What Determines Teacher Quality in a High-Performing Education System? Value-Added Estimates for Primary Schools in Vietnam

Pedro Carneiro* Paul Glewwe† Anusha Guha*
Sonya Krutikova‡

June 2021

*University College London and Institute for Fiscal Studies
†University of Minnesota
‡Institute for Fiscal Studies

1 Introduction

There is a growing body of literature showing that access to high quality teachers has not only short-term gains for students at all levels of schooling but can also have lasting positive effects on later-life outcomes (Chetty, Friedman, and Rockoff, 2014). Based on these findings, estimates monetizing potential gains show that in the U.S. context improving teacher quality could yield substantial benefits. Many gaps remain, however. There is still very little evidence for contexts outside the U.S, especially for Lower and Middle Income countries (LMIC), where there is a particularly great need to understand how to improve learning outcomes in school. Even in the U.S. context, the bulk of evidence is limited to effects that teachers have on numeracy and literacy skills; in light of growing evidence of the importance of a broader set of skills and capabilities for long-run outcomes (Heckman, Stixrud, and Urzua, 2006), an important question is how much schools and teachers contribute to the development of these skills.

In this paper we study teacher effectiveness in Vietnam. Vietnam is an outlier among Lower-Middle Income countries with respect to the very high levels of learning achieved for its income level (Dang et al., 2021). However, there is very little prior evidence on the role that teachers play in this system. This is a particularly interesting question as in addition to being high performing, Vietnam's education system is highly centralised with respect to teacher recruitment, training and curriculum. It is not clear, a-priori, therefore, how much variation there is in teacher effectiveness and whether having a higher or lower quality teacher matters as much as has been found in other LMIC contexts, such as Pakistan, India and Ecuador (Bau and Das, 2020; Singh, 2015; Araujo et al., 2016).

We study teacher effectiveness in raising a range of skills, including not only core academic skills - literacy and numeracy - but also a broader set of higher order cognitive and non-cognitive capabilities. These were selected to include capabilities that are considered to be important dimensions of child development in the psychology and education literature and have been identified as priority skills in the new primary school curriculum currently being developed in Vietnam - Vietnam New Curriculum. Teacher's effectiveness is the teacher's contribution to student's learning in an academic year. We estimate the class effect using the test scores of students and with the sub-sample of two cohorts taught by the same teacher we are able to separate teacher effect from classroom effect. Beyond assessing the role that teacher quality plays in shaping these skills, we are able to investigate whether the dimensions of teacher quality that are relevant for formation of core academic cognitive skills are similar to those relevant for broader cognitive and non-cognitive development by studying whether teachers who are effective at raising the former are also effective at raising the latter.

We utilise a rich longitudinal data set from 140 primary schools and over 5,000 primary school pupils in Vietnam. In this study we use data for grades 2 and 3 of primary school. A key challenge with studying teacher effects using observational data is to address potential sorting of students and teachers. We sample two classes per cohort per school in order to be able to estimate classroom specific effects within a school-grade-year, thus addressing concerns about non-random selection of pupils into schools. In order to address concerns about sorting within a school across classes we are able to rely on teacher and principal reports on how children and teachers are assigned to classes.

Another important challenge is to separate variance in test scores due to teachers from that due to classroom random shocks. In a sub-set of the schools in our sample, we observe two different cohorts of children being taught by the same teacher in grade 2. We use the covariance in class effects across the classes taught by the same teacher to estimate the variance of the teacher effect purged of classroom shocks like random compositional classroom changes overtime or noise next to a classroom, following Hanushek and Rivkin, 2012 and McCaffrey et al., 2009. A limitation of our design is that we are able to estimate teacher value added in this way only for academic skills in grade 2. For academic skills in grade 3, as well as higher order cognitive and non-cognitive skills, our estimates combine effects of teachers and classroom shocks.

Our preliminary results are as follows. First, we find that teachers have a moderate effect on children's learning. A one standard deviation increase in classroom quality for students in grades 2 and 3 increases students' mathematics scores by 0.08 - 0.10 standard deviations. This suggests that if a pupil moves from a class in the bottom 5th percentile of the class value added distribution to the top 95th, his/her math test scores would increase by up to a third of a standard deviation. For both grades 2 and 3, value-added estimates for Vietnamese are somewhat lower, at between 0.062-0.083 standard deviations, or up to 0.28 of a standard deviation improvement in response to movement from a class in the bottom 5th percentile of the class value added distribution to the top 95th. These class effects are similar in magnitude to the class effects estimates in Araujo et al for Ecuador, but much lower than the estimates in Bau and Das for Pakistan.

Separating out the effects of differences in teacher quality from classroom shocks, we estimate grade 2 teacher effects of 0.054 standard deviations for math and Vietnamese. This impact is only 14-19 percent of the impact of one year of schooling on median test scores (0.28 to 0.39 of a standard deviation for grade 2 students). In terms of moving a student from a teacher in the bottom 5th percentile of quality to 95th percentile, this move would improve scores on math and Vietnamese by just under 0.18 of a standard deviation. This teacher effect of 0.054 standard deviations is a much lower than that found by Bau and Das for Pakistan: they estimate that a 1 standard deviation improvement in teacher quality results in a 0.15 standard deviation improvement in test scores, which is almost one half of an annual test score gain of 0.33. It is also lower that the 0.09 standard deviation teacher effect reported by Araujo et al in the context of Ecuador.

Second, we estimate class effects for a set of additional higher order cognitive skills - executive functions (EF) - which capture children's ability to maintain attention, focus on the task at hand, and maintain relevant information in mind. These skills are related to two sets of competencies targeted in the Vietnam New Curriculum - Self-study and Self-managed learning, as well as Problem Solving and Creativity. EF consists of three domains; inhibitory control, working memory and cognitive flexibility. We use validated direct assessment tasks that have been used internationally to capture each of these domains. We find that, in fact, class value added is somewhat higher for these skills than academic skills: one standard deviation increase in classroom quality results in a 0.12-0.17 standard deviation increase in different indicators of executive function. This is higher than the 0.07 standard deviation classroom effect estimated for Ecuador.

Third, we collected measures of "non-cognitive" skills, including self-perception across several domains (scholastic, social and physical) as well as propensity for independent mastery of school subjects rather than reliance on teachers. Existing studies suggest that more independent students with higher perceived competence attain better success at school. In addition, these competencies link to the Vietnam New Curriculum focus on academic, physical and social competence, as well as competence in self-study and self-managed learning. We find class effects in the range of 0.13-0.16 standard deviations.

Finally, we find that our measures of cognitive and non-cognitive skills are not highly correlated with each other, in two distinct ways. First, while students' cognitive skills, in particular mathematics and Vietnamese, are highly correlated with each other (correlation coefficient of 0.66), cognitive and non-cognitive skills are at most weakly correlated (correlation coefficients between -0.04 and 0.12). Second, the same holds for correlation of the classroom effects that we estimate: the correlation between the mathematics and Vietnamese class effects is 0.50, while the correlations between these two cognitive skills and the self-perception and mastery scores range from -0.09 to 0.04. This suggests that the classes, and likely the teachers, that are relatively effective at developing students' cognitive skills are not the same classes, and likely not the same teachers, that are most effective at developing students' non-cognitive skills.

2 Education in Vietnam

Vietnam's primary and secondary education system is divided into primary school (grades 1-5, starting at age 6), lower secondary school (grades 6-9), and upper-secondary school (grades 10-12). Vietnam also has pre-primary education (for ages 3-5), secondary vocational training schools, and many different post-secondary institutions. In 2014, Vietnam had more than 15,000 primary schools, 10,000 lower-secondary schools, and 2,300 upper secondary schools.

Virtually all primary schools in Vietnam are state-managed and thus are public schools. In 2013, about half of primary schools were providing "full day" (6 hours) instruction; the other half received only "half day" (3.5 hours) instruction, with schools usually operating two shifts. An explicit goal of the government is to extend full-day schooling to poorer localities, but this has proceeded slowly. At first glance, Vietnamese children's time in school seems very low; it has one of the shortest school days, and one of the shortest school years, in the world. Yet in most areas of Vietnam – including rural areas – parents send their primary-age children to varying hours of "extra study" classes, though the time and resources devoted to this activity vary across provinces.

Administratively, the Ministry of Education and Training (MoET) in Hanoi retains formal authority over the entire education system. MoET works with other line ministries to determine investments in education, and plays the leading role in education planning and in determining the content of curricula (London, 2011).

This an especially policy relevant time for this study as the Vietnamese government is keen to reform the current education system in order to equip students with the skills needed for a modern economy. In particular, the government has issued various guidelines that encourage students to develop their non-cognitive skills, including social skills, creativity, and self-learning ability, and that discourage teachers from using ineffective teaching practices such as passive learning and rigid memorization (Government of Vietnam, 2014). In 2018, the Vietnamese government announced a major education reform – the "Fundamental and Comprehensive Reform of Education" – that has started to implement major revisions to the official curriculum, pedagogical methodology, and teacher professional development in order to provide Vietnamese children with the skills they need to be effective participants in Vietnam's economy and society. Our study is able to shed light on the role that schools and teachers currently play in shaping this wider set of skills.

3 Data and Measurement

The data used in this paper were collected from 140 primary schools that are approximately nationally representative of all of Vietnam's primary schools. To date, data have been collected for three school years, 2017-18, 2018-19 and 2019-20; a fourth year of data has just been collected (in April-May of 2021) for the 2020-2021 school year but is not yet available for analysis. For each of the 140 schools, data were collected for two adjacent cohorts of students, those who started grade 2 in the 2017-18 school year (henceforth cohort 1), and those who started grade 2 in the 2018-19 school year (cohort 2). Data were collected from cohort 1 for the 2017-18, 2018-19 and 2019-20 school years, when they were in grades 2, 3 and 4, and data were collected from cohort 2 for the 2018-19, 2019-20 and 2020-21 schools years, when they were in grades 2, 3 and 4. Table 1 shows when the data were collected from the two cohorts.

Table 1: **Dates of Data Collection for Two Cohorts of Students**

Date	Nov 2017	April 2018	Nov 2018	April 2019	April 2020	April 2021
Cohort 1	Grade 2	Grade 2		Grade 3	Grade 4	
Cohort 2			Grade 2	Grade 2	Grade 3	Grade 4

For both cohorts, more than 5,000 students were tested in mathematics and Vietnamese in their classroom when they were in grade 2 and grade 3. The data was collected from a random sample of 20 students per class and each test lasted an hour. Table A2 in Appendix provides the attrition rate over different rounds. We collected more detailed data on a random sub-sample of slightly more than 1600 students in each cohort. These pupils, and their parents, were interviewed. The pupils also completed assessments for an extended set of skills, including EF and a battery of non-cognitive skills. Table 2 provides some basic information on both cohorts when they were in grade 2. As can be seen, the two adjacent cohorts are quite similar; none of the differences is significant at the 5% level, although the differences for birth order and gender are significant at the 10% level.

A key issue regarding estimation is whether the cohort 2 students have the same teachers that the cohort 1 students had, for example whether the two grade 2 teachers in the two classrooms from which data were collected for cohort 1 in the 2017-18 school year were teaching grade 2 in the following school year (2018-19) when cohort 2 was in grade 2. Estimation of the variance of the teacher fixed effects, and not just the

Table 2: Sample Statistics (Grade 2)

	Cohort 1		Cohort 2	
	mean	sd	mean	sd
Age in years	7.416	0.558	7.386	0.564
Proportion Ethnic minority	0.232	0.422	0.249	0.432
Proportion Male	0.539	0.499	0.506	0.500
Birth order	1.985	1.130	1.898	1.116
Mother's highest grade	8.554	3.361	8.577	3.470
Father's highest grade	8.688	3.266	8.496	3.247
Father's age	36.977	6.508	36.812	5.955
Mother's age	34.007	5.859	34.031	5.557
Students tested (Math,Vietnamese)	5070		5209	
Students interviewed (including Ncog)	1654		1673	
Schools	140		140	
Classrooms	276		279	

variance of the class fixed effects, requires that the same teachers teach students in the same grade in two adjacent years. Table 3 provides this information for grades 2 and 3.

Table 3: Extent to which Both Cohorts had the Same Teacher, Grades 2 and 3

	Grade 2	Grade 3
Cohort 1	276	268
Cohort 2	268	295
Schools	140	140
Classes	276	276
Common across cohort	165	103
Schools	104	80
Common (both classes)	124	46
Schools	62	23

Consider the information on grade 2 in Table 3. The plan for both cohorts was to collect data on two grade 2 teachers, in two separate classrooms, for each of the 140 schools, which implies collecting data on 280 teachers. For cohort 1, data were collected in 2017-18 for two grade 2 teachers in 136 of the 140 schools, while data were collected for only one teacher in 4 of those 140 schools (because those schools had only one grade 2 class in 2017-18), so data were collected for 276 grade 2 classes. For cohort 2, the number of grade 2 classes in 2018-19 was 279 (only one school had only one grade 2 class in 2018-19), so we have 275 grade 2 classrooms (but not necessarily the same number of teachers) in which cohort 1 was taught in 2017-18 and cohort 2 was taught in 2018-19. The key issue is whether these two cohorts of students had the same grade 2 teachers. This is the case for 165 teachers, who are spread across 104 of the 140 schools. Yet in only 62 of these schools are *both* grade 2 teachers the same for both cohorts. While this reduces the sample by almost half, this is a sufficient sample to carry out the estimation procedure described in Section 4.

The analogous figures for grade 3 are shown in the last column of Table 3. The discouraging situation here is that in only 23 of the 140 schools did both of the grade 3 teachers in the 2018-19 school year continue to teach grade 3 in the 2019-20 school year. This attrition is rather severe and so we do not estimate the variance of teacher effects for grade 3 (although we do estimate class fixed effects, which do not require following the same teacher over time).

Data were also collected from about 280 teachers using a teacher questionnaire. Table 4 presents some basic characteristics of teachers, separately for grades 2, 3 and 4. The typical teacher for these grades is about 42 years old, and the vast majority (90% in grade 2, and 78% in grades 3 and 4) are women. About one out of eight is a member of an ethnic minority group, and about 90% are married. About two thirds have a university education, and well over 90% have had pedagogical training at the university level. Almost all (92% to 94%) have permanent positions.

Table 4: Teacher Characteristics of Older Cohort Students, by Grade

	Grade 2		Grade 3		Grade 4	
	mean	sd	mean	sd	mean	sd
Teacher's age	41.179	8.736	42.873	8.262	41.717	8.140
Male teachers	0.099	0.299	0.216	0.413	0.221	0.416
Ethnic minority	0.139	0.346	0.142	0.349	0.107	0.310
Married	0.909	0.288	0.884	0.320	0.903	0.296
Univ education	0.620	0.486	0.634	0.483	0.689	0.464
Univ pedagogical training	0.916	0.278	0.933	0.251	0.962	0.192
experience	20.226	9.380	21.705	8.829	20.498	8.895
Permanent position	0.916	0.278	0.922	0.269	0.941	0.236
Observations	276		268		289	
Tchrs moving with students			31		29	

3.1 Academic Skills Assessments

Information on the cognitive and non-cognitive tests given is provided in Table A1 in Appendix. The initial math and Vietnamese tests given at the beginning of grade 2 both had 25 items (questions). Later tests in those subjects had 30 items. Common (linking) items were included in order to put students' performance on these tests on a common scale, using item response theory (IRT).

For both the mathematics and Vietnamese tests, items were developed for each grade by the Vietnam Institute of Educational Sciences. All tests administered during the RISE project assess domains relevant to the Vietnamese curriculum. The Maths test assesses three skills: arithmetic, measurement and quantities and geometry and Vietnamese test assesses four language skills: vocabulary, rhetoric, grammar and reading comprehension. They both assess the following cognitive domains - knowledge, understanding and application. We conducted two pilot tests, the first was done in May of 2017, before the first round of data collection, and the second was in January of 2019, before collecting data from grade 4. For both pilot tests, about 300 students and their teachers participated for each grade. The best performing test items (questions) were selected by using IRT diagnostic analysis. Final scores on (the final versions of) each test were constructed using a 2-parameter IRT model, and items with poor performance were excluded from the dataset.

To compare students across grades and across the two cohorts, all test results were combined into a single data set (separately for math and Vietnamese) and IRT analysis was used to construct latent measures of mathematics and Vietnamese ability for each student that are comparable across grades and across the two cohorts. For ease of interpretation, these math and Vietnamese latent scores were normalized to have a standard deviation on one and mean of zero within each grade.

3.2 Executive Functioning Measures

In addition to these academic skills we also administered measures intended to capture high-level cognitive processes that are within the group of "executive functions" that enable individuals to concentrate (Burgess and Simons, 2005, Espy, 2004, Miller and Cohen 2001). Students use executive functioning skills to maintain concentration in class, consolidate taught material and apply it to complex problems. Executive functioning was measured using two different tests: "hearts and flowers" and "backward digit span". The former is a computerized task where a stimulus appears on the right or left of the screen. The rules are: (a) For one stimulus, press on the same side as the stimulus (called the congruent condition); (b) For the other stimulus, press on the side opposite the stimulus, which requires inhibiting the natural tendency to activate the hand on the same side as the stimulus (called the incongruent condition). This task requires both working memory and inhibitory control. The backward digit span tests only working memory; respondents are required to repeat a series of numbers back to the assessor in reverse order, and the final score is equal to the longest sequence of numbers that a student was able to say backwards.

3.3 Non-Cognitive Skill Measures

The first measure of non-cognitive skills is the self-perception profile. It asks students to compare themselves to hypothetical children with high and low skills or characteristics of various types (e.g. academic performance, social skills, and physical appearance). This measures students' self-perceptions of these skills. The second measures intrinsic vs. extrinsic motivation skills, of which only one scale was administered: independent mastery vs. dependence on the teacher. This measures the extent to which the student likes to solve problems independently or prefers to depend on the teacher for help.

4 Methodology

This section presents the methodology used to estimate school and teacher value added, and to estimate the impact of teacher characteristics and behavior on student learning.

The overall objective of this paper is to understand the characteristics and behavior of teachers in Vietnam that make some of them more productive than others. The first step is to estimate teacher productivity, and the second step is to examine what observable characteristics of those teachers appear to make them more productive.

In the current version of the paper we focus only on estimating the teacher productivity, which is defined as the value added of a teacher on students' learning. The starting point is to estimate a classroom fixed effect for a given time period. This can be done by estimating the following regression equation:

$$Y^k_{icst} = \alpha^k_{cs} + \beta Y^k_{icst-1} + \epsilon^k_{icst}$$

where Y^k_{icst} is end-of-school-year score of student i in classroom c in school s at time t on a test for subject k (mathematics, Vietnamese), α_{cs} are classroom indicators, Y^k_{icst-1} is beginning-of-school-year score on subject k and ϵ^k_{icst} is an i.i.d. error term. The classroom fixed effects (α_{cs}) are estimates of classroom value added under the assumption that, conditional on controls (primarily the test score at the beginning of the school year or at the end of the previous grade) students are randomly assigned to classrooms. It is not plausible that students are randomly assigned to schools, even conditional on past learning, but it is generally accepted that the assumption of conditional random assignment within school is reasonable (e.g., Chetty, Friedman, and Rockoff, 2014). This requires, however, calculating classroom value added relative to the school mean, which ignores any cross-school variation in school quality.

Therefore, since we have two classrooms per school, as in most papers in this literature, we redefine each classroom effect relative to the school average to address the issue of sorting of teachers and/or students into schools. This will estimate classroom effects using only variation across classrooms within schools. The demeaned classroom effect, denoted by λ^k_{cs}, is:

$$\lambda^k_{cs} = \alpha^k_{cs} - \frac{\sum_{c=1}^{C_s} N_{cs}\alpha^k_{cs}}{\sum_{c=1}^{C_s} N_{cs}}$$

where C_s is the number of classrooms in a school and N_{cs} is the number of students in the classroom c in school s (in our analysis C_s always equals 2).

To measure the overall contribution of (variation in) classroom quality to (variation in) student learning, it is useful to estimate the variance of the classroom effect, which can be denoted by $(V(\lambda^k_{cs}))$. In order to avoid overestimating the true variance of λ^k_{cs}, due to sampling error in the estimates of λ^k_{cs}, we apply a shrinkage procedure, following, for example, Chetty, Friedman, Hilger, et al., 2011. This correction can be expressed as follows:

$$V(\lambda^k_{cs}) = V(\hat{\lambda}^k_{cs}) - E\left\{\frac{(\sum_{d=1}^{C_s} N_{ds}) - N_{cs}}{N_c s(\sum_{d=1}^{C_s} N_{ds})}\sigma^2\right\}$$

where σ^2 is the within-classroom variance of residual student learning (the variance of ϵ^k_{icst} in equation (1)).

The (within-school) variance in class effects in equation (3) confounds both variation in teacher quality across classes as well as random "classroom shocks", which reflect random differences in (average) pupil characteristics over time (over cohorts) and random events that happen on the day of the test. To separate out variation in teacher quality (variation in teacher fixed effects) from "classroom shocks", one can estimate the covariance over time of the classroom effect for classrooms that have the same teacher in the two time periods. That is, one can estimate:

$$Cov(\lambda^k t_{cs}, \lambda^{k,t+1}_{cs})$$

If students are randomly assigned to teachers within schools, then the square root of this covariance is an estimate of the standard deviation of the teacher effects alone; the classroom shocks are uncorrelated over time and so drop out of this covariance term.

5 Results

This section presents our (preliminary) findings.

5.1 Academic Skills

We start by estimating the proportion of the overall variance of student test scores in math and Vietnamese that is accounted for by variance in classroom fixed effects, for both cohorts for grades 2 and 3.

Table 5 presents estimates of the within school standard deviations of classroom and teacher effects, adjusted for sampling error, for Mathematics and Vietnamese scores for both cohorts for grades 2 and 3. Class effects on mathematics for grade 2 pupils are very similar across the two cohorts and suggest that a one standard deviation increase in classroom quality results in around 0.10 standard deviation improvement in mathematics scores. Estimates for Vietnamese are somewhat lower, ranging between 0.083 for Cohort 1 and 0.062 for Cohort 2. These estimates are based on almost all of the schools for both cohorts. We find very similar size class effects for Grade 3.

For Grade 2, we are able to separate effects of classroom shocks from teacher quality, using a sub-sample of 62 primary schools for which we have the same grade 2 teachers teaching the two cohorts (so that we observe each teacher teaching two different classes). As expected, teacher effects are smaller than class effects, yet this reduction is much more pronounced for math (a reduction of almost one half) than for Vietnamese (a reduction of only about 13%). These teacher effects of 0.054 for both math and Vietnamese are a little more than one half of the effects of 0.09 for both math and language found by Araujo et al. in Ecuador. Furthermore, these teacher effects can be used to estimate the standard deviation of the shocks, which are 0.084 for math and 0.030 for reading,[1], suggesting that these effects can be quite large, as been found in other studies. In fact, for math the standard deviation of the classroom shock effect is larger than that of the teacher effect. These large classroom shock effects may be indicative of the presence of significant peer effects.

[1] These are calculated as (class effect^2 − teacher effect^2)$^{0.5}$ where the class effects are averaged over the two cohorts.

Table 5: **Within school standard deviations of classroom and teacher effects on mathematics and Vietnamese**

	Grade 2 Class Effects		Grade 3 Class Effects		Grade 2 Teacher Effect
	Cohort 1	Cohort 2	Cohort 1	Cohort 2	
Math	0.104	0.100	0.093	0.081	0.054
Vietnamese	0.083	0.062	0.078	0.073	0.054
Students	4556	4728	4148	4363	4700
Schools	127	130	119	124	62
Classes	254	260	238	249	124

5.2 Executive Functioning

Table 6 presents our estimates of the within-school standard deviation of classroom effects on two different measures of EF. The Hearts and Flowers (HF) task is a measure of functioning in the three core domains that make up executive functioning - working memory, switching and inhibition. The Backward Digit Span Task (BDST) measures working memory. Class effects on executive functioning can be estimated only for Cohort 1 Grade 3 as executive functioning tasks were administered only at the end of grades 2 and 3 and only for Cohort 1. Furthermore, they were administered to a sub-sample of around 7 children per class. Due to some compositional changes in classes between grades 2 and 3, however, our sample contains some classes that have even fewer than 7 children in grade 3 who completed the executive functioning tasks. We restrict our analysis sample to classes that had at least 3 children who completed these tasks, leaving us with a sample of 128 schools, 257 classes and 1,456 children. The raw scores on each task have been standardised to have a mean zero and standard deviation of 1 within the full sample of children who completed this assessment.

Our estimates suggest that classroom effects on EF are, if anything, slightly higher than on mathematics and Vietnamese. A 1 standard deviation increase in classroom quality leads to an improvement in executive functioning of between 0.12 and 0.17 standard deviations. If we assume that the classroom shock effect that we imputed for mathematics is the upper bound and the one for Vietnamese is the lower bound of what it is for EF, then these bounds suggest a teacher effect for the composite measure of EF (HF) of between 0.097 and 0.11 standard deviations.

Table 6: **Within school standard deviation of classroom effect on Executive Functioning**

Assessment score	Class Effect
EF: HF standardized raw score for round 3	0.123
EF: BDST standardized raw score	0.170
Students	1,456
Classes	257
Schools	128

5.3 Non-cognitive skills

The final set of classroom effect estimates is presented in Table 7. It shows effects on our measures of non-cognitive skills, including self-perception in the scholastic, social and physical domains, an aggregate self-perception score, and a mastery score (see Section 3 for further details). These scales were administered to the same sub-set of children who completed the EF assessment in Cohort 1, as well as to a sub-sample of children in Cohort 2 (also about 7 per class). In order to estimate the classroom effects on non-cognitive skills, we pool the two cohorts (to maximise sample size) but (as before) we exclude classes in which less than 3 children completed the non-cognitive scales. We used Item Response Theory (IRT) to construct the scores and standardised them to have mean zero standard deviation 1 within the grade.

These estimates are in the same range of magnitude as those for EF: A one standard deviation improvement in classroom quality results in a 0.14 standard deviation improvement in self-perception and 0.15 standard deviation increase in independent mastery. We know of no comparable estimates for either self-perception or independent mastery.

5.4 Classroom Effect Correlations

Having estimated classroom effects on math, Vietnamese, EF and non-cognitive skills we can now ask whether good teachers are good for acquisition of competencies in all domains or whether some are good at, for example, teaching academic skills while others are good at fostering non-cognitive skills. To do this, we simply look at the correlations between the classroom effects estimated for each our measures of cognitive and non-cognitive skills. Table 8 shows these estimates for Cohort 1, Grade 3 (as this is the cohort-grade combination for which we have the most complete data).

The first results to point out in this table is the high positive correlation between

Table 7: **Within school standard deviation of classroom effect on non-cognitive skills**

Non-cognitive domain	Class Effect
Self-Perception: Scholastic	0.153
Self-Perception: Social	0.126
Self-Perception: Physical	0.155
Self-Perception: Total	0.142
Independent mastery	0.153
Students	3127
Classes	544
Schools	140

math and Vietnamese classroom effects (0.501), which indicates that classrooms, and presumably the teachers in them, that are successful in raising students' math skills are also successful in increasing their reading skills. A second result of interest is the moderate positive correlation between EF and math and Vietnamese skills; this also makes sense because EF is hypothesised to support students' learning skills. Finally, it is interesting that there is almost no, or even negative, correlation between cognitive and non-cognitive class effects. This suggests that teachers, or, at least, classroom environments, that are effective for students' acquisition of cognitive skills are different from those that are good for their acquisition of non-cognitive skills.

Table 8: **Classroom Effect Cross-correlations for Cohort 1 Grade 3**

Variables	HF	BDST	SPsch	SPsoc	SPphys	SP	MAS	MATH	VT
HF	1.000								
BDST	0.014	1.000							
	(0.830)								
SPsch	-0.122	-0.105	1.000						
	(0.052)	(0.093)							
SPsoc	-0.023	-0.128	0.577	1.000					
	(0.715)	(0.040)	(0.000)						
SPphys	-0.090	-0.062	0.149	0.220	1.000				
	(0.153)	(0.320)	(0.017)	(0.000)					
SP	-0.099	-0.122	0.846	0.875	0.402	1.000			
	(0.116)	(0.052)	(0.000)	(0.000)	(0.000)				
MAS	0.033	0.126	-0.108	-0.162	-0.112	-0.129	1.000		
	(0.604)	(0.045)	(0.084)	(0.009)	(0.073)	(0.038)			
MATH	0.138	0.157	-0.182	-0.172	-0.096	-0.216	0.055	1.000	
	(0.027)	(0.012)	(0.003)	(0.006)	(0.126)	(0.000)	(0.379)		
VT	0.181	0.184	-0.206	-0.203	-0.090	-0.228	0.072	0.501	1.000
	(0.004)	(0.003)	(0.001)	(0.001)	(0.148)	(0.000)	(0.250)	(0.000)	

6 Conclusion (preliminary)

High quality teachers increase student learning at all levels of schooling, and also have long-lasting positive impacts on later-life outcomes. Yet most of the evidence for this comes from the U.S., and almost nothing is known about this for Lower and Middle Income countries. Even the U.S. evidence is mostly limited to the effects teachers have on students' numeracy and literacy skills, with very little known about broader sets of skills and capabilities. Finally, and most importantly, evidence is lacking on what distinguishes more effective teachers from less effective ones.

This paper addresses these questions for Vietnam, using a rich longitudinal data set from 140 primary schools and over 5,000 primary school pupils in that country. First, we estimate that a one standard deviation increase in classroom quality for students in grades 2 and 3 increases students' mathematics scores by 0.08-0.10 standard deviations, and increases Vietnamese scores by 0.06-0.08 standard deviations. Furthermore, we find that parsing apart classroom shock effects from teacher quality effects reveals that classroom shocks, especially for math, are quite substantial, at around 0.071 standard deviations and thus, the teacher effect is lower than that found in other studies, at around 0.05 standard deviations.

Second, turning to EF, we find that a one standard deviation increase in classroom quality results in 0.12 standard deviation increase in a composite measure of EF and a 0.17 standard deviation increase in a measure of working memory. These magnitudes are higher than those estimated for Ecuador (Araujo et al., 2016). Estimates of the impact of classroom quality on several domains of self-perception and independent mastery are in the same order of magnitude, ranging between 0.13 and 0.15 standard deviations.

These are the first estimates of teacher effects on academic skills for Vietnam. They are lower than estimates that exist for other countries, suggesting that variation in teacher quality may play a smaller role in driving inequalities in attainment at the primary school level in Vietnam compared to other settings for which estimates exist. This is an especially striking finding in the context of evidence that overall Vietnam achieves much higher levels of proficiency in mathematics and verbal skills, even when compared to countries that are several orders of magnitude richer (Dang et al., 2021). It suggests that Vietnam manages to achieve not only a high standard of teaching but also a higher degree of uniformity in the quality of teaching that primary school children receive than do other countries.

Furthermore, this is one of very few studies that estimate classroom effects on a wider range of competencies than core school subjects. We show that classroom quality explains as much if not more variation in these competencies, though our design does not allow us to identify how much of this is due to teacher quality. We find that there is a reasonably strong overlap between classroom environments which are good for learning mathematics and Vietnamese and a weaker but still positive and statistically significant correlation in class effects on mathematics and Vietnamese on the one hand and EF on the other. This is consistent with EF being an important set of skills for being able to learn.

A striking finding is that there is no correlation between classroom environments that foster development of independent mastery and confidence in own skills and those which are most conducive to development of cognitive skills - EF, mathematics and Vietnamese. In future work, we will also make use of direct classroom observations that has already been collected from all Grade 3 and Grade 2 classes for both the cohorts and is currently being coded. The classroom and teacher level data from teacher interviews and direct classroom observations will help us to identify what factors explain variation in teacher and classroom quality and how these differ depending on whether it is quality that is relevant for the development of cognitive or non-cognitive skills. This is an especially pertinent question in light of ongoing efforts to increase focus on fostering non-cognitive skills within the Vietnam as part of its ongoing education reforms.

References

Araujo, M Caridad et al. (2016). "Teacher quality and learning outcomes in kindergarten". In: *The Quarterly Journal of Economics* 131.3, pp. 1415–1453.

Bailey, Charles E (2007). "Cognitive accuracy and intelligent executive function in the brain and in business". In: *Annals of the New York Academy of Sciences* 1118.1, pp. 122–141.

Bau, Natalie and Jishnu Das (2020). "Teacher value added in a low-income country". In: *American Economic Journal: Economic Policy* 12.1, pp. 62–96.

Blankenship, Tashauna L et al. (2015). "Working memory and recollection contribute to academic achievement". In: *Learning and individual differences* 43, pp. 164–169.

Borella, Erika, Barbara Carretti, and Santiago Pelegrina (2010). "The specific role of inhibition in reading comprehension in good and poor comprehenders". In: *Journal of Learning disabilities* 43.6, pp. 541–552.

Burgess, Paul W and Jon S Simons (2005). "18 Theories of frontal lobe executive function: clinical applications". In: *The effectiveness of rehabilitation for cognitive deficits* 211.

Chetty, Raj, John N Friedman, Nathaniel Hilger, et al. (2011). “How does your kindergarten classroom affect your earnings? Evidence from Project STAR”. In: *The Quarterly journal of economics* 126.4, pp. 1593–1660.

Chetty, Raj, John N Friedman, and Jonah E Rockoff (2014). “Measuring the impacts of teachers II: Teacher value-added and student outcomes in adulthood”. In: *American economic review* 104.9, pp. 2633–79.

Dang, Hai-Anh et al. (2021). “What Explains Vietnam’s Exceptional Performance in Education Relative to Other Countries? Analysis of the 2012 and 2015 PISA Data”. In:

Davidson, Matthew C et al. (2006). “Development of cognitive control and executive functions from 4 to 13 years: Evidence from manipulations of memory, inhibition, and task switching”. In: *Neuropsychologia* 44.11, pp. 2037–2078.

Diamond, Adele and Andy Wright (2014). “An effect of inhibitory load in children while keeping working memory load constant”. In: *Frontiers in psychology* 5, p. 213.

Espy, Kimberly Andrews (2004). “Using developmental, cognitive, and neuroscience approaches to understand executive control in young children”. In: *Developmental neuropsychology* 26.1, pp. 379–384.

Government of Vietnam (2014). *Resolution No. 44/NQ-CP dated June 09, 2014*.

Hanushek, Eric A and Steven G Rivkin (2012). “The distribution of teacher quality and implications for policy”. In: *Annu. Rev. Econ.* 4.1, pp. 131–157.

Harter, Susan (1981). “A new self-report scale of intrinsic versus extrinsic orientation in the classroom: Motivational and informational components.” In: *Developmental psychology* 17.3, p. 300.

— (2012). “Self-perception profile for adolescents: Manual and questionnaires”. In: *Denver, CO: Univeristy of Denver, Department of Psychology*.

Heckman, James J, Jora Stixrud, and Sergio Urzua (2006). “The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior”. In: *Journal of Labor economics* 24.3, pp. 411–482.

Lemos, Marina S and Lurdes Verıssimo (2014). “The relationships between intrinsic motivation, extrinsic motivation, and achievement, along elementary school”. In: *Procedia-Social and Behavioral Sciences* 112, pp. 930–938.

London, Jonathan D (2011). *Education in Vietnam*. Institute of Southeast Asian Studies.

Marchant, Gregory J (1991). “A profile of motivation, self-perception, and achievement in Black urban elementary students”. In: *The Urban Review* 23.2, pp. 83–99.

McCaffrey, Daniel F et al. (2009). “The intertemporal variability of teacher effect estimates”. In: *Education finance and Policy* 4.4, pp. 572–606.

Oberle, Eva and Kimberly A Schonert-Reichl (2013). "Relations among peer acceptance, inhibitory control, and math achievement in early adolescence". In: *Journal of applied developmental psychology* 34.1, pp. 45–51.

Obradović, Jelena et al. (2018). "Assessing students' executive functions in the classroom: Validating a scalable group-based procedure". In: *Journal of Applied Developmental Psychology* 55, pp. 4–13.

Rosenthal, Eve N et al. (2006). "Digit Span components as predictors of attention problems and executive functioning in children". In: *Archives of Clinical Neuropsychology* 21.2, pp. 131–139.

Singh, Abhijeet (2015). "Private school effects in urban and rural India: Panel estimates at primary and secondary school ages". In: *Journal of Development Economics* 113, pp. 16–32.

Will Crescioni, A et al. (2011). "High trait self-control predicts positive health behaviors and success in weight loss". In: *Journal of health psychology* 16.5, pp. 750–759.

A Appendix

A.1 Additional Tables

Table A1: Cognitive and Non-Cognitive Tests Administered

Instrument	Description
Cognitive Test R1 (G2)	Maths Test with 25 items and Vietnamese Reading Comprehension test with 25 items
Cognitive Test R2 (G2)	Maths- 30 items and Vietnamese-30 items (M: 11 common items ; V: 12 common items from R1)
Non-Cognitive Test R2 (G2)	Executive Functioning, Self Perception Profile, Intrinsic vs.Extrinsic Motivation Scales, Competence Beliefs and Subjective Task Values
Cognitive Test R3 (G2)	Maths-25 items and Vietnamese-30 items (M: 16 common items; V: 12 common items from R1)
Cognitive Test R4 (G3)	Maths-30 items and Vietnamese-30 items (M: 11 common items; V: 13 common items from R2)
Cognitive Test R4 (G2)	Maths-30 items and Vietnamese-30 items (M: 12 common items ; V: 15 common items from R3)
Non-Cognitive Test R4 (G2 and G3)	EF, SPP, Intrinsic vs.Extrinsic Motivation Scales, CBST
Cognitive Test R5 (G3)	Math-30 items and Vietnamese-30 items (M: 11 common items ; V: 17 common items from R4(G2))
Non-Cognitive Test R5 (G3)	SPP, Intrinsic vs.Extrinsic Motivation Scales, CBST

Table A2: Sample Size and Missing Responses

Instrument	Max. Sample	Sample	% Missing
Older Cohort			
Cognitive Tests G2 B	5,186	5,008	3.4%
Cognitive Tests G2 E	5,186	5,012	3.4%
Cognitive Tests G3	5,186	4836	8.0%
Student Interview G2	1,654	1,654	0%
Student Non-Cognitive Tests G2	1,654	1,654	0%
Student Non-Cognitive Tests G3	1,654	1,591	3.8%
Parent Interview G2	1,654	1,628	3.87%
Teacher Interview G2	277	274	1%
Teacher Interview G3	277	268	3.2%
Younger Cohort			
Cognitive Tests G2 B	5,257	5,140	2.2%
Cognitive Tests G2 E	5,257	5,169	1.6%
Cognitive Tests G3	5,257	4,990	5.0%
Student Interview G2	1674	1,673	0%
Student Non-Cognitive Tests G2	1,674	1,665	0.5%
Student Non-Cognitive Tests G3	1,674	1594	4.7%
Parent Interview G2	1,674	1,638	2.1%
Teacher Interview G2	277	275	0.7%
Teacher Interview G3	278	273	2.0%

A.2 Non-Cognitive Assessments

A.2.1 Selection of Non-cognitive measurses

Selected domains of non-cognitive skills were chosen to address key questions such as: 'What are the effect of non-cognitive skills on the acquisition of cognitive skills?' and 'What are the effects of individual teachers on the acquisition of non-cognitive skills relative to the prior ability of students and parental investments?'. Therefore, domains were chosen based on their importance to learning in the classroom environment and the potential for these domains to be influenced by teachers.

Executive functioning skills are predictors of long-term outcomes such as schooling achievement (Borella, Carretti, and Pelegrina, 2010), health (Will Crescioni et al., 2011) and labour market success (Bailey, 2007). In the context of educational attainment executive functioning is essential. Students use executive functioning skills to maintain concentration in class, consolidate taught material and apply it to complex problems. Children with low levels of executive functioning will be unable to focus and can disrupt the learning environment of their peers. The RISE Vietnam Research Project aims to quantify the impact of different attributes of individual teachers on the executive functioning of their students in order to investigate which specific teacher attributes contribute to development of executive functioning skills.

Self-esteem is a key non-cognitive variable predictive of long term life outcomes including academic attainment, lifetime earnings, employment and the likelihood of engaging in risky behaviours and hence the self perception profile measure was selected. This measure can be used to assess teacher effectiveness in terms of non-cognitive outcomes and can also be a control variable in terms of cognitive outcomes. Self-characterization has been proved to change with age and with the area of a child's life. Differentiation starts when children approach middle childhood in areas such as scholastic competence, social competence, physical appearance and behavioural conduct.

The Intrinsic and Extrinsic Motivation instrument was selected, as an individual's motivation is a determinant of cognitive outcomes since it is related to the effort children exert and their ability to conduct independent study. In addition, the teacher's non-cognitive value-added effect on student's motivation can be computed to investigate the influence of teaching on students' motivation.

A.2.2 Domains measured by non-cognitive assessments

The domains measured by the assessments are executive functioning, self-perception and motivation in the classroom.

- **Self Perception Profile**

Self-esteem is related to competences such as self-managed learning, self-assert, career self-orientation, self-study, social acceptance, communication and cooperation. In the tests, the specific domains of scholastic competence, social acceptance and physical appearance were assessed.
The studies to validate the instrument have been conducted in different states of United States: Colorado, New York, California and Connecticut (Harter, 2012; Marchant, 1991). In Colorado it was applied to lower middle and upper middle class children where most of them were Caucasian, white. For reliability, we depend on Cronbach alpha to estimate the internal consistency of the scales, as done by most papers. Cronbach's alpha ranges from .80 to .85 for scholastic competence, .75 to .80 for social acceptance and 76 to .82 for physical appearance.

- **Intrinsic vs. Extrinsic Motivation** The sub-scale measured in our assessment is independent mastery vs. dependence on teacher. It assesses children's attitudes toward learning and mastery in classroom. The study used to build the instrument was administered to 3000 pupils in different phases focusing mainly of grade 3 to 6 in United States (Harter, 1981). Previous studies have used modified versions of Harter's scale to identify the relationship between intrinsic motivation, extrinsic motivation and achievement. The study by Lemos and Verıssimo, 2014 in Portugal found that both type of motivation co-exist in elementary school and that intrinsic motivation was positive related with children achievement opposite to the relationship observed between extrinsic motivation and student's achievement. For reliability, we have relied upon internal consistency indices (i.e., Cronbach's alpha).

- **Executive functioning** Executive functioning is made up of three domains-inhibitory control, working memory, and cognitive flexibility. Together these cognitive skills are required for reasoning, problem solving, and planning. The measures we use are Hearts and Flowers and Backward Digit Span Task. Cronbach's alpha measures the reliability of these instruments, which is the extent to which individual items are measuring the same construct.
  The HF task (Diamond and Wright, 2014), is designed to assess inhibitory control and cognitive flexibility skills and has been widely used and validated with elementary school students in US and Canada (Davidson et al., 2006;Obradović et al., 2018; Oberle and Schonert-Reichl, 2013).
  The Backward Digit Span is a standard measure of working memory drawn where respondents are required to repeat a series of numbers back to the assessor in reverse order. It has been commonly used in middle childhood and also children with ADHD, dyslexia (Blankenship et al., 2015; Rosenthal et al., 2006).

A.3 Test Scoring

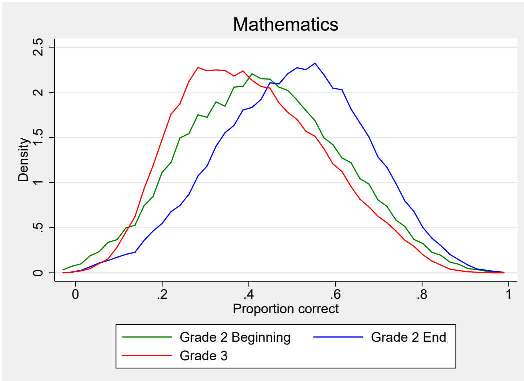
[h] The cognitive tests are the primary outcome variables which will be used to measure student learning. Below we describe the distribution of the cognitive test scores and an analysis of the quality of the test data using Item Response Theory.

All items administered under Maths and Literature tests were multiple choice questions, responses to which were coded as "1" for correct and "0" for incorrect. Table A3 shows the raw cognitive test scores, and Figure 1 their distributions. We see that the raw test scores exhibit adequate variation. The distributions of the raw Maths test scores are approximately normal for all grades in both the cohorts. The high density to the right in older cohort's Vietnamese Grade 2 end of the year test indicates that the test was too easy. The distribution of the younger cohort's Grade 3 Vietnamese test scores have a similar shape, but it is less pronounced.

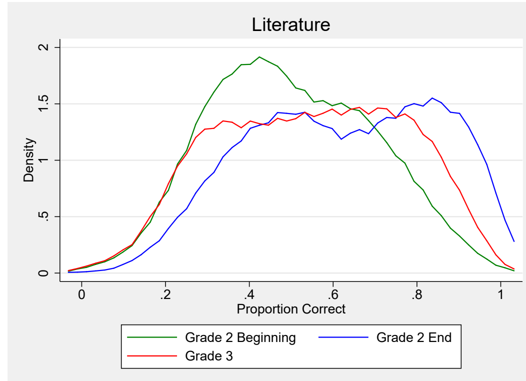
Table A3: Raw Cognitive Test Scores

	Mean	Std. Dev	Min	Max	N
Older Cohort					
Total Math G2 B	10.83	4.40	0	24	5008
Total Math G2 E	15.01	5.00	0	29	5041
Total Math G3	12.22	4.85	1	27	4836
Total Lit G2 B	12.52	4.86	0	25	5031
Total Lit G2 E	18.75	6.64	0	30	5012
Total Lit G3	16.39	6.55	0	30	4846
Younger Cohort					
Total Math G2 B	10.86	4.59	0	24	5159
Total Math G2 E	13.28	5.48	0	29	5148
Total Math G3	13.07	5.18	0	28	4961
Total Lit G2 B	14.08	6.21	0	30	5140
Total Lit G2 E	14.80	6.75	0	30	5149
Total Lit G3	16.54	6.60	0	30	4970

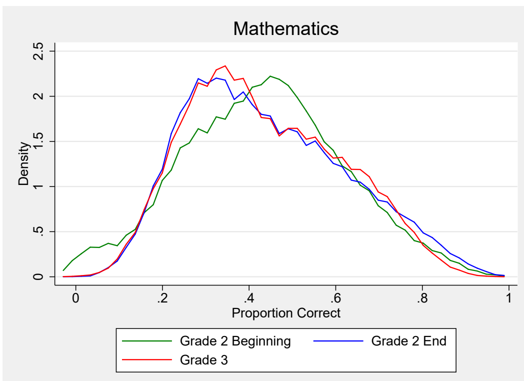
The raw test scores however are not comparable across grades and cohorts as different questions were administered in each test. There were however some common (linking) items that were included in order to put students' performance on these tests on a common scale, using item response theory (IRT).



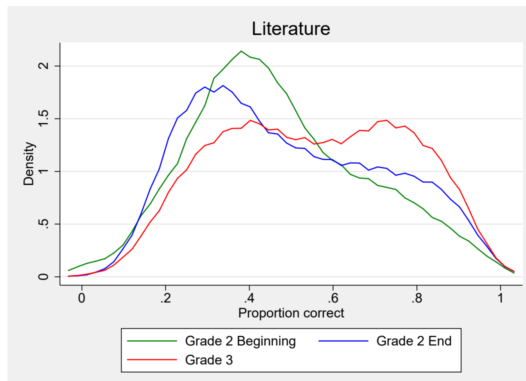
Older Cohort



Older Cohort



Younger Cohort



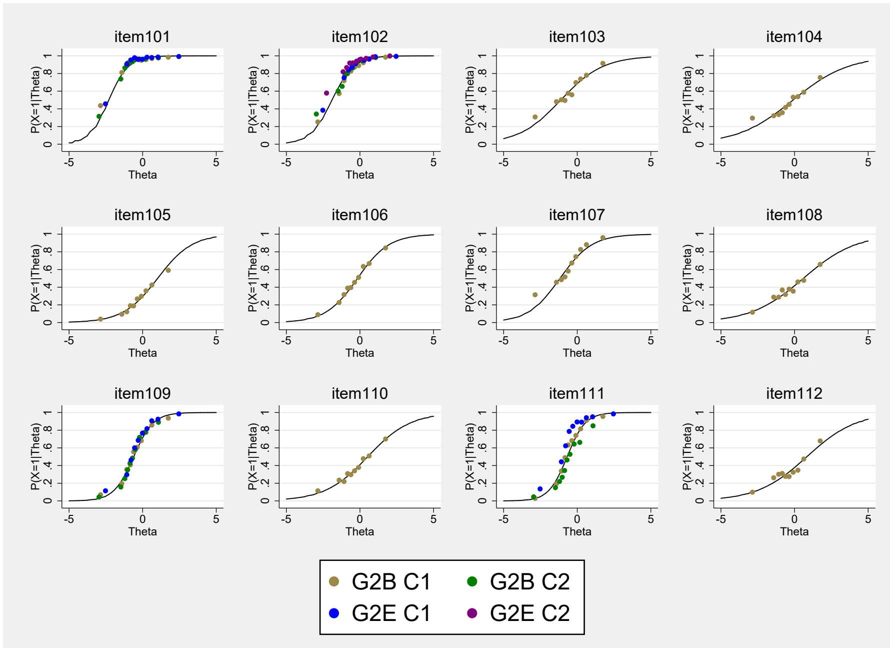
Younger Cohort

Figure 1: Distribution of raw percentage correct scores

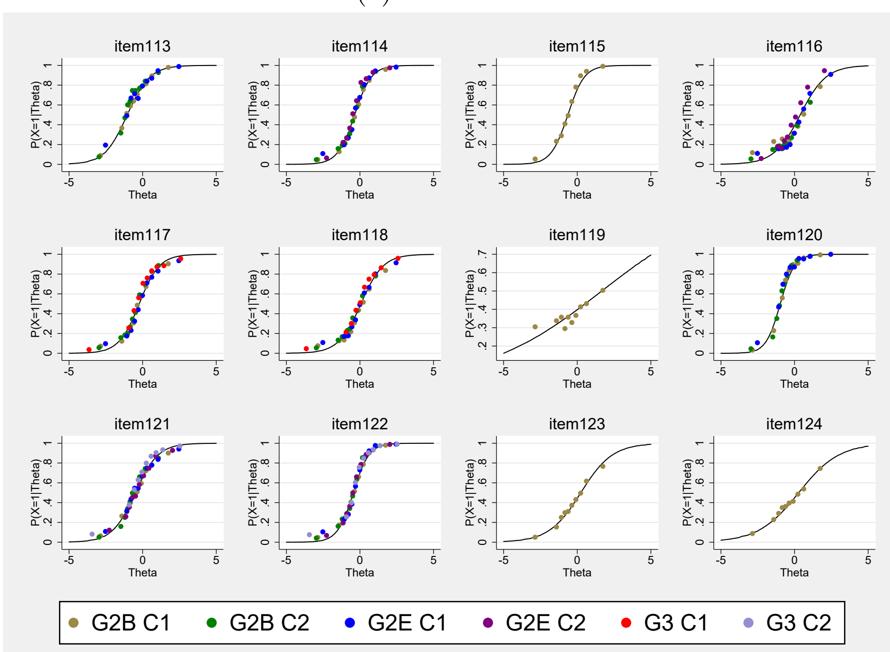
Two parameter IRT models were fitted on the data to evaluate the ability of the tests to measure the intended latent trait and individual item functioning.To compare students across grades and across the two cohorts, all test results were combined into a single data set (separately for Math and Vietnamese) and IRT analysis was used to construct latent measures of mathematics and Vietnamese ability for each student that are comparable across grades and across the two cohorts. For ease of interpretation, these Math and Vietnamese latent scores were normalized to have a standard deviation on one and mean of zero within a grade.

In Figures 3 and 5 we plot the item characteristics curve (ICCs) for each individual item (for both Math and Vietnamese) , i.e., plotting the probability of attaining the correct answer against theta, along with the empirical fit (the dots which indicate the empirical deciles of theta). We then get rid of the items with poor performance and identify linking items with Differential Item functioning(DIF) (items for which different subgroups have different probability of answering an item correctly, even after controlling for level of ability).

A good performing item is the one that generally lie in the -2.5 to 2.5 difficulty range and that which helps discriminate between low and high ability students. In order to assess the performance of the items, we check the estimated fit of the data to the estimated item characteristics. In general, the fit of the items are good and they don't exhibit DIF. We remove the poor performing items and for the few that do exhibit DIF, we generate a new item for each subgroup with different behaviour.

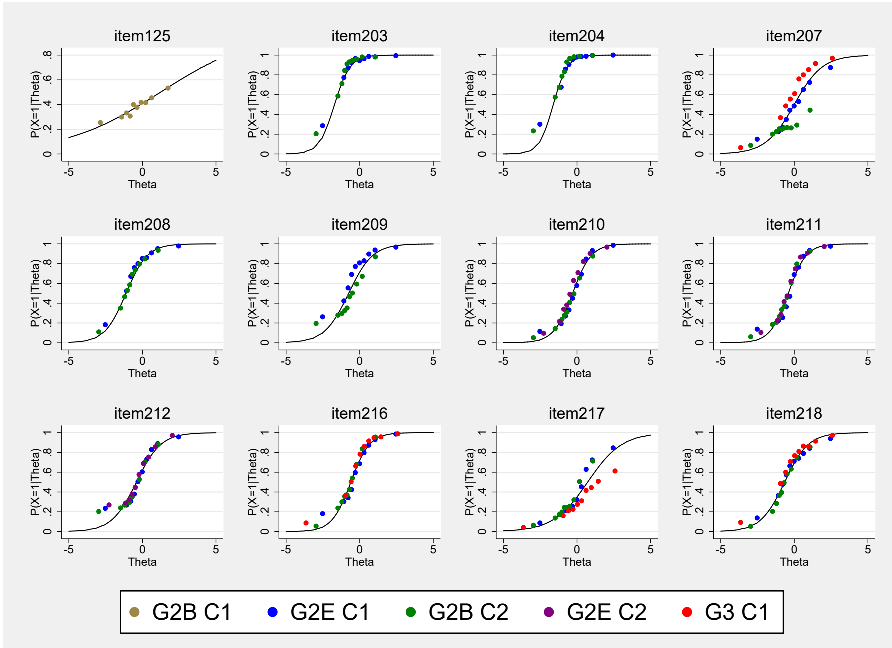


(a) Vietnamese

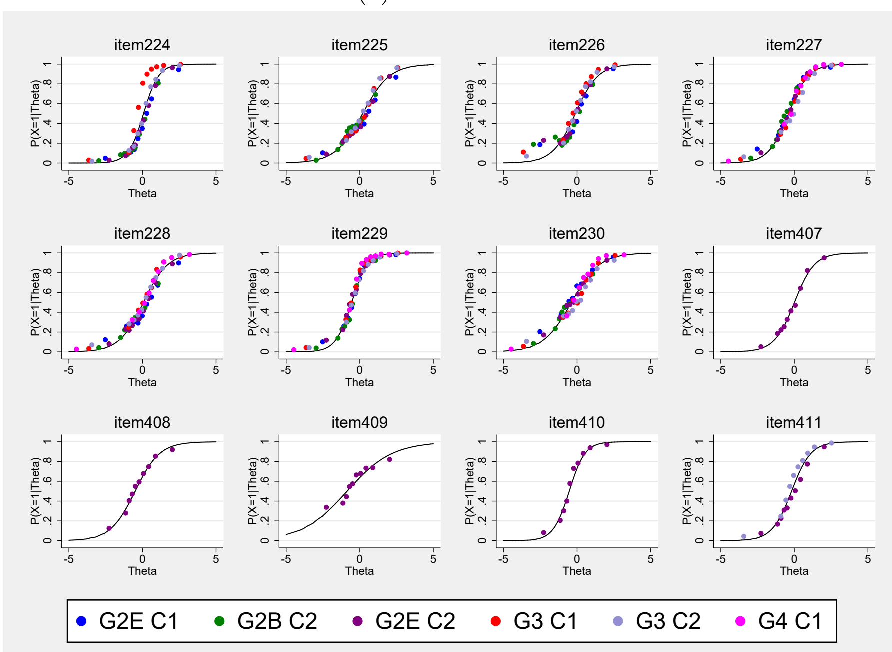


(b) Vietnamese

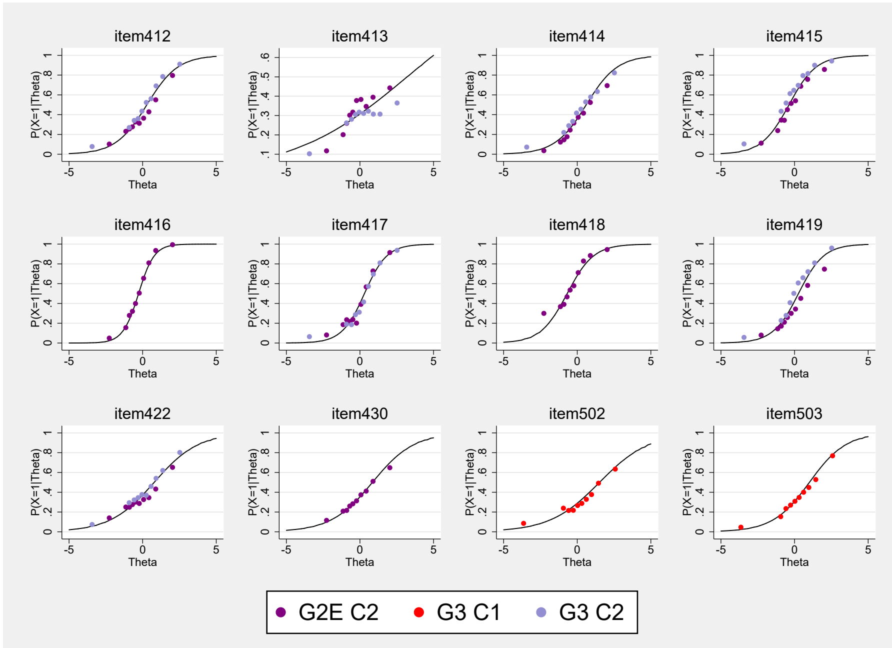
Figure 3: Item Characteristics Curve - Vietnamese



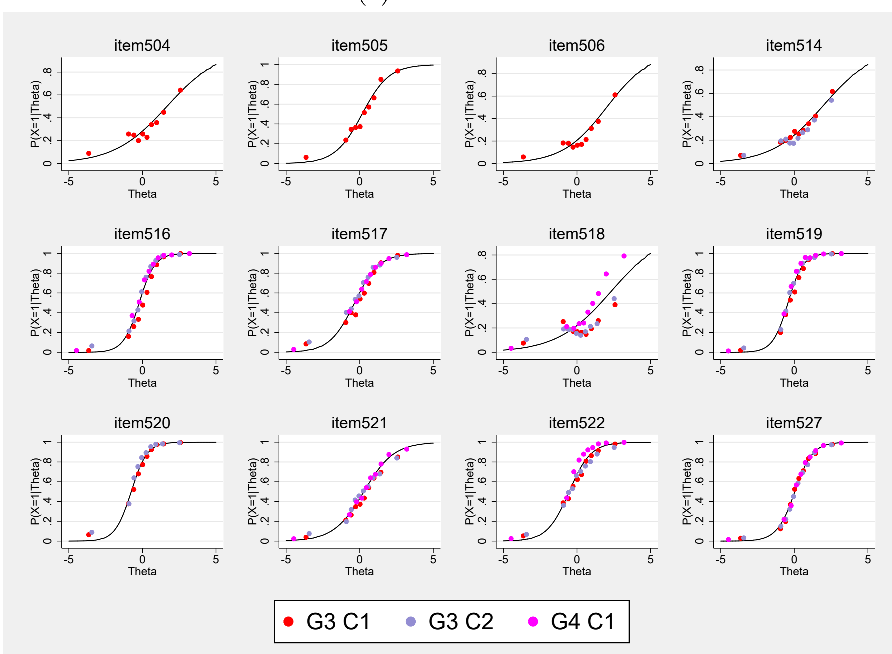
(c) Vietnamese



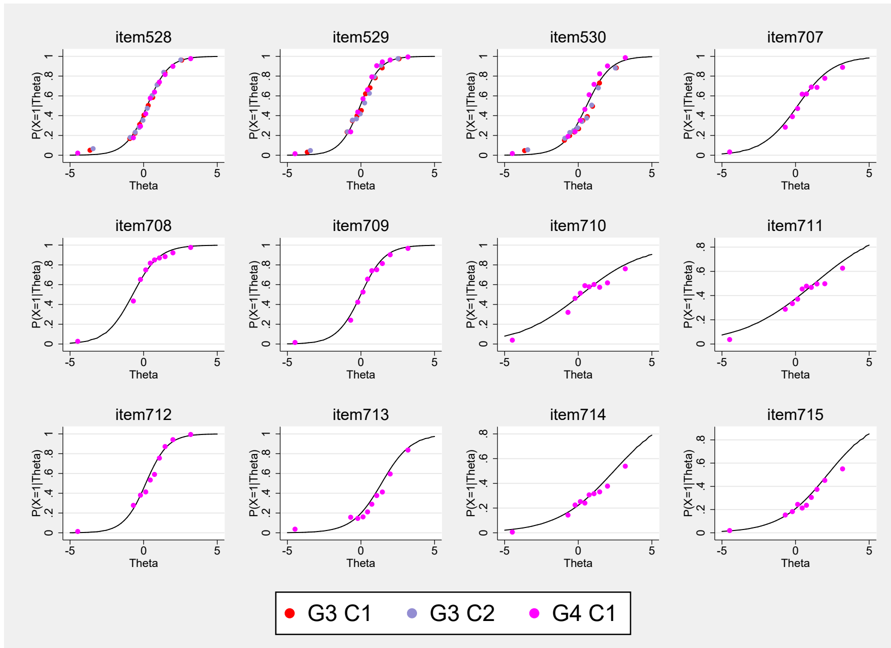
(d) Vietnamese



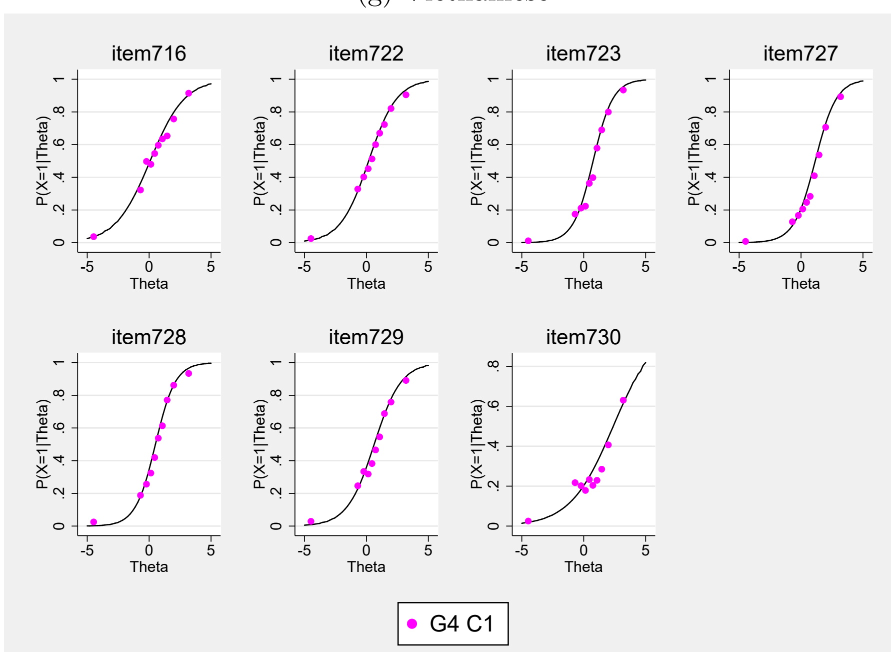
(e) Vietnamese



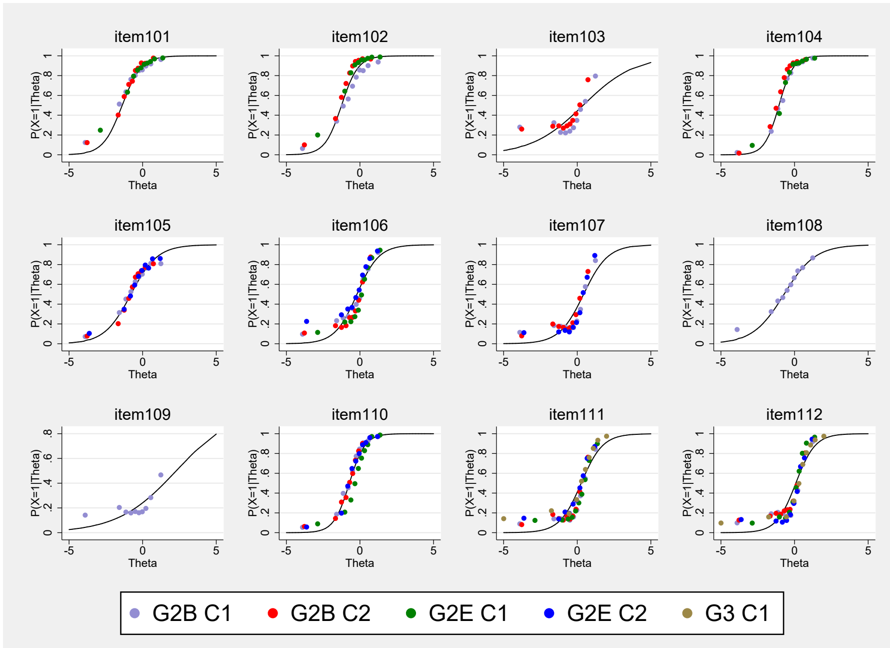
(f) Vietnamese



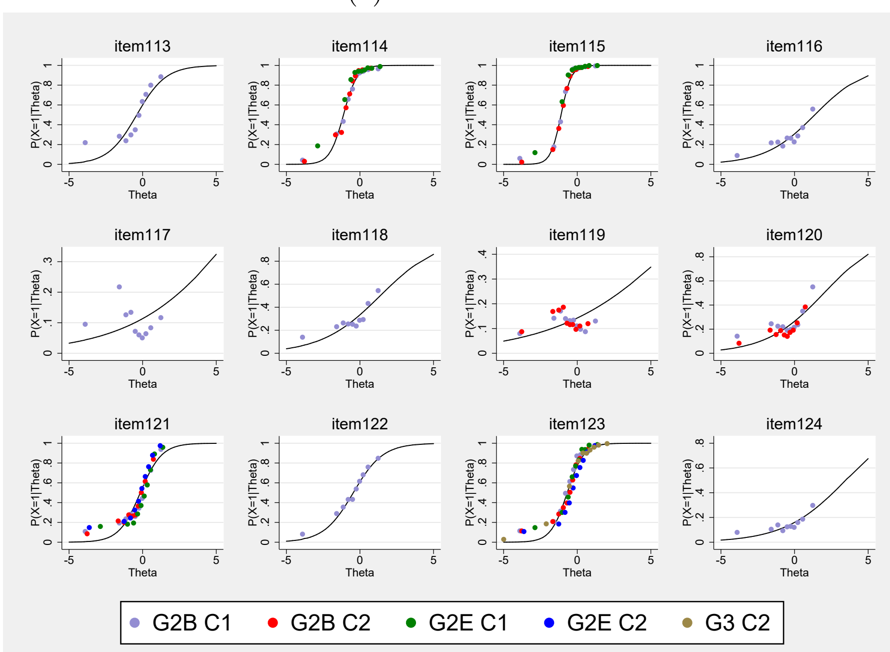
(g) Vietnamese



(h) Vietnamese

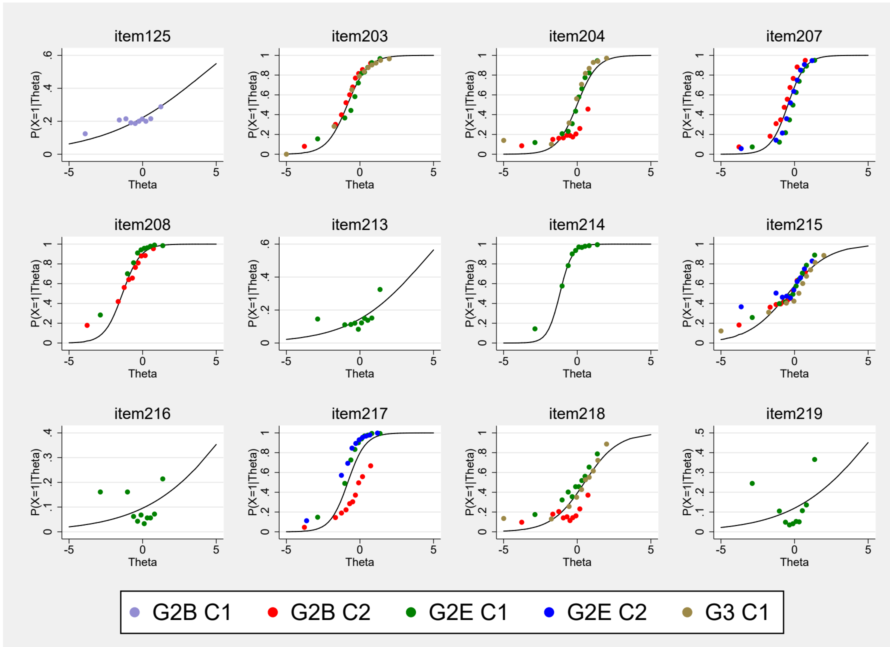


(a) Mathematics

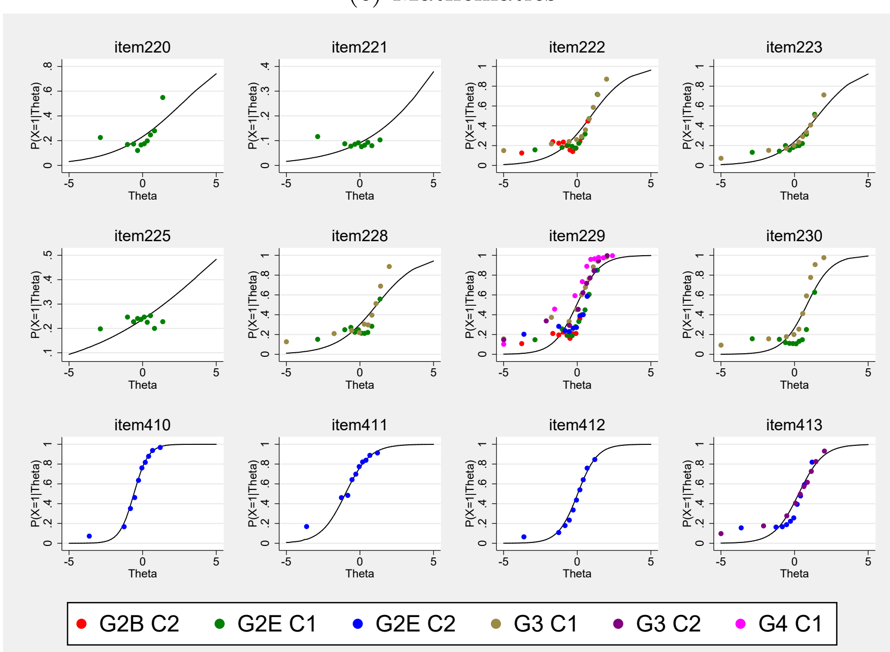


(b) Mathematics

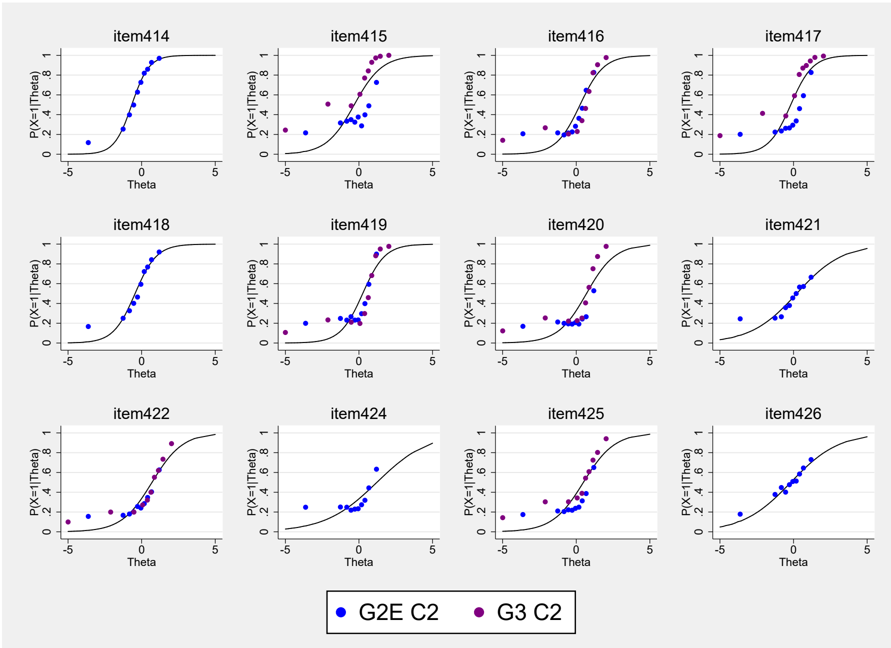
Figure 5: Item Characteristics Curve - Mathematics



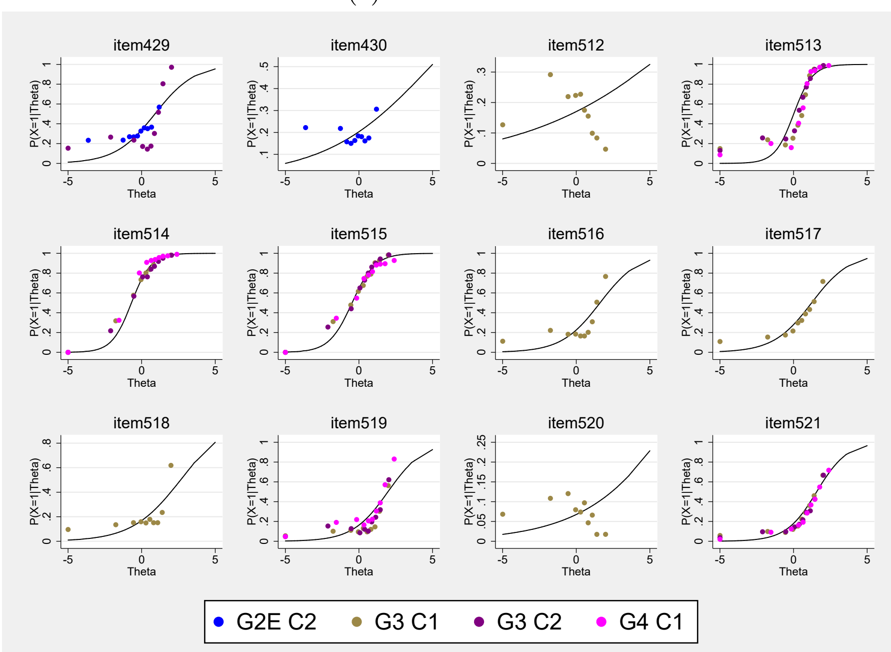
(c) Mathematics



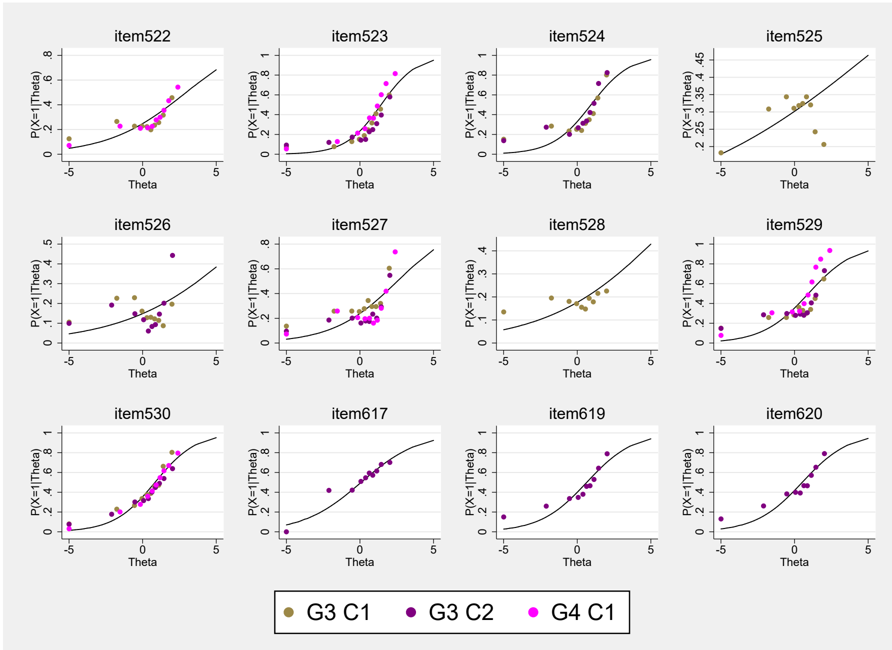
(d) Mathematics



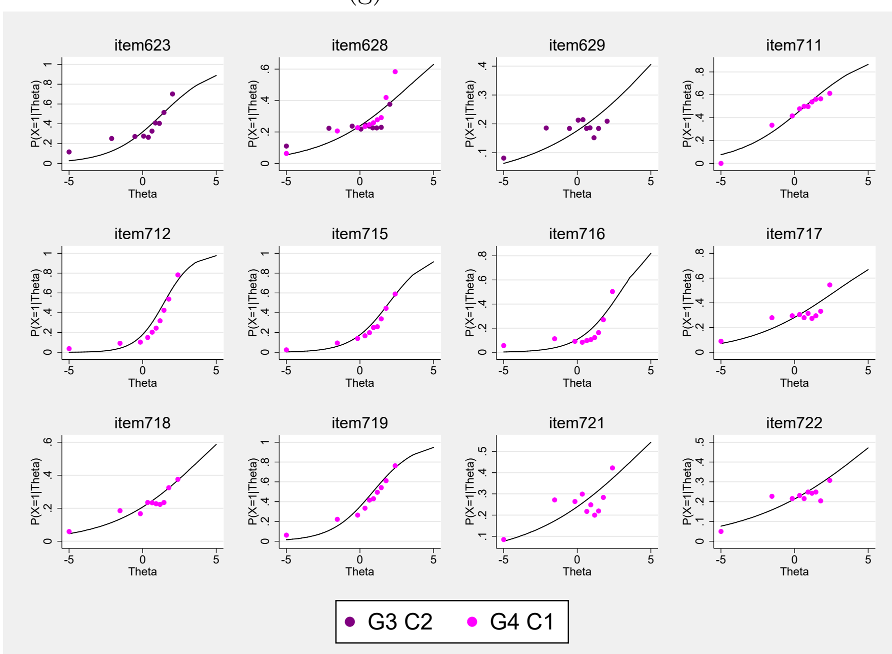
(e) Mathematics



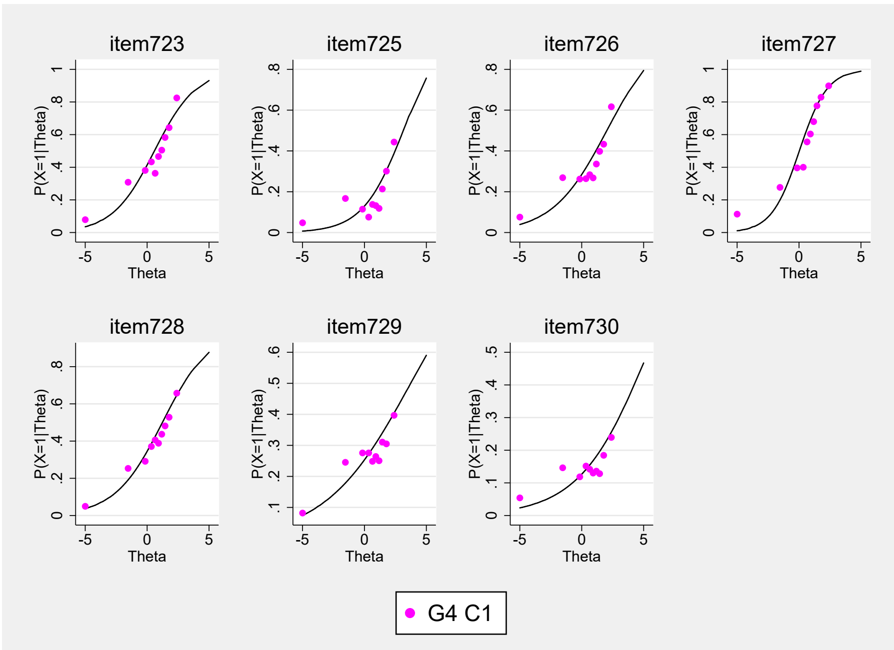
(f) Mathematics



(g) Mathematics



(h) Mathematics



(i) Mathematics