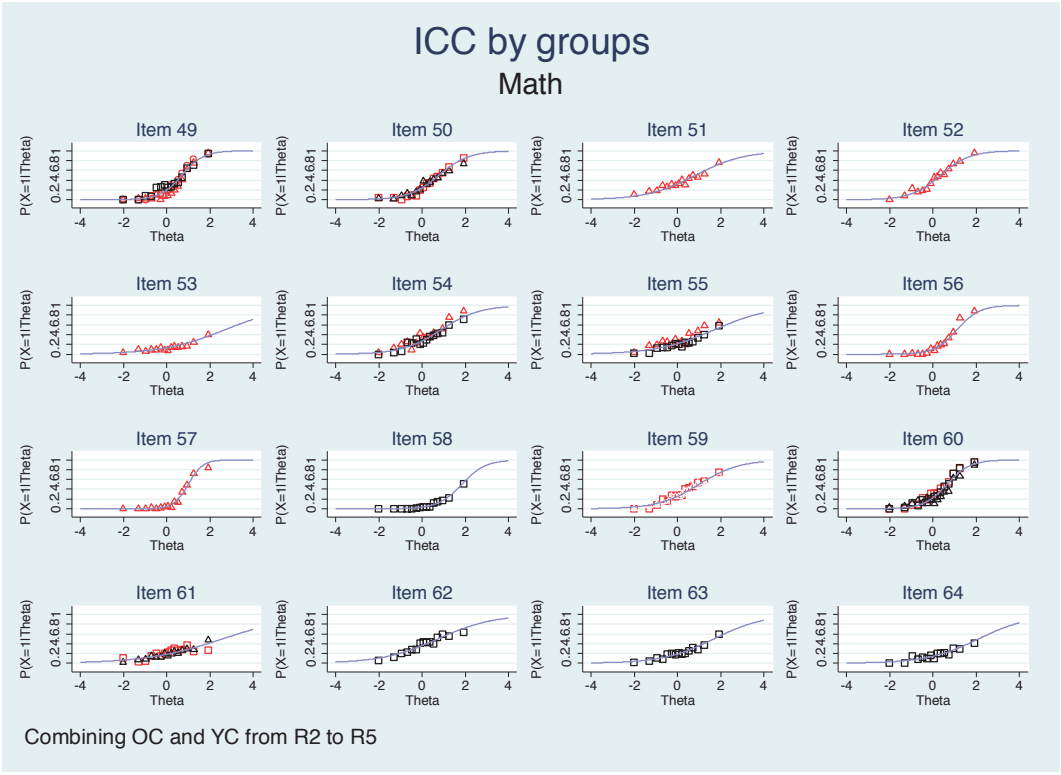
Combining OC and YC from R2 to R5

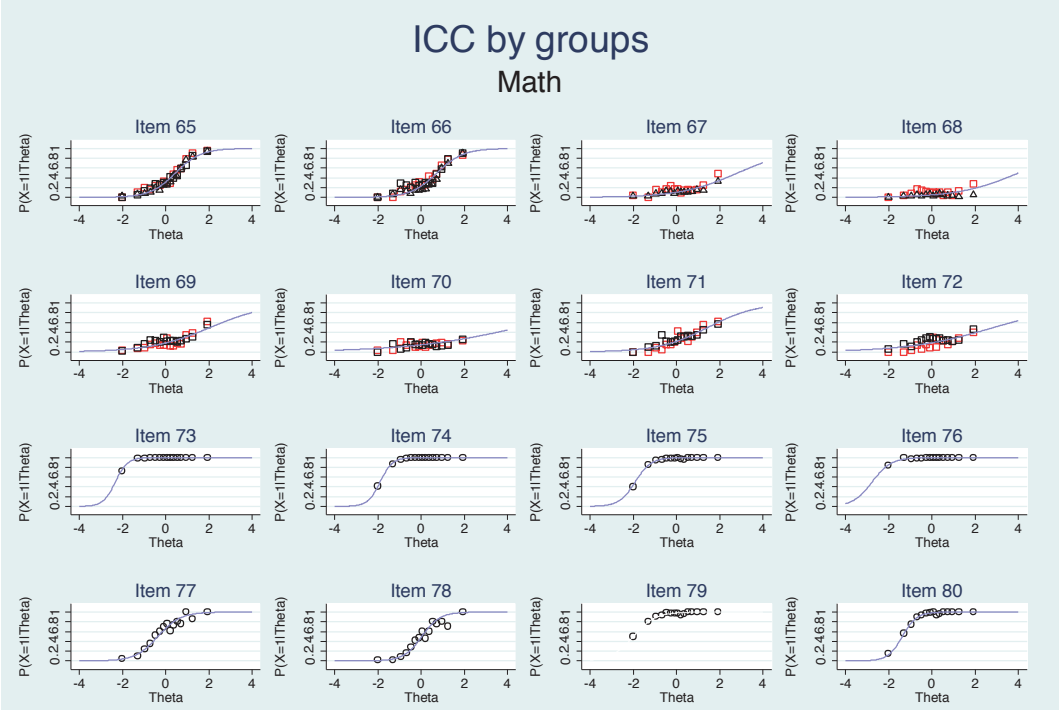
Figure 8. *Differential Item Functioning for maths analysis, India*



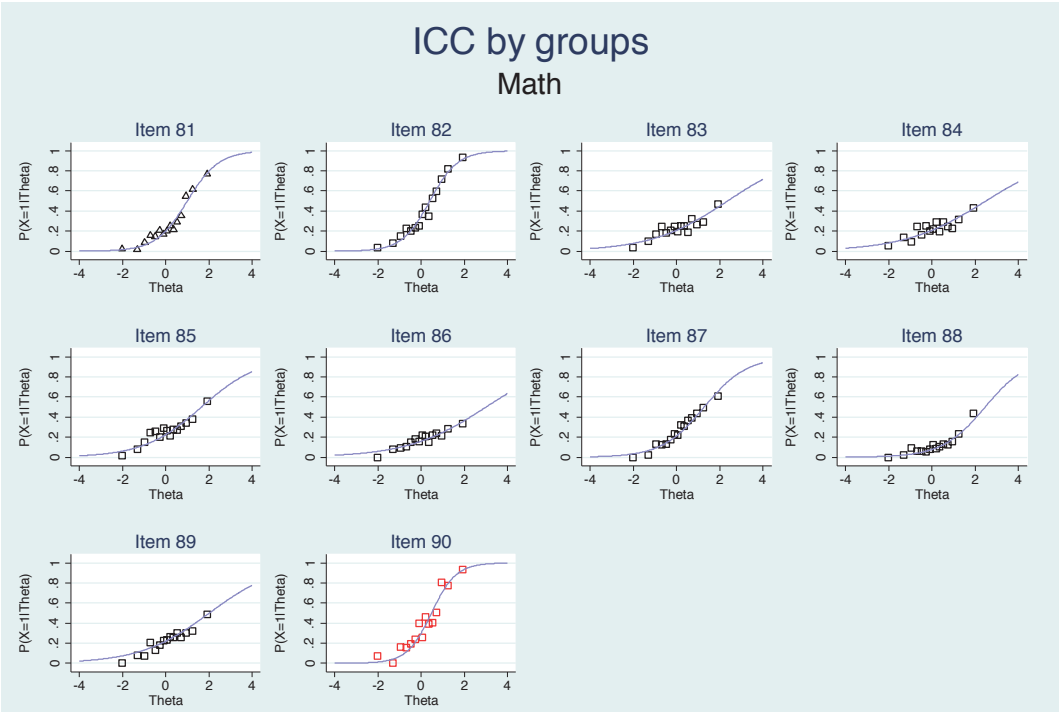



ICC by groups
Math
Item 33
Item 34
Item 35
Item 36
Item 37
Item 38
Item 39
Item 40
Item 41
Item 42
Item 43
Item 44
Item 45
Item 46
Item 47
Item 48
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5


ICC by groups
Math
Item 49
Item 50
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5

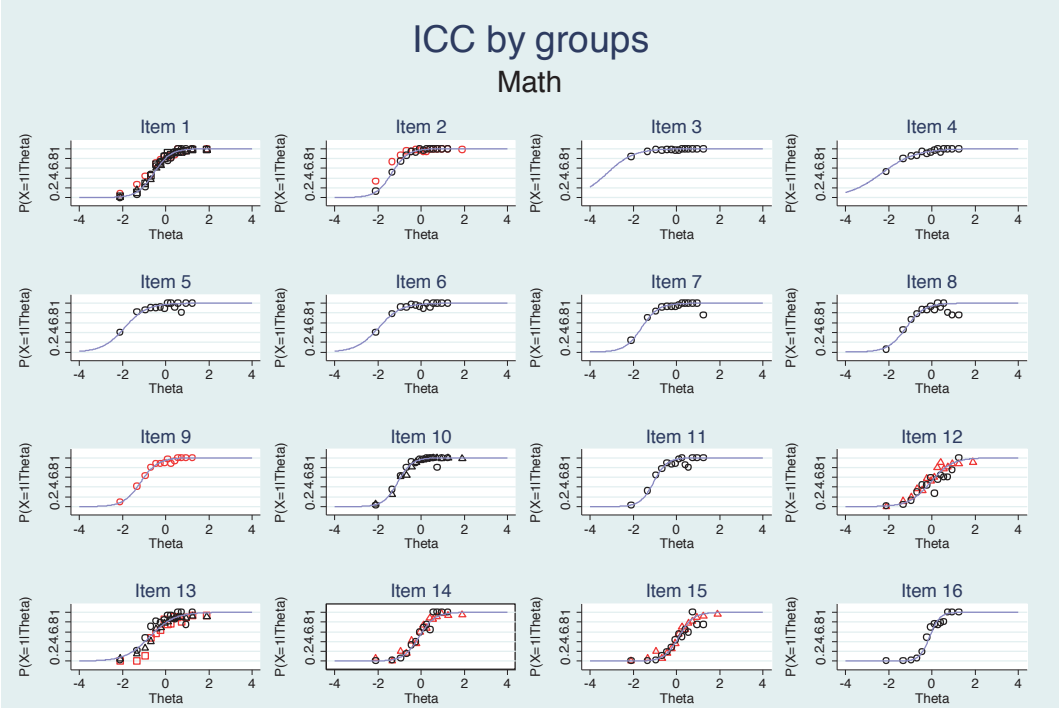


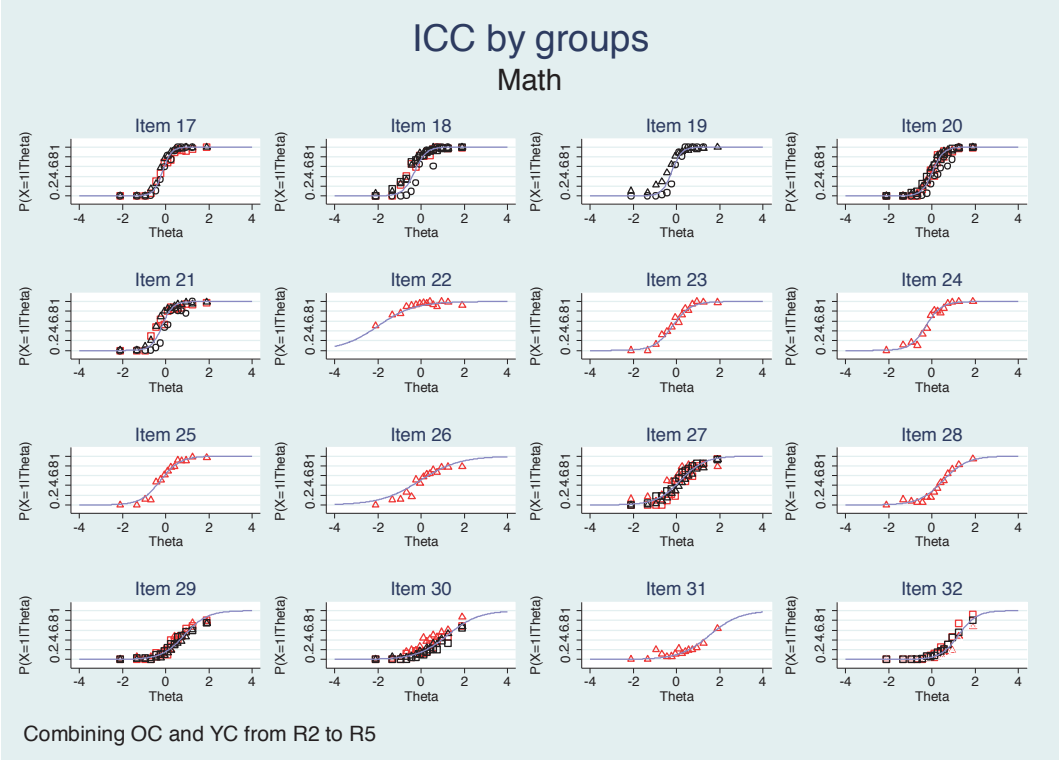
Combining OC and YC from R2 to R5

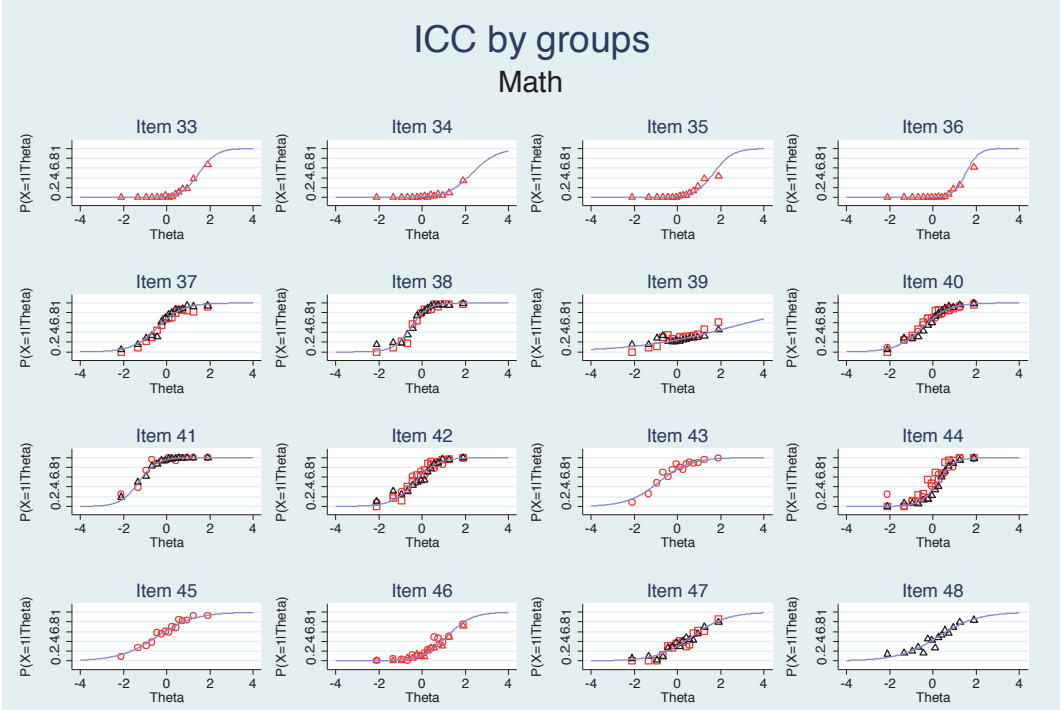


Combining OC and YC from R2 to R5

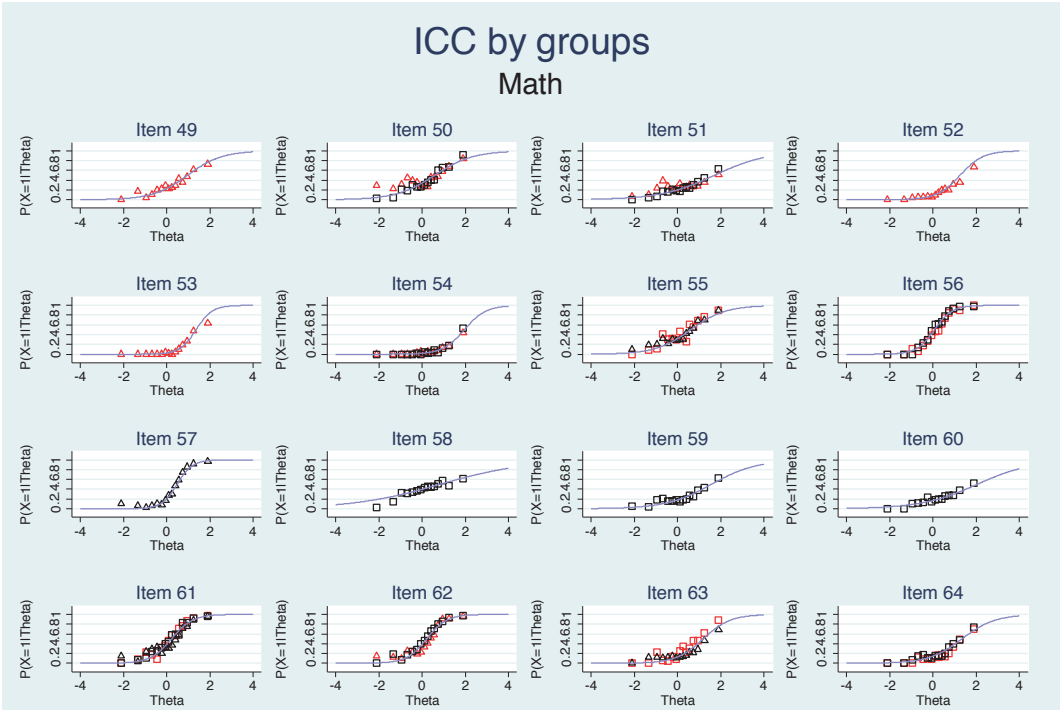
Figure 9. *Differential Item Functioning for maths analysis, Peru*



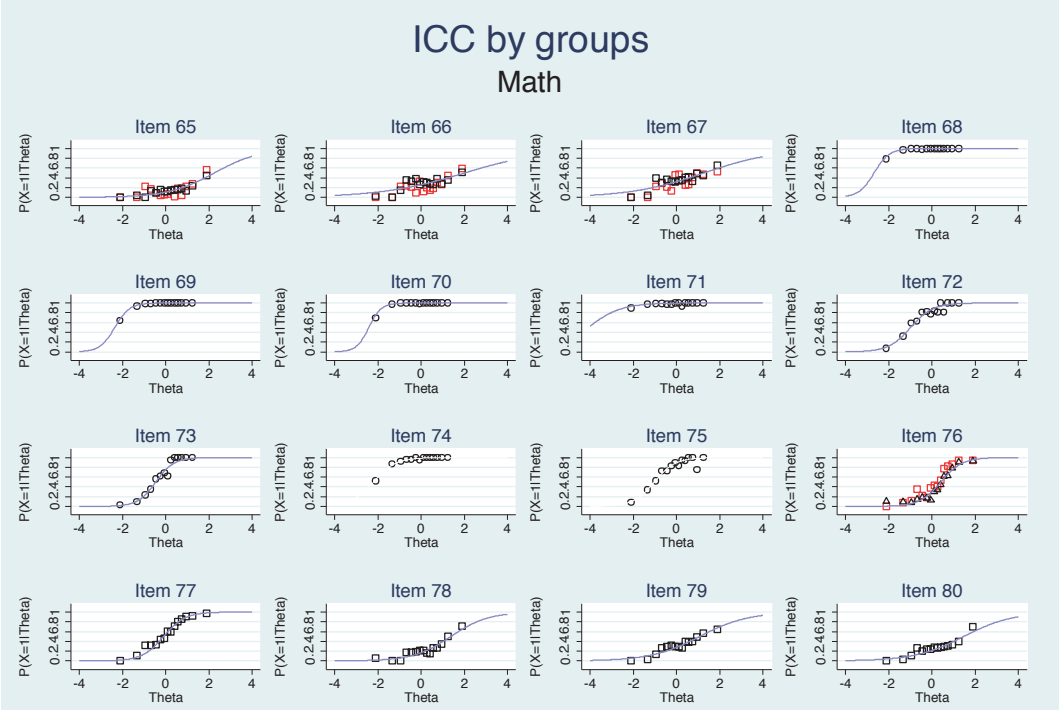




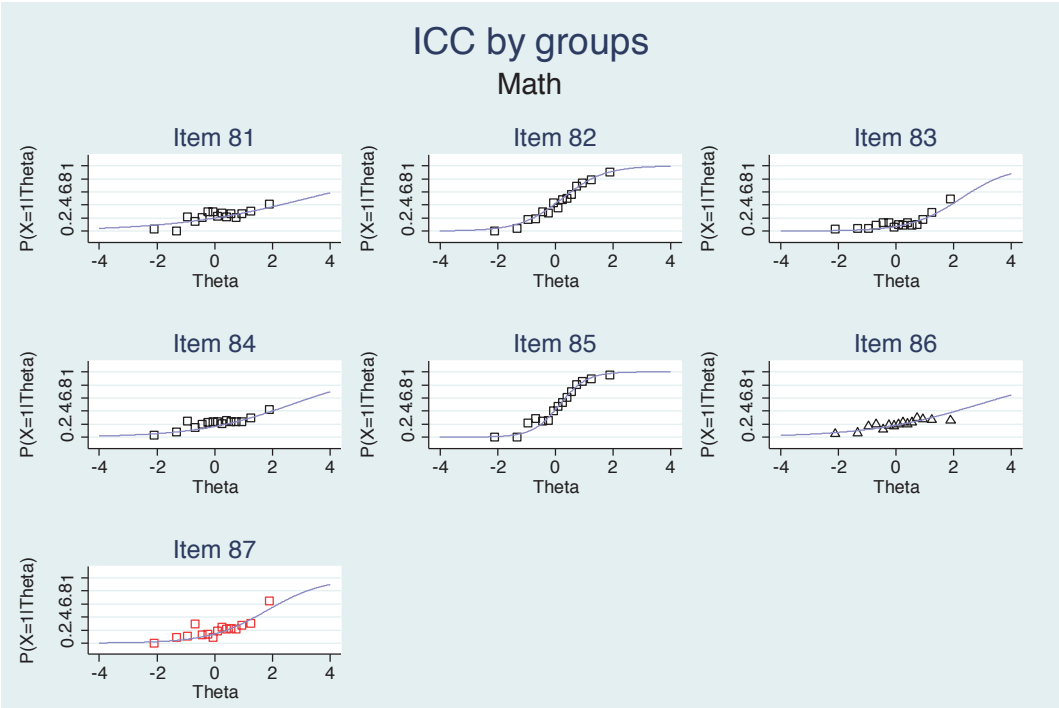
Combining OC and YC from R2 to R5



Combining OC and YC from R2 to R5

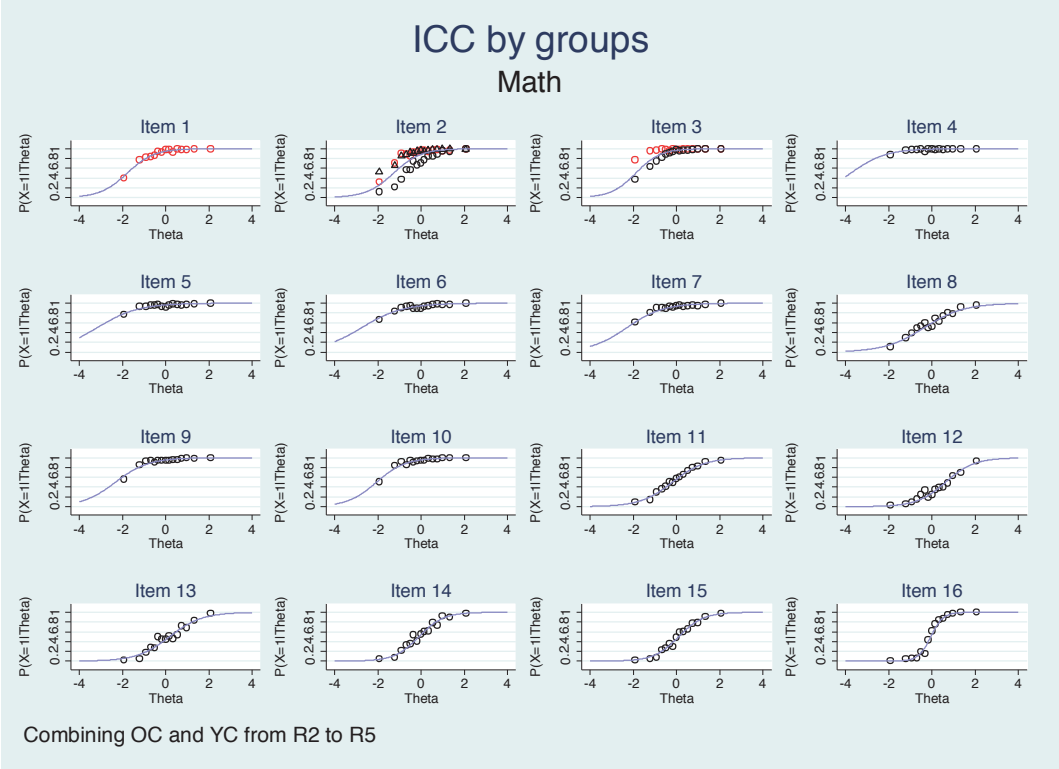


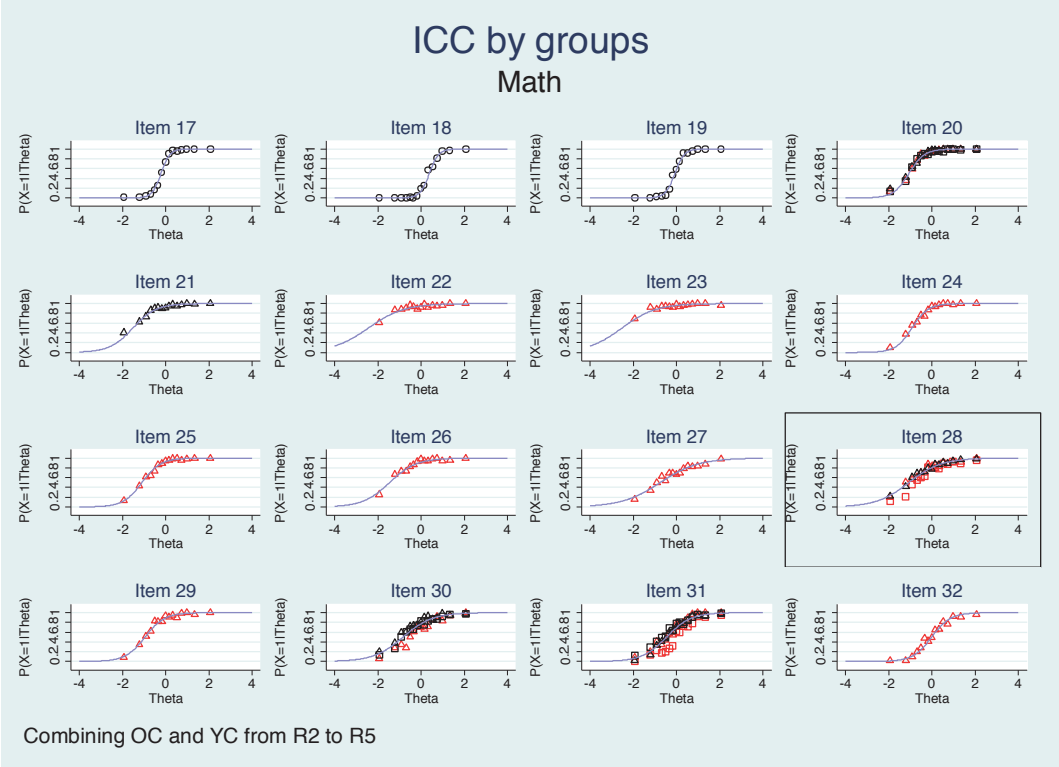
Combining OC and YC from R2 to R5

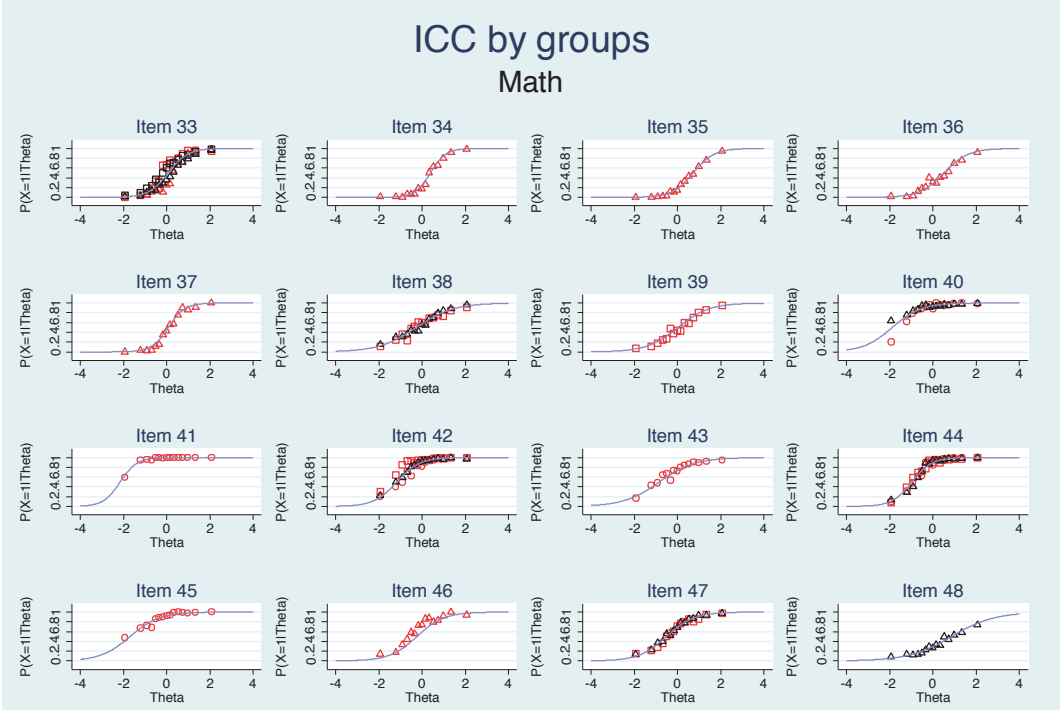


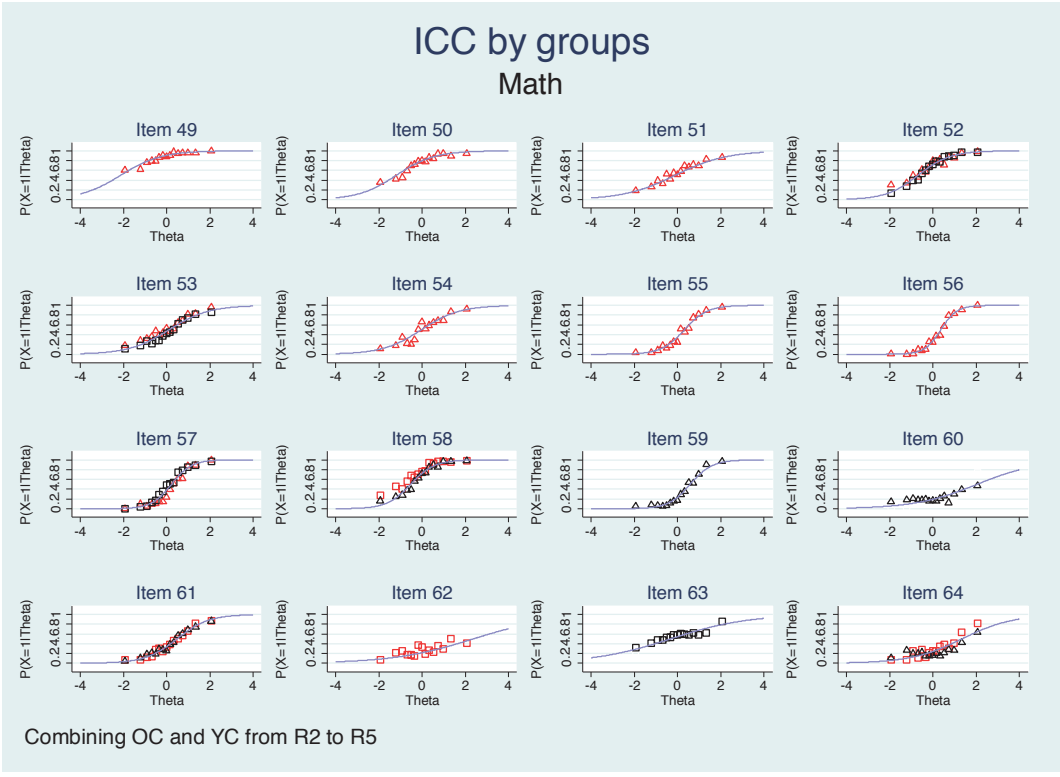
Combining OC and YC from R2 to R5

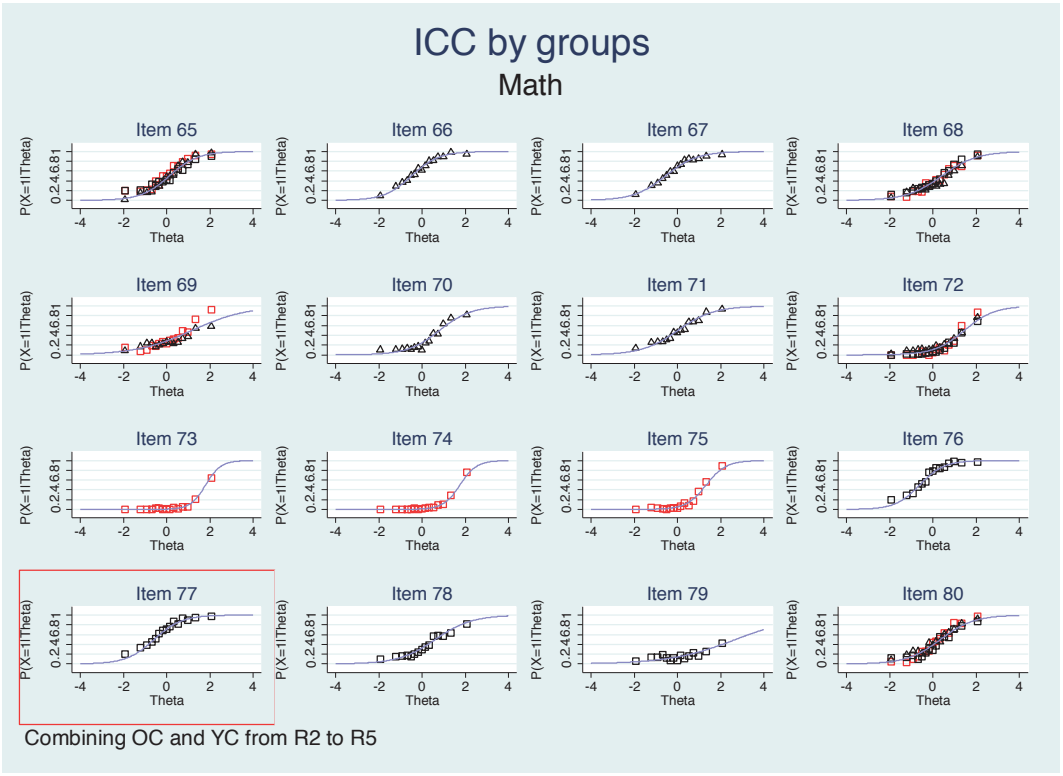
Figure 10. *Differential Item Functioning for maths analysis, Vietnam*

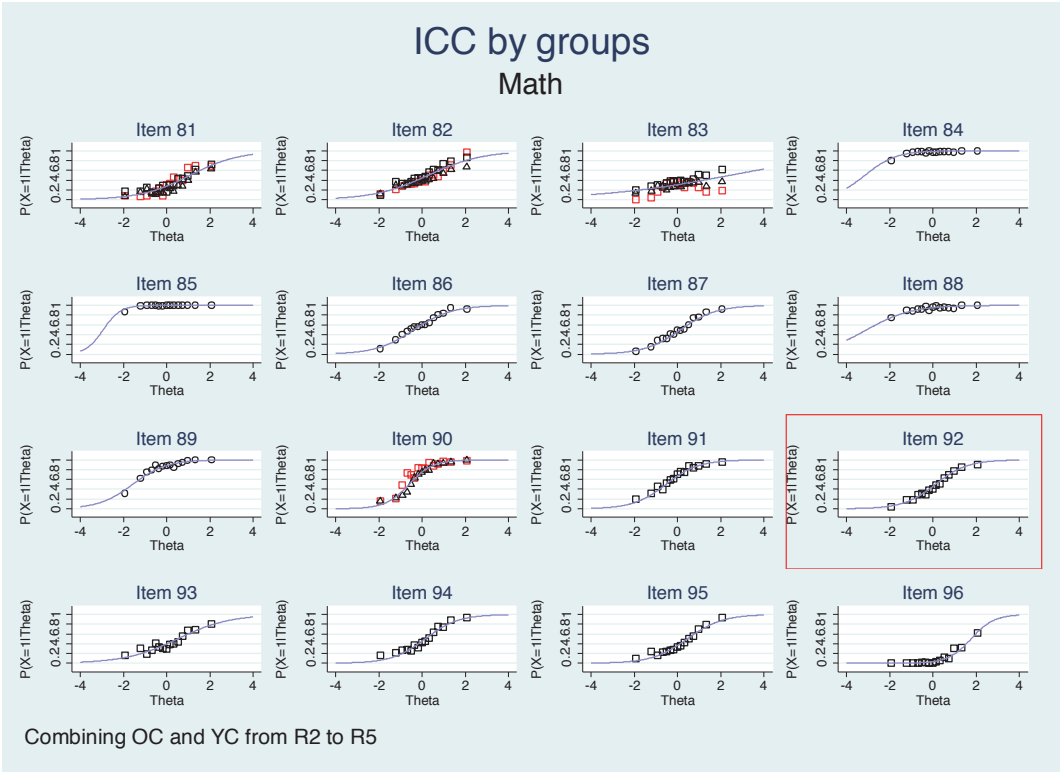





ICC by groups
Math
Item 33
Item 34
Item 35
Item 36
Item 37
Item 38
Item 39
Item 40
Item 41
Item 42
Item 43
Item 44
Item 45
Item 46
Item 47
Item 48
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5


ICC by groups
Math
Item 49
Item 50
Item 51
Item 52
Item 53
Item 54
Item 55
Item 56
Item 57
Item 58
Item 59
Item 60
Item 61
Item 62
Item 63
Item 64
P(X=1|Theta)
0 .2 .4 .6 .8 1
-4 -2 0 2 4
Theta
Combining OC and YC from R2 to R5


ICC by groups
Math
Item 65
Item 66
Item 67
Item 68
Item 69
Item 70
Item 71
Item 72
Item 73
Item 74
Item 75
Item 76
Item 77
Item 78
Item 79
Item 80
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5


ICC by groups
Math
Item 81
Item 82
Item 83
Item 84
Item 85
Item 86
Item 87
Item 88
Item 89
Item 90
Item 91
Item 92
Item 93
Item 94
Item 95
Item 96
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

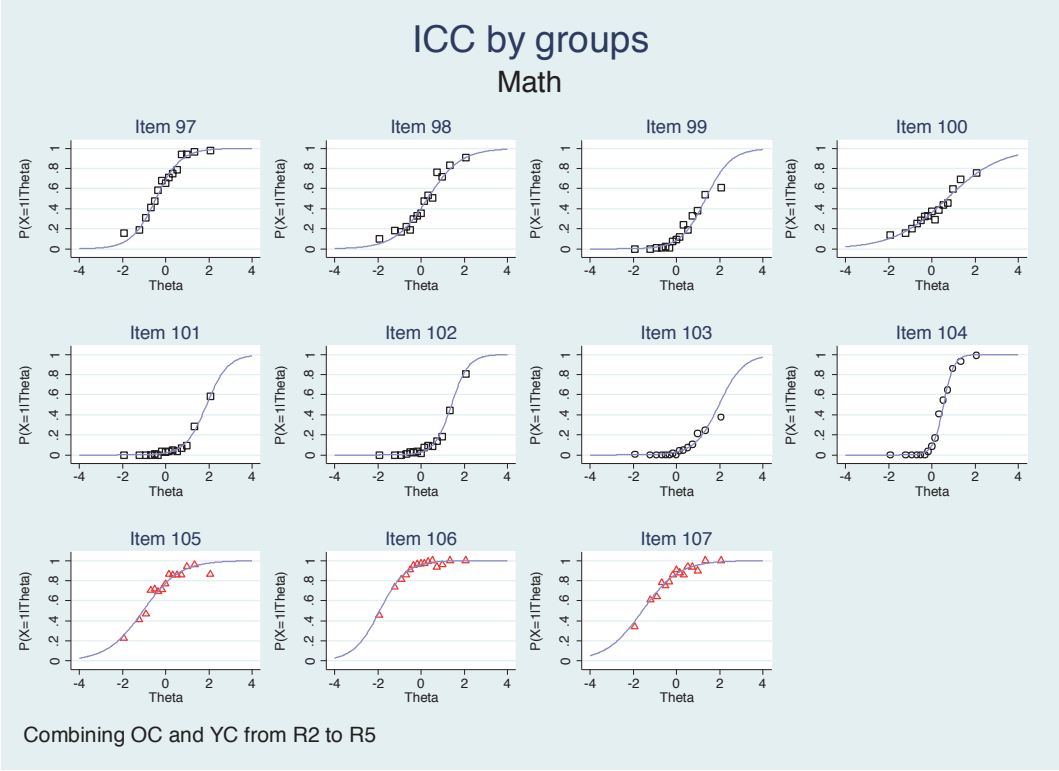
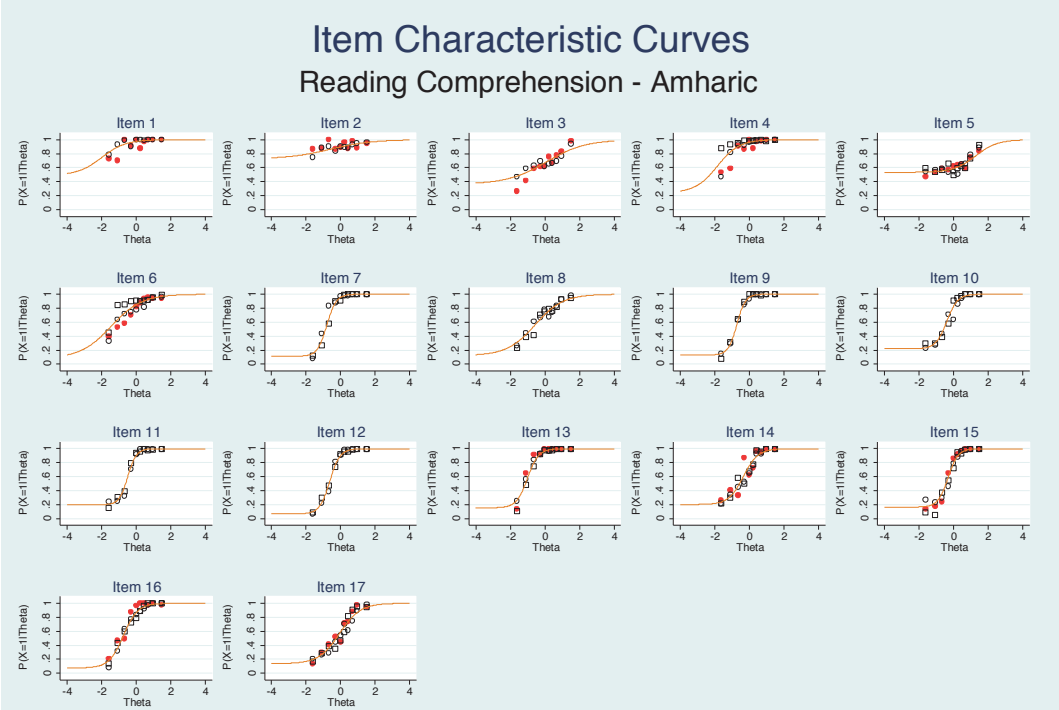

ICC by groups
Math
Item 97
Item 98
Item 99
Item 100
Item 101
Item 102
Item 103
Item 104
Item 105
Item 106
Item 107
P(X=1|Theta)
Theta
Combining OC and YC from R2 to R5

Figure 11. *Differential Item Functioning for reading comprehension analysis, Amharic*



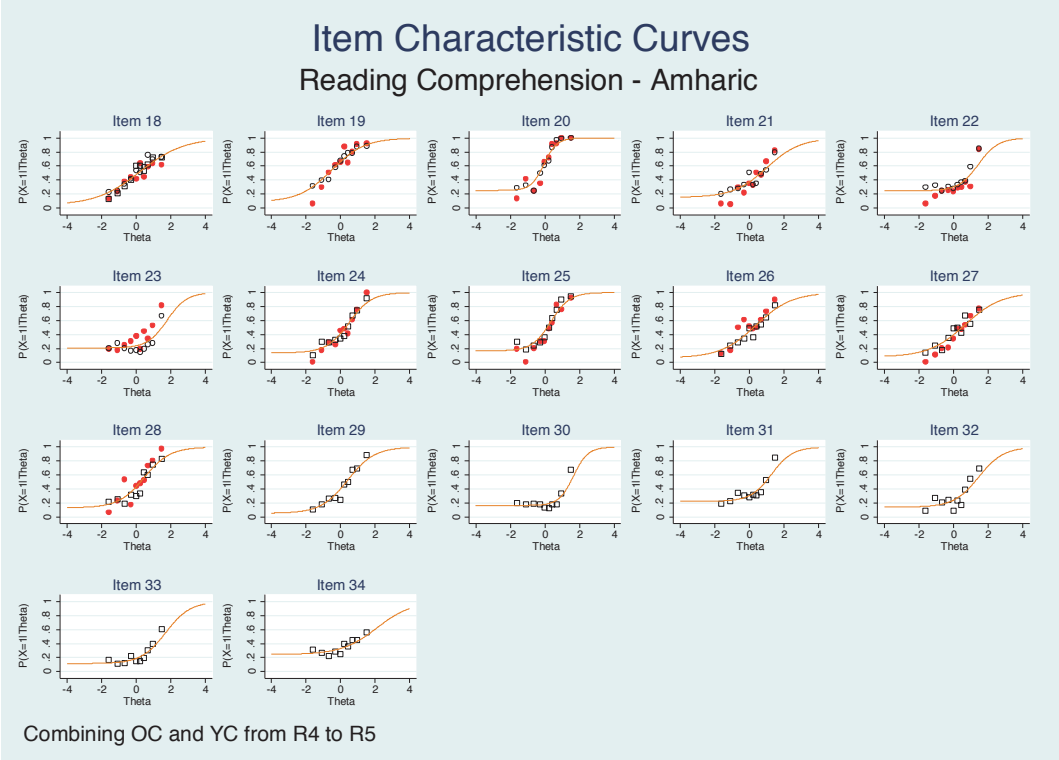
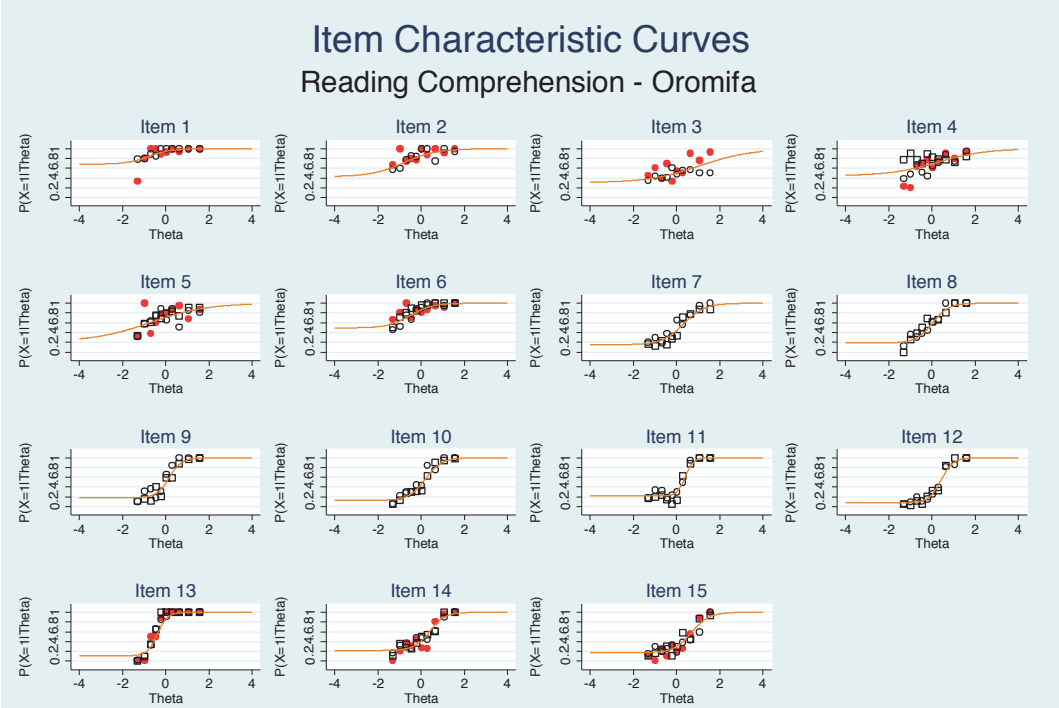


Figure 12. *Differential Item Functioning for reading comprehension analysis, Oromifa*



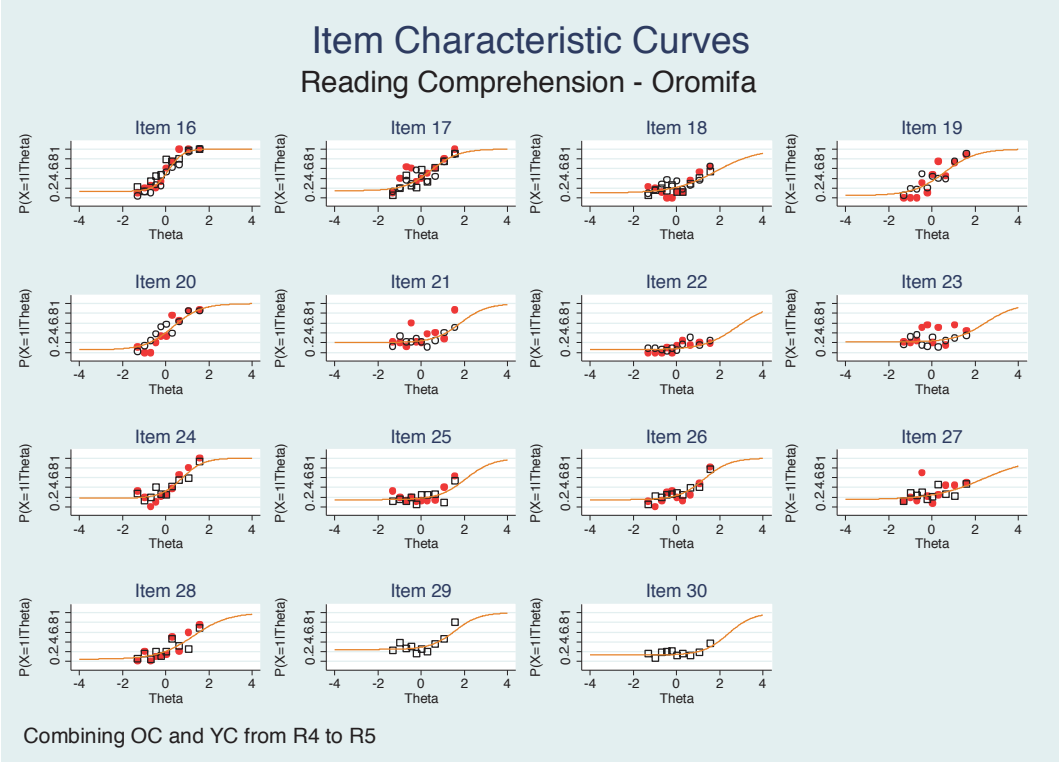
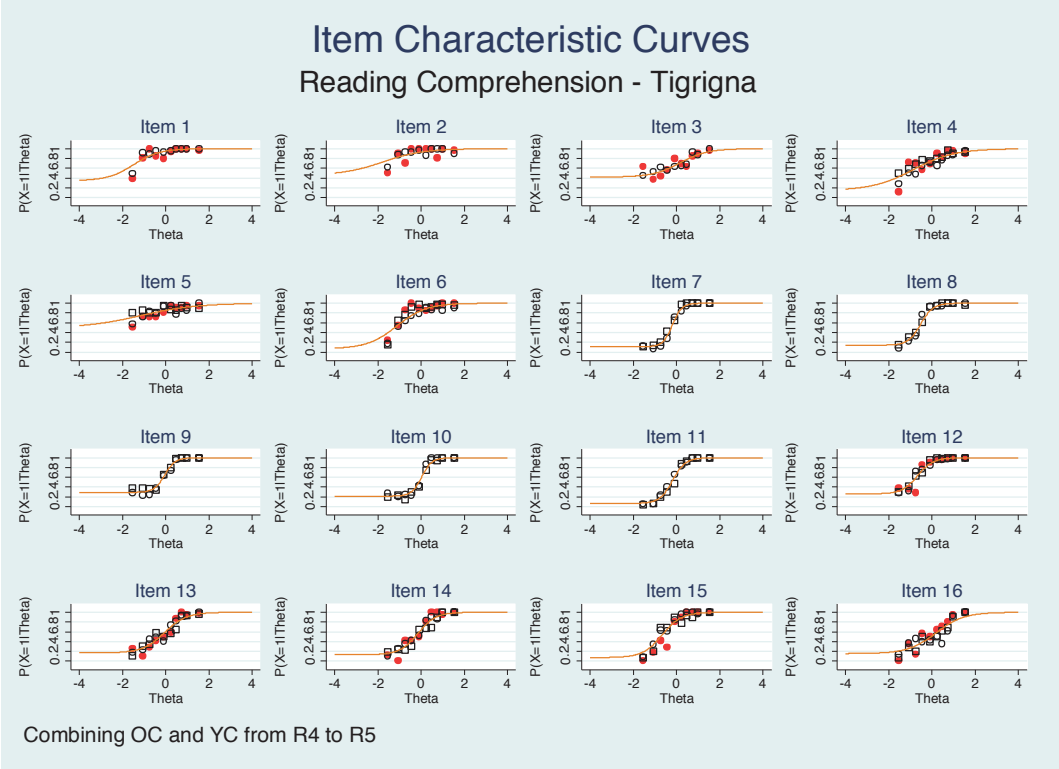


Figure 13. *Differential Item Functioning for reading comprehension analysis, Tigrigna*



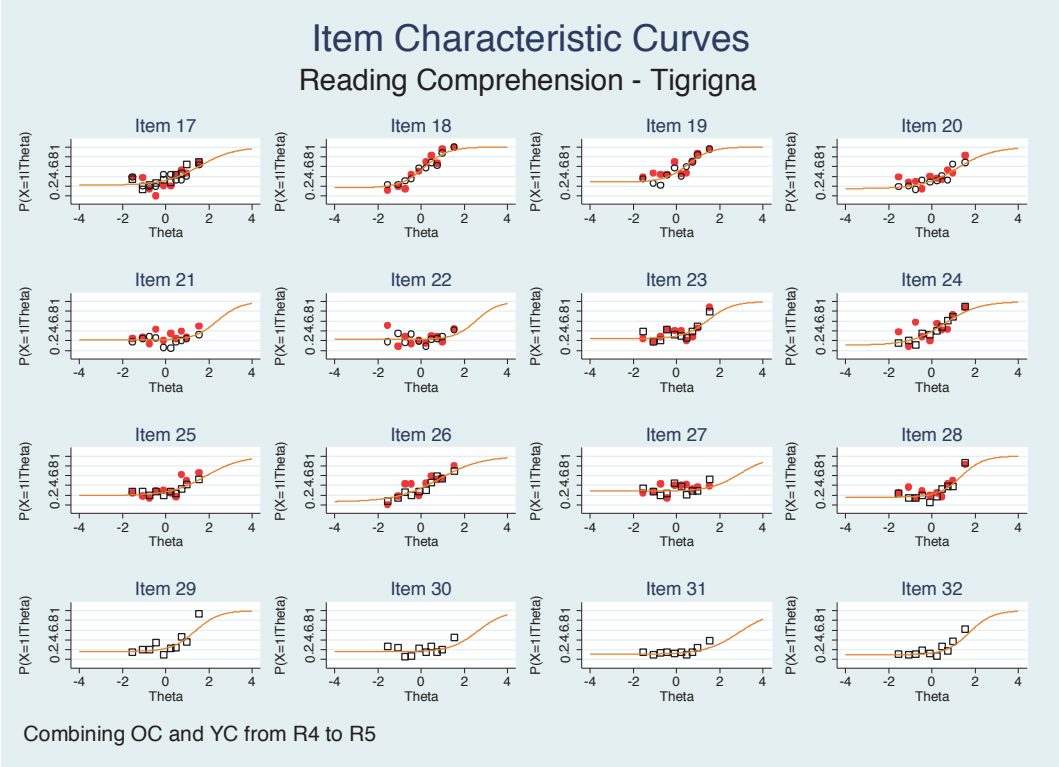
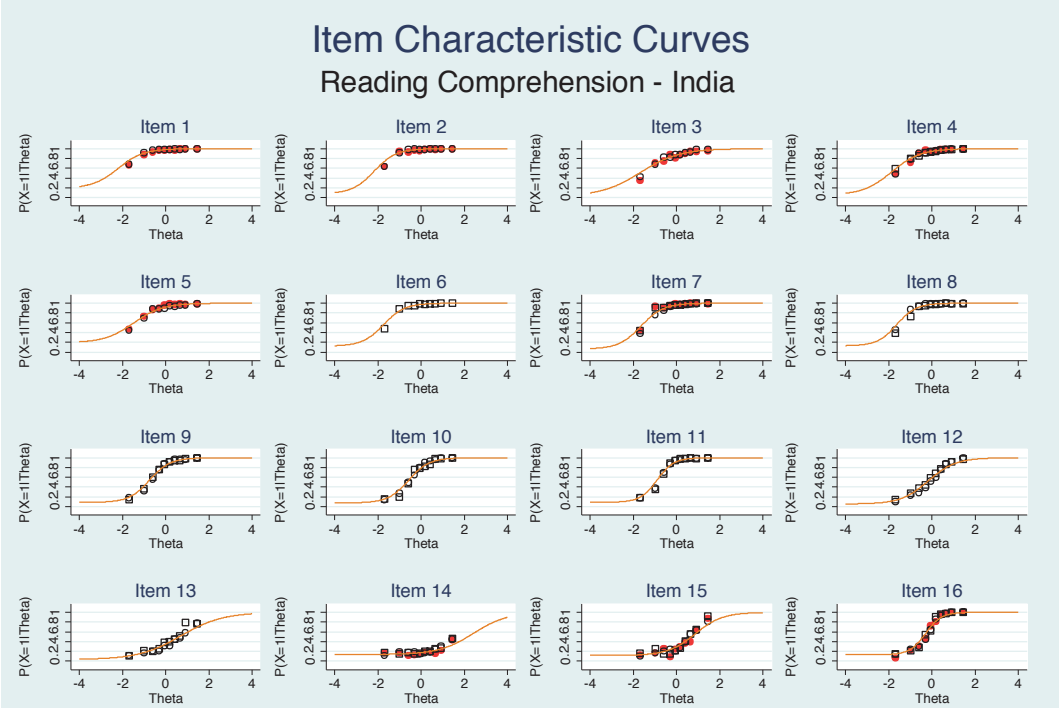


Figure 14. *Differential Item Functioning for reading comprehension analysis, Telugu*



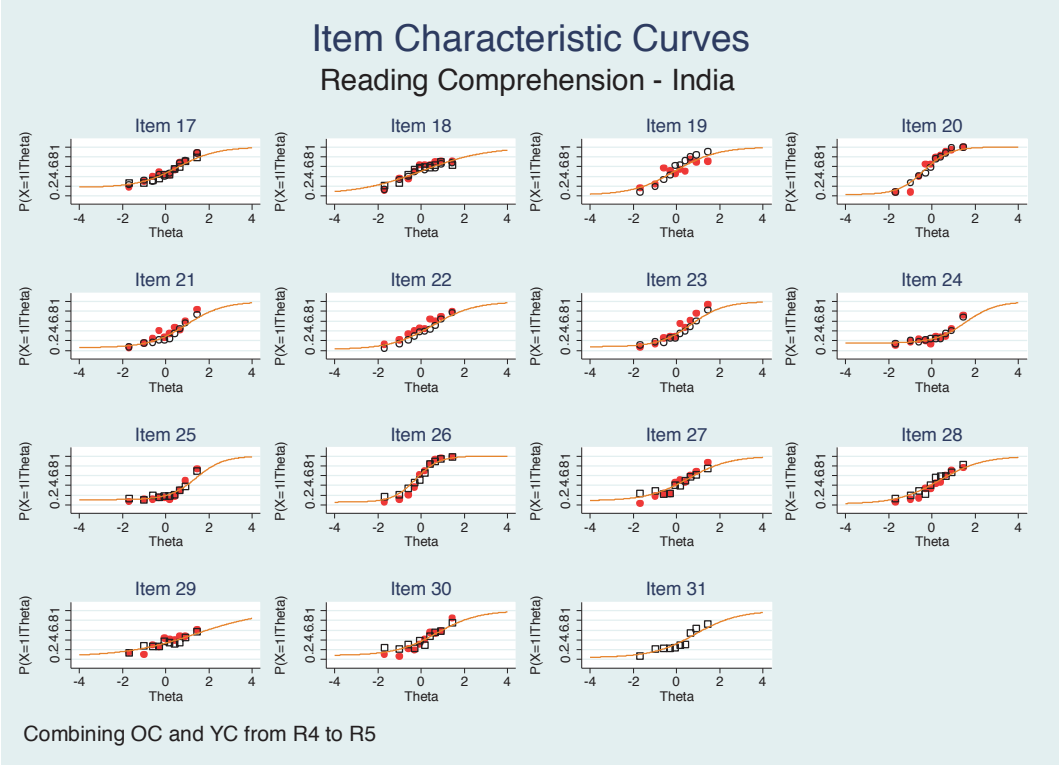
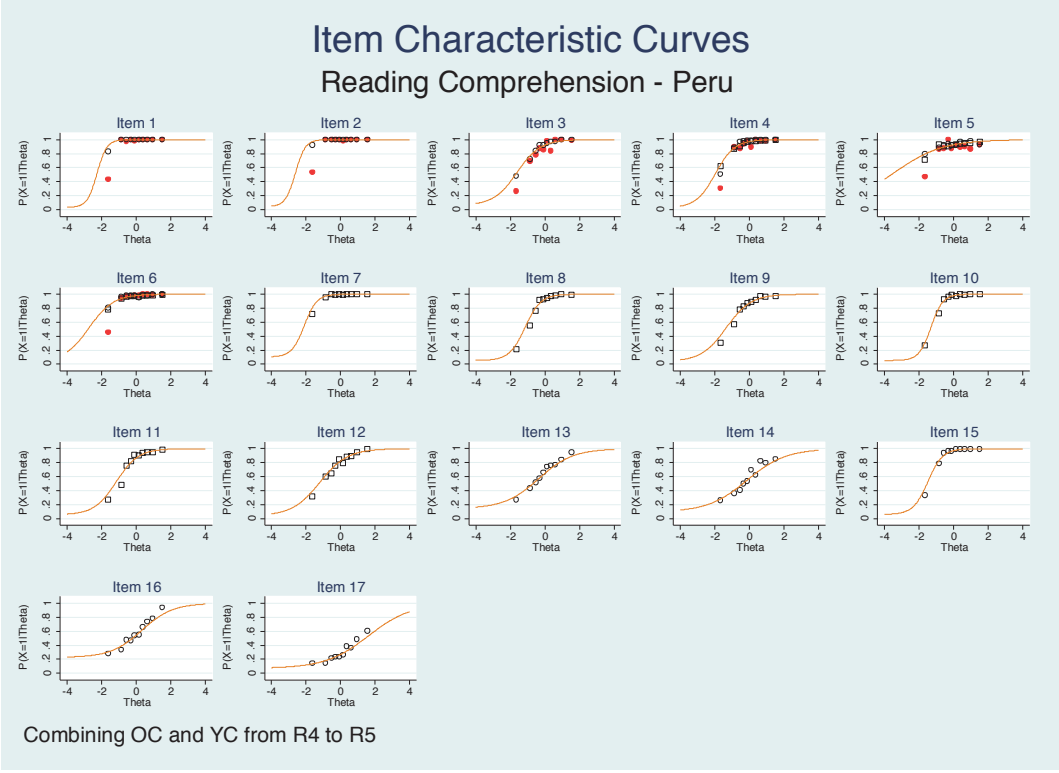


Figure 15. *Differential Item Functioning for reading comprehension analysis, Spanish*



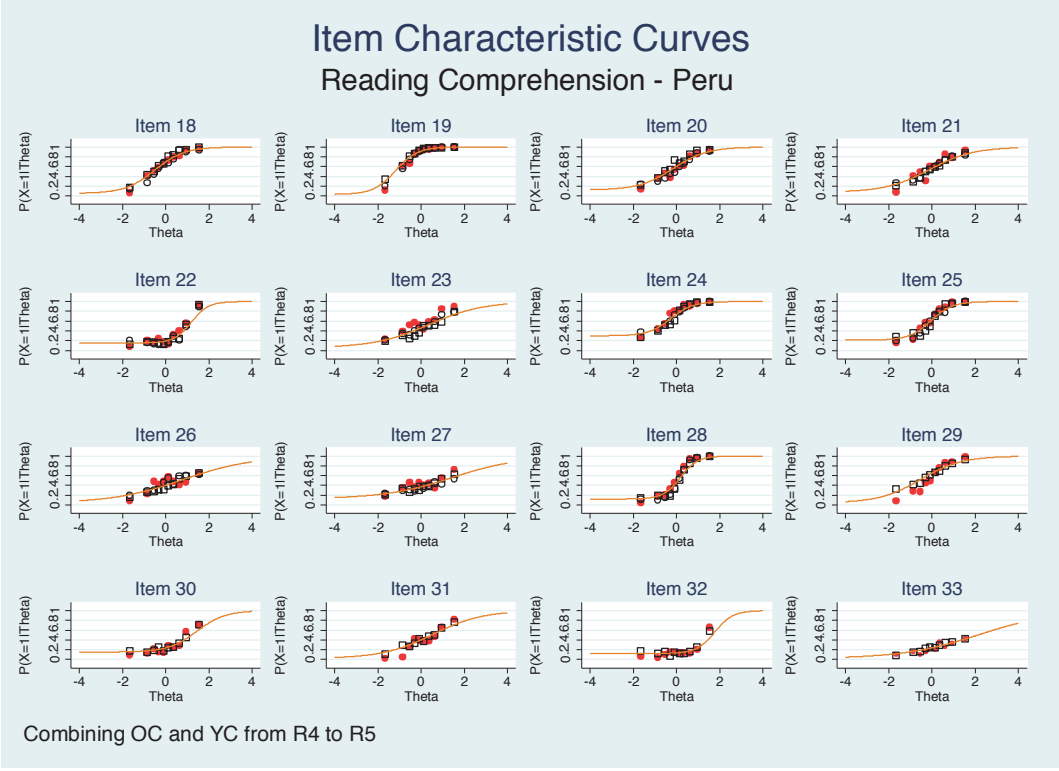
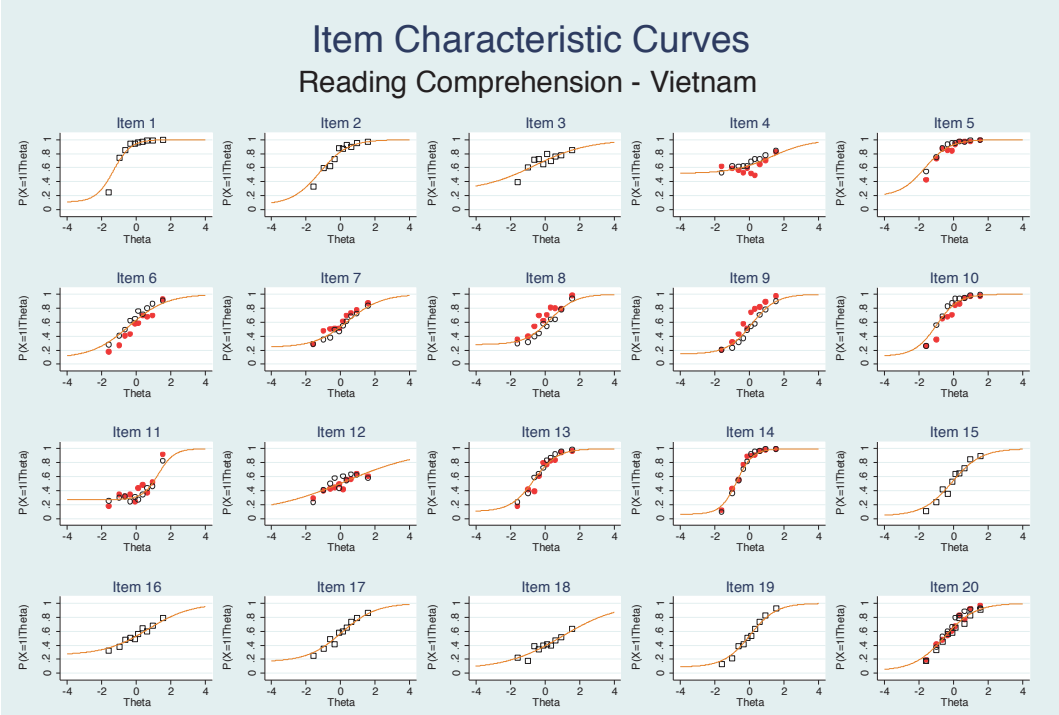
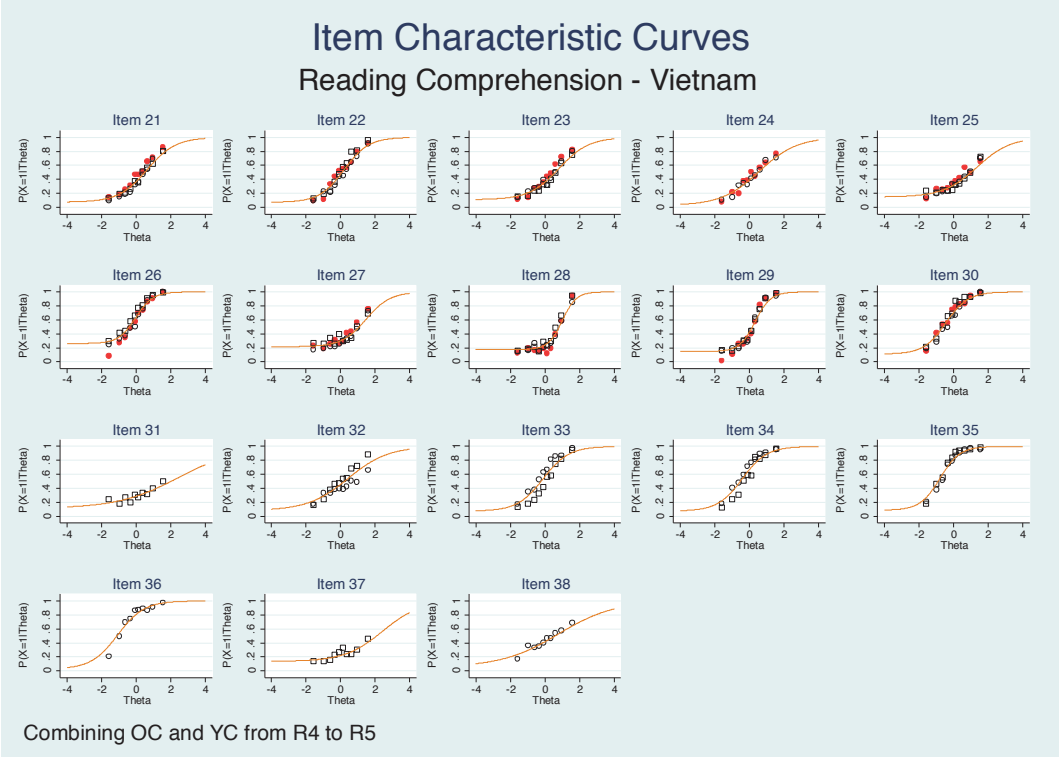


Figure 16. *Differential Item Functioning for reading comprehension analysis, Vietnamese*





Equating Cognitive Scores across Rounds and Cohorts for Young Lives in Ethiopia, India, Peru and Vietnam

For longitudinal studies such as Young Lives, getting comparable measures of children's cognitive abilities over time is essential for identifying individual, family, school or contextual variables that affect children´s development. Few longitudinal studies that follow birth/age cohorts have comparable cognitive measures over time; of those that are available, most are from developed countries and there are almost none from developing countries. For example, studies such as The National Education Longitudinal Study (NELS), the Early Childhood Longitudinal Study – Kindergarten (ECLS-K), Education Longitudinal Study (ELS), and the Rochester Longitudinal Study in the United States have achievement measures (maths and reading comprehension) that are comparable across waves. To help fill this gap, this technical note outlines the statistical procedures that Young Lives has followed to achieve comparable measures across rounds and age cohorts in Ethiopia, India, Peru and Vietnam.

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 18 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.

Young Lives Partners

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

- *Policy Studies 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@qeh.ox.ac.uk
Website: www.younglives.org.uk

Young Lives