RISE
RESEARCH ON IMPROVING SYSTEMS OF EDUCATION

WORKING PAPER
November 2018

Indonesia Got Schooled: 15 Years of Rising Enrolment and Flat Learning Profiles

Amanda Beatty, Emilie Berkhout, Luhur Bima, Thomas Coen, Menno Pradhan, Daniel Suryadarma

RISE-WP-18/026

Funded by:

Indonesia Got Schooled: 15 Years of Rising Enrolment and Flat Learning Profiles

Amanda Beatty[*], Emilie Berkhout[^], Luhur Bima[#], Thomas Coen[*], Menno Pradhan[^], Daniel Suryadarma[#]

[*]***Mathematica***

[^]***AIGHD***

[#]***The SMERU Research Institute***

RISE Working Paper

November 2018

Contents

Tables

Figures

Abstract

Indonesia has instituted wide-ranging educational reforms over the past twenty years, but recent international assessments of student learning indicate that these reforms may not have translated into learning gains—the country is performing comparatively poorly and worse than its regional neighbours. To examine the relationship between schooling completed and learning gains, and how that changed over time, we developed learning profiles using five rounds of data from the Indonesian Family Life Survey (IFLS). We show that Indonesia has succeeded in achieving high levels of school enrolment and attainment, with particular gains concentrated in junior secondary and senior secondary school between 2000 and 2014. However, we also find a large gap between students' mathematical ability and what they are supposed to know based on the education curriculum. Absolute learning levels as well as marginal learning levels are low, meaning that students are learning little as they are promoted from grade to grade. Even high school graduates struggle to correctly answer numeracy problems that they should have mastered in primary school. We also find that learning is decreasing slightly over time. We extend our analysis by identifying characteristics of children who are educationally left behind: children who are performing particularly poorly compared to their peers. Children with low numeracy levels are more likely to live in Eastern Indonesia, in rural areas, and be older and male. Our findings, albeit limited to a narrow set of test items, demonstrate the incredibly slow pace of learning occurring throughout Indonesia, and reiterate the importance of focusing system reforms on learning progress.

Keywords: education, schooling, learning, numeracy, Indonesia

1. Introduction

Over the past twenty years, Indonesia has made dramatic progress in increasing access to education and has also attempted to improve school quality. In 2001, the government decentralised policy control, including for education policy, to allow districts to shape how education is delivered and adjust policy for local context and needs. While the central government retains authority over the education curriculum, assessments, and school accreditation, other aspects of education policy, including personnel management, schooling infrastructure, and allocation of budgetary funds, are now under the purview of sub-national governments. In 2002, the Constitution was amended to require that 20% of the budget be allocated to education spending. And in 2005, the government passed a new teacher certification policy that instituted higher standards for new and existing teachers connected to compensation.

Despite these reforms that increased educational resources, adjusted policy incentives, and increased school access, the quality of education in Indonesia appears to have only slightly improved. The country continues to rank near the bottom of international education assessments. The 2015 Programme for International Student Assessment (PISA) ranked Indonesia 64th out of 72 countries that participated in the test.[1] Even with an improvement in PISA's results from 2012 to 2015, Indonesia is still ranked below neighbouring countries. The Trends in International Mathematics and Science Study (TIMSS) assessment in 2015 showed similar results with Indonesia ranking among the lower-achieving countries.

This paper seeks to unpack the learning crisis in Indonesia by analysing what children are learning in school and retaining as young adults. We examine learning by school- and grade-level, and also assess how learning varies across gender, economic status, and geographic location. To do this, we develop a set of descriptive mathematics learning profiles that analyses changes in children's math skills as they progress through school and become older (Pritchett and Beatty, 2015). To our knowledge, this paper is one of only two studies (see also Afkar, de Ree and Khairina, forthcoming) looking at learning accumulation in Indonesia over time. Afkar, de Ree and Khairina (forthcoming) examine changes in math learning as children progress through primary and junior secondary school, using ten math anchor items that were similar across grades. They find that approximately 40% of students do not master basic numeracy questions after three years in school and that in many schools, learning does not keep up with curriculum expectations.

Our focus on learning profiles draws on a body of literature from other countries, which is dominated by cross-sectional data. Pratham, a non-governmental organisation in India, began collecting yearly cross-sectional data on student learning throughout India at the household level in 2005. This initiative led to the creation of the ASER (Annual Status of Education Report) Centre, responsible for over ten years of such assessments, and inspired many other similar student assessment systems, such as EGRA/MA (Early Grade Reading Assessment and Early Grade Math Assessment) and Uwezo that were used to create the first examples of learning profiles in developing countries (Banerji, Bhattacharjea and Wadhwa, 2013). For example, Jones et al. (2014) using Uwezo data showed that in Kenya, Tanzania, and Uganda more than half of 10-year-olds and one-third of 13-year-olds could not recognise a single written word or recognise numbers. Spaull and Kotze (2015) showed grave learning deficits in South Africa, also utilizing learning profiles. Pritchett and Beatty (2015) used ASER data to illustrate the concept of learning profiles and incongruence between curriculum pace and actual student learning.

[1] Note that most countries that participate in the PISA are developed countries.

In a cross-country context, several papers originating from Young Lives, a child-level panel data study utilizing similar questions from Ethiopia, India, Peru, and Vietnam, demonstrated vast differences in learning gains over time across countries using learning profiles (Rolleston, 2014; Rolleston and James, 2015; Singh, 2017). Others construct learning profiles from adult survey data. For example, Kaffenberger and Pritchett (2017) created learning profiles across ten countries using Financial Inclusion Insight data with young adults ages 18 to 37 and Pritchett and Sandefur (2017) used DHS literacy data from women aged 25 to 34 in 51 countries. Crouch and Gustafsson (2018) examined learning over time using international assessments like TIMSS, SACMEQ (Southern and Eastern Africa Consortium for Monitoring Education Quality), and PISA, but focus on how inequality in cognitive outcomes is changing within and across countries.

To construct the learning profiles for Indonesia, we use mathematics test data from four waves of the Indonesian Family Life Survey (IFLS), a longitudinal survey that is representative for 83% of the Indonesian population and was administered in 1993, 2000, 2007, and 2014. We also analyse IFLS East test data from 2012, a cross-sectional household survey covering seven provinces in Eastern Indonesia. As there are two versions of the numeracy test—an easy version and a more difficult version—we also apply a test equating procedure using Item Response Theory (IRT) to generate a measure of numeracy skills that is comparable between the two versions of the test and adjusts for question difficulty.

We find that Indonesia has succeeded in achieving high levels of school enrolment and attainment. Although by 2000 Indonesia already achieved near universal enrolment in primary school (PS, Grades 1 to 6), the country saw large enrolment gains in junior secondary school (JSS, Grades 7 to 9) and senior secondary school (SSS, Grades 10 to 12) between 2000 and 2014. By 2014, more than 93% of students who start a level of schooling complete that level. However, those gains in enrolment and attainment have not been accompanied by meaningful improvements in student learning. There is a gap between what students are supposed to learn and what they are actually learning. Only a small portion (11%) of individuals who graduated from SSS were able to answer all numeracy problems that should have been mastered by fourth grade. Students have low marginal gains in numeracy as they advance through the education system, meaning that students are learning little as they are promoted from grade to grade. We also find that learning levels have decreased slightly over time. The 2000 sample achieved higher learning outcomes compared to the 2014 sample and these differences are similar across subgroups. Our results are descriptive and do not take into account selection as students are promoted through the school system. Since the literature shows that the decision to drop-out of school is correlated with student performance (Hanushek, Lavy and Hitomi, 2008), a causal learning profile would likely show even smaller learning gains than what we currently find.

We find some characteristics are associated with low numeracy in Indonesia. Our estimation results show that individuals who are older and male have a higher probability of low numeracy. We also find that the numeracy level is not influenced much by wealth. The estimation reveals that individuals with higher education attainment are less likely to have low numeracy. However, the highest educational benefits to learning occurs during the first five years of schooling.

This study contributes to existing literature by enriching the discussion on learning profiles of developing countries. More importantly, the IFLS allows us to examine how the learning profile has changed over a period of fourteen years. To our knowledge, no other studies in the literature has been able to do this. Furthermore, this paper helps policymakers in Indonesia understand what students are actually learning and to inform policy decisions to improve the education system.

The rest of the paper is organised as follows. We begin by describing our data sources and analytical approaches (section 2). We then present results on educational enrolment, attainment, and school types (section 3). Section 4 details our key findings of the learning profiles for Indonesia by currently enrolled grade-level, recently completed education-level, and changes over time. Next, we investigate the characteristics of those with low numeracy (section 5) before summarizing our findings and concluding the paper (section 6).

2. Data and Methodology

2.1. Data

We use data from the first wave and last three waves of IFLS (Frankenberg et al., 1995; Strauss et al., 2004, 2009; Strauss, Witoelar and Sikoki, 2016) as well as the IFLS East (Sikoki et al., 2013) to produce the learning profiles. The IFLS is a panel dataset that is representative for 83% of the Indonesian population, containing over 30,000 individuals living in 13 of the 27 provinces in the country. It covers four provinces in Sumatra (North Sumatra, West Sumatra, South Sumatra, and Lampung), all five of the Javanese provinces (DKI Jakarta, West Java, Central Java, DI Yogyakarta, and East Java), and four provinces on the remaining major island groups (Bali, West Nusa Tenggara, South Kalimantan, and South Sulawesi).

The IFLS waves used for our analysis took place in 1993 (IFLS 1), 2000 (IFLS 3), 2007 (IFLS 4), 2012 (IFLS East), and 2014 (IFLS 5). The households from the first wave in 1993 and their split offs are tracked over time. A split off household arises when part of an original household forms a new household, potentially with new household members. For instance, when a son gets married and moves out of his parents' home he forms a split off household with his wife. Among the main IFLS 1 respondents, the re-contact rate is 82% in IFLS 5. IFLS East, meanwhile, is a 2012 cross-sectional household survey of around 10,000 individuals in 2,500 households living in seven provinces in Eastern Indonesia: East Nusa Tenggara, East Kalimantan, Southeast Sulawesi, Maluku, North Maluku, West Papua, and Papua. These provinces are less developed and thus important to consider for a comprehensive picture of learning in Indonesia. We only use IFLS East in section 5 of this paper.

IFLS collects information at the community, household, and individual level. Key variables that we use are school enrolment, educational attainment, age, gender, household assets, and region. To measure learning, we use a numeracy test that was administered in the third, fourth, and fifth wave of the IFLS. There are two versions of the test: an easy version (version 1) and a more difficult version (version 2). Both versions contain five multiple choice questions as shown in Table 2.1. The first three questions of version 1 have three answer options, while the other questions have four. There is one overlapping question (56/84). The items reflect numeracy skills on addition, subtraction, fractions and percentages. Although the scope of the test is limited, we would expect the respondents to gain these skills during their basic education.

Table 2.1 IFLS numeracy questions

Version 1	Version 2
Questions	
EK13: 49-23 a. 25 b. 26 c. 27	EK18: $\frac{56}{84}$ a. 5/7 c. 3/4 b. 2/3 d. 5/6
EK14: 267+112-189 a. 180 b. 188 c. 190	EK19: (412+213)/(243-118) a. 125 c. 25 b. 75 d. 5
EK15: (8+9)*3 a. 34 b. 45 c. 51	EK20: 0.76-0.4-0.23 a. 0.11 c. 0.13 b. 0.12 d. 0.16
EK16: 56/84 a. 5/7 c. 3/4 b. 2/3 d. 5/6	EK21: (100-65)% of 160 million (in text) a. 35 million c. 48 million b. 40 million d. 56 million
EK17: 1/3-1/6 a. 2/3 c. 1/6 b. 1/3 d. 1/9	EK22: 5% interest on Rp75,000 (in text) a. Rp7,500.00 c. Rp750,00 b. Rp3,750.00 d. Rp375,00
Administered to	
7-14 years old	15-24 years old
15 years old and above that participated in version 1 in the previous wave (IFLS 4 and 5 only)	24-35 years old if participated in version 2 of previous wave (IFLS 4)
	24-60 years old (IFLS 5)

Source: IFLS 3, 4, East and 5
Note: Question numbers are taken from the IFLS questionnaire.

The age of the respondent determines the version that s/he had to answer, but the age range differs between the waves. In IFLS 3, all respondents between 7 and 14 years old filled in the first version, while respondents between 15 and 24 years old filled in the second version.[2] The same was done in IFLS 4 and IFLS 5, but additional age cohorts were included as summarised in Table 2.1. First, those who participated in the second version of the test in IFLS 3 also participated in the second version of the test in IFLS 4, as long as they were under 35 years in age. Second, the age range for the second version of the test was extended to 60 years old in IFLS 5. Finally, both in IFLS 4 and 5, the first version of the test was re-administered to those respondents above 15 years old that had taken the easier version of the test in the previous wave. Hence, 15- to 21-year-old panel respondents in IFLS 4 and 15- to 28-year-old panel respondents in IFLS 5 answered both versions of the test. In this paper, we focus on the age cohorts between 7 and 30 years old in 2014 and age cohorts between 7 and 24 years old in 2000 and 2007, because the samples in 2000 and 2007 are not representative for older cohorts. The percentage of those age cohorts above 15 years old that answered both versions of the test is 60.3% in 2007 and 51% in 2014.

The results for the numeracy test that contain missing values and missing data patterns are systematic. Non-response is higher among the youngest and oldest cohorts and among those who live on other islands than Java-Bali. Moreover, it negatively correlates with years of education and wealth. We also find that the share of missing values generally increases as the question difficulty increases, measured by the grade in which the items are expected to be mastered according to the curriculum, and that the highest share of missing values is concentrated among the youngest respondents (results available upon request). This provides evidence that the missing value patterns are associated with low skills and so we infer that respondents likely left these questions blank because they didn't know the answer. Because leaving these values out of our analysis would bias the results, we impute the missing items as if the respondent applied random guessing. We exclude from the study those individuals for whom the complete numeracy test is missing because they refused, could not be contacted, did not have enough time, or any other reason unrelated to competencies. We also exclude those individuals for whom educational attainment is missing.[3] Descriptive results by item are presented in Appendix 1: Tables. The results with and without imputations are mainly similar for students in Grades 4 through 12. We find the largest differences for students in Grade 1, which had the largest number of missing values and where the un-imputed results seemed artificially high. Imputing by guessing substantially drops the percent correct for students in Grade 1. We see a small decrease in the percent correct when imputing by guessing for students in Grades 2 and 3.

Table 2.2 shows the sampling results for the numeracy test. The table splits the age cohorts in two groups to show the sample sizes for each version of the test separately. The bottom row of the table presents the percentage of the numeracy scores, which is described in the next section that were generated with at least one imputed item.

[2] This is the first wave that includes the numeracy test.
[3] Educational attainment is only missing for 0.5% of the sample.

Table 2.2 Numeracy question sample sizes

Number of observations	2000		2007		2012 (EAST)		2014	
	7–14	15–24	7–14	15–24	7–14	15–24	7–14	15–30
Individuals in sampled households	6805	9144	7306	8506	2221	1505	8798	14,329
Respondents interviewed	6176	7968	6702	7537	2216	1499	8165	13,058
Respondents with numeracy score	5973	7411	6466	6975	2116	1303	7776	11615
Percentage generated with at least one imputed item	22.6	20.4	18.9	14.8	10.2	7.1	18.1	8.5

Source: IFLS 3, 4, East and 5
Note: 'Individuals in sampled households' presents the total number of household members in the sampled households, and 'respondents interviewed' shows how many of those household members were interviewed. 'Respondents with numeracy score' shows the number of interviewed respondents for which we could generate a numeracy score and for whom educational attainment is not missing.

2.2. Methodology

We use Item Response Theory (IRT) to create a link between the two tests and to generate a score for numeracy skills that is comparable between the two versions of the test. To link the test, we employed a horizontal test equating procedure using the group of respondents that answered both versions, that we called anchor respondents. For IFLS East, we combine the data with IFLS 5 before implementing the IRT, as none of the IFLS East respondents took both versions of the test.

Responses from the anchor respondents generated the difficulty level and discrimination power of each of the ten items, which are comparable between tests. In Table 2.1 we can see that there is an item in version EK 1 (56/84) that is almost identical to an item in EK2 ($\frac{56}{84}$). We chose to treat the overlapping question as separate questions in each version because a third of the respondents that answered both versions gave two different answers.

To estimate each respondent's IRT score, we use a two-parameter logistic model (two parameters to characterize the distribution, one for student ability), which includes parameters for discrimination, difficulty, and ability. The difficulty parameter relates to the ability of an individual, such that if the difficulty parameter is equal to the ability parameter, the individual is equally likely to answer correctly or incorrectly. The discrimination parameter reflects how fast the probability of success changes with ability near the item difficulty. The higher the discrimination parameter, the better the item can differentiate high ability students with those with low ability. Putting these parameters in a formula, the probability of person *j* providing a positive answer to item *i* is given by

$$\Pr\left(Y_{ij} = 1 \middle| \theta_j\right) = \frac{\exp\{\alpha_i(\theta_j - b_i)\}}{1 + \exp\{\alpha_i(\theta_j - b_i)\}} \qquad \theta_j \sim N(0,1) \qquad (1)$$

where α_i represents the discrimination of item i, b_i represents the difficulty of item i, and θ_j is the latent trait (or ability) of person j (StataCorp, 2017). IRT estimates these three parameters using maximum likelihood.

We predict the probability that someone gives the correct answer for each item and take the mean of these probabilities to calculate their numeracy score. The score can be interpreted as the probability of answering any of the items correctly. Other studies often present the ability parameter. We chose to present the mean predicted probability of answering correctly for easier interpretation of the results. However, results would look similar if we presented the ability parameter.

The score aims to measure numeracy skills. All presented items are about only one domain of numeracy skills, which is number operation. Though numeracy skills also comprise other domains, such as geometry and statistics, number operation skills are the foundation for solving more complex numeracy problems. Moreover, number operation is one of the main domains taught during basic education. Therefore, we think that the score based on these items provides a relevant reflection of the numeracy skills of the respondents.

Psychometric properties of the numeracy score are presented in the Appendix 3: Psychometrics of Numeracy Score. We check the validity of the score with factor and infit and outfit analysis, and we examine the reliability using Cronbach's alpha and the IRT discrimination coefficients. In addition, we run tests on the IRT assumptions of unidimensionality, no differential item functioning, and conditional local independence. The validity analysis shows that the numeracy score is unidimensional, contains little noise, and does not follow unexpected patterns, which means that the numeracy score achieves test validity. For reliability, the item-test correlations are good (between 0.46 and 0.63), but the Cronbach's alpha (0.66) and the discrimination coefficients of the relative difficult items (between 0.5 and 0.9) are rather low. The low alpha could mean that the test needs more items to achieve acceptable internal consistency, which is understandable as we only have five or ten items per respondent for the numeracy score. The low discrimination coefficients for the relatively difficult items are not surprising either. Only a small number of respondents answered those items correctly, so the data are limited in distinguishing the ability of the respondents based on those items. However, the coefficients are significantly different from zero, so they still add information on the ability of the respondents to the score. For the IRT assumption analysis we do robustness checks by excluding items from the IRT estimation that might violate the assumptions, but we find that our results are robust to these tests. The appendix contains a more detailed explanation of the methods. Based on these results, we conclude that the test items are adequate for measuring numeracy, even though the limited number and scope of the items pose constraints.

We correct the mean numeracy score for guessing such that, in expectation, a zero is given for those who randomly guessed and a 1 is given for those who knew the correct answer. As the test items are multiple choice, respondents could correctly answer a question by chance alone. We use the following method by Afkar, de Ree and Khairina (forthcoming). If α is the fraction that knows the answer and y is the fraction that answered correctly, then:

$$y = (1 - \alpha) \times \frac{1}{K} + \alpha \times 1 \tag{2}$$

for K answer options. Those who guess have a probability of 1/K to answer correctly, while those who know the answers have a probability of one. In this paper, we present results for the separate items and for the numeracy score. Therefore, the y is either the result for an individual item or the probability of a correct answer as predicted by IRT, both based on actual and imputed values. Because some items have three answer options while others have four, we first correct the predicted probability of a correct answer for each item. The numeracy score is the mean of these corrected probabilities. We present the results for α and weight them using sampling weights. We bottom code negative mean values to 0.

In the next chapters, we show the numeracy score by gender, region, and wealth. For the differences by region, we divide the sample in residents of Java-Bali and residents from other islands, the lowest geographic level at which the IFLS is representative (Rokx et al., 2010). For the differences by wealth, we generate an asset index using Principal Component Analysis (PCA). PCA is a common method for generating socio-economic status indices (Filmer and Pritchett, 2001), and converts the asset data into independent components of which the first one explains the largest amount of variance of the original data. The first component is used as the asset index. The items included in the asset index are dummy variables for ownership of certain types of assets in order to limit the number of missing values. The included assets are house and land, other buildings, non-farm land, poultry, livestock or fishpond, hard stem plant not used for business, vehicles (cars, boats, bicycles, motorbikes), household appliances (radio, television, fridge, etc.), savings or certificate of deposit or stocks, receivables, jewellery, and household furniture and utensils.

We test the significance of the sub-group differences by regressing the numeracy score on subgroup indicators, controlling for highest grade completed and IFLS wave, as shown in the following formula for individual *i* from enumeration area *c*.

$$NUMERACY_{i,c} = \beta_1 + \beta_2 SUBGROUP_{i,c} + \gamma_{wave} + \gamma_{grade} + \varepsilon_{i,c} \qquad (3)$$

In which NUMERACY is the ability that follows from IRT, SUBGROUP is one or multiple dummy variables indicating the subgroups, γ_{wave} and γ_{grade} are wave and grade fixed effects and ε is an error term. The coefficient of interest is β_2. We estimate the model using Ordinary Least Squares (OLS) and the standard errors are corrected for clustering. When we are only interested in the significance of the subgroup differences for 2014, we run this regression on the data from 2014 only without the wave fixed effects.

3. School Types, Enrolment, and Attainment

Primary and secondary education in Indonesia is divided into three levels. For the first six years, students attend a *Sekolah Dasar* or primary school (PS), in which students are generally between the ages of 7 and 12. If students pass a school examination at the end of PS, they move onto three years of *Sekolah Menengah Pertama* or junior secondary school (JSS). Students are generally 13 to 15 years old in JSS, and must pass a national standardised exam before being able to enrol in *Sekolah Menengah Atas* or senior secondary school (SSS). There are three years of SSS and students generally complete them by age 18 after passing another national standardised exam. Alternatively, students can enrol in three years of vocational schools called *Sekolah Menengah Kejuruan*.

Throughout primary and secondary school, students can choose between attending “regular” or *madrasah* schools—Islamic religious schools under the jurisdiction of the Ministry of Religious Affairs (MoRA). Here we define “regular” schools as any school that is not a madrasah, and we categorise these schools as public and private in Table 3.1 below. Madrasahs largely follow the Ministry of Education and Culture (MoEC) curriculum as they are required to use 70% of that curriculum in their schools. Even private madrasahs, which comprise the vast majority of madrasahs in PS and JSS and are largely run by foundations affiliated with Islamic organisations, may receive some public funding and are supported by the MoRA. This unique dual structure with a special distinction for Islamic schools exists in Indonesia because during colonial times public schools were targeted at educating the Dutch elite, while Islamic schools sought to educate the broader population (Ali et al., 2011).

Table 3.1 shows enrolment and completion rates for those respondents that are within the anticipated age ranges as described above. Indonesia has achieved near universal enrolment in primary school. Of children between 7 and 12 years old, 89.4% are enrolled in and 10.1% already completed primary school. However, enrolment and completion numbers slightly drop for junior secondary (89%) and senior secondary (71%). Part of the reason for the drop-off could be lower demand for education at higher school levels. Students who repeat a grade and/or have a lower household income have an increased risk of dropping out of school (Wicaksono and Witoelar, 2018).

Table 3.1 School enrolment, total, and by type

School type	School level		
	PS (%)	JSS (%)	SSS (%)
Net enrolment	89.4	73.6	54.8
Regular – Public*	83.1	67.0	72.6
Regular – Private*	6.8	10.0	16.2
Madrasah – Public*	1.4	7.7	5.3
Madrasah – Private*	8.8	15.3	5.9
Completed school level	10.1	15.9	15.9

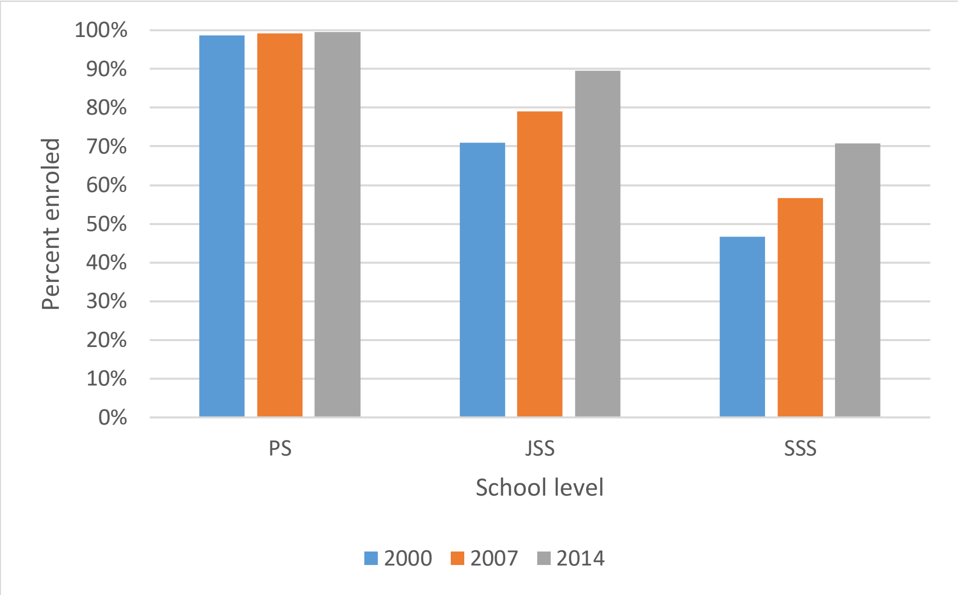
Source: IFLS 5, 2014
Note: Net enrolment and completion rates are calculated as a percentage of respondents who are within the anticipated age range: 7- to 12-year-olds for PS, 13- to 15-year-olds for JSS, and 16- to 18-year-olds for SSS.
*Enrolment by school type shows the share of children enrolled in a particular school type out of children who enrolled in any school.

The type of school students attend does not vary greatly by gender, region, or household economic status (see Appendix 1 Table A.3:). Children living outside the islands of Bali and Java are 9 percentages-points more likely to enrol in regular public school compared to other school choices; while children living on those two islands are more likely to enrol in private schools (including madrasahs). Wealthier children are less likely to enrol in private madrasah, and more likely to enrol in regular schools, as madrasahs are perceived as providing a low-quality education, outside of religious study. We find no statistical difference in school enrolment choices between girls and boys.

Indonesia has low rates of out-of-school children. According to our analysis of IFLS 5 data, less than 1% of 7- to 12-year-olds, which we would expect to be enrolled in PS, are out-of-school and only 7% of 13- to 15-year-olds, which we would expect to be enrolled in JSS, are out-of-school. There is a large drop off for SSS, with more than 35% of 16- to 18-year olds reporting they are not currently enrolled in school and have not completed SSS.[4] While Indonesia has witnessed marked success in high enrolment for PS and JSS, it still has a way to go to improve SSS enrolment.

Looking at changes over time, school enrolment has expanded markedly as the government has improved access and eliminated school fees from public schools. Just using the three most recent rounds of IFLS, we can see in Figure 3.1 that there was little change in PS enrolment since it started at nearly 100% in 2000. However, enrolment in JSS and SSS has increased. JSS enrolment jumped from 71% in 2000 to 90% in 2014. There was an even larger jump for SSS—from 47% in 2000 to 71% in 2014. In a span of 14 years, Indonesia made large improvements in educational access for Grades 7 through 12.

Figure 3.1 Educational enrolment by year and school level



Source: IFLS 3 (2000), 4 (2007) and 5 (2014)
Note: The figure shows the total of net enrolment and completion rates. Net enrolment and completion rates are calculated as a percentage of respondents who are within the anticipated age range and who (1) ever enrolled in the specified school level and are still enrolled, or (2) ever enrolled in the specified school level and finished that school level: 7- to 12-year-olds for PS, 13- to 15-year-olds for JSS, and 16- to 18-year-olds for SSS. This methodology provides a more realistic image of enrolment than net enrolment alone, because 10 to 16 percent of children within these age ranges already finished and are no longer enrolled in the specified school level (Table 3.1).

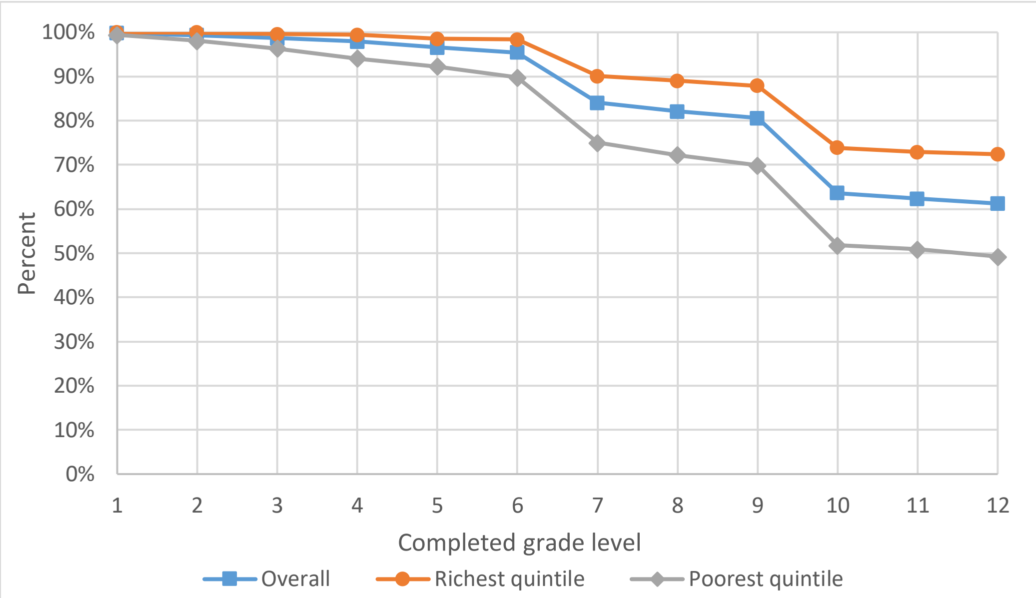
Similar to enrolment, Indonesia has also achieved near universal educational attainment for PS (Figure 3.2). Ninety-five percent of Indonesians between the ages of 20 and 30 have successfully completed PS. However, attainment decreases for JSS (81%) and SSS (61%). Students may decide not to continue to the next level of schooling based on their exam scores, economic pressure to find a job, or disinterest in education. Educational attainment rates are similar between 2000 and 2014 through Grade 9. However, SSS attainment has

[4] We report the percentage of children within the reported age ranges that are out of school, including those that finished a certain school level but then dropped out afterwards. Note that some children between 13 and 15 years old are still enrolled in primary school.

inched up during that 14-year period with 61% completing SSS in 2014 compared to 51% in 2000.

Figure 3.2 shows educational attainment by grade level for the overall population and separately for the richest and poorest quintiles. We see a marked difference when examining educational attainment between the richest quintile and poorest quintile of Indonesians. The difference in attainment between these two groups increases with the level of schooling, extending to a 23 percentage point difference by the end of SSS.[5]

Figure 3.2 Educational attainment by grade level and income



Source: IFLS 5, 2014
Note: Results are fractions of all respondents between 20 and 30 years old who completed each grade level. The quintiles are based on the asset index described in section 2.

We also examine attainment rates by gender and found that they are almost identical for PS and JSS in all IFLS waves. For SSS we observe higher graduation rates among boys, though the gender gap narrows over time. In the latest 2014 wave, boys have a four-percentage point higher SSS graduation rate as compared to girls (Appendix 1 Table A.4).

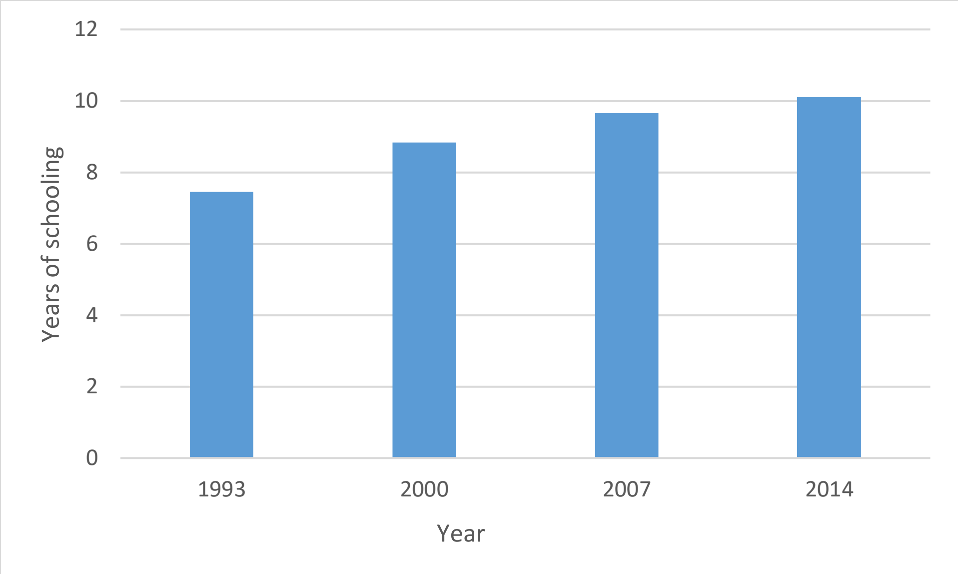
We find that students who started a level of schooling completed it. In 2014, 95% of 20- to 30-year-olds who ever enrolled in PS completed it. That figure is 94% for JSS and 93% for SSS. Although we find that a large fraction of students' drop out of school after completing PS or JSS, graduation rates for those enrolled in JSS and SSS are still over 90%.

We also want to examine how education levels have changed over time. We analyse IFLS data going back to 1993 to calculate the average years of schooling through senior

[5] When examining the 2000 cohort from IFLS 3, we found a similar Grade 12 attainment gap of twenty three percentage points between the poorest and richest quintile of young adults. Both groups had lower Grade 12 attainment rates compared to the 2014 cohort.

secondary.[6] We find that average years of schooling have steadily increased from 7.5 years in 1993 to 10.1 years in 2014 (Figure 3.3).

Figure 3.3 Average years of school attainment increases over time



Source: IFLS 1, 3, 4, and 5
Note: Results based on the years of schooling that correspond to the highest educational attainment reported among 20- to 25-year-olds in each IFLS survey. For those who continued past senior secondary school, educational attainment is capped at twelve years.

Finally, we investigate grade retention. Only a small percentage of students report repeating a grade in school. Approximately 15% of 18-year-olds repeated at least one grade in school (authors' analysis of IFLS 5 data). Students who repeat a grade tend to repeat the first grade of PS, indicating they may not be school or age ready at the time of enrolment.

There has been a change in age for school entry over time. In 2000, 50% of 5-year-olds were enrolled in PS. By 2007, only 4% of 5-year-olds were enrolled, a percentage that stays consistent for the 2014 cohort. School may be more strictly enforcing an age cut-off for first grade based on grade retention rates for primary school. For currently enrolled students in 2014, 7% of first graders and 5% of second graders reported repeating that grade. Grade retention rates for JSS and SSS are less than 1%.

4. Learning Outcomes

Indonesia has succeeded in achieving high levels of school enrolment and attainment, particularly at the PS and JSS levels, and has markedly increased access to JSS and SSS since 2000. Just increasing access to education though, does not necessarily correlate into increased student learning. Schools could have resource challenges, human capital

[6] We focus this analysis on 20 to 25 year olds since they are old enough to have completed SSS and the age range is narrow enough that respondents will not be counted in subsequent IFLS surveys within that same age range (survey is conducted every seven years).

constraints, and management or organisational issues that detract from student learning. In this section we attempt to shed light on what students are actually gaining from their educational experience based on a set of mathematical problems associated with Indonesia's primary school curriculum from the IFLS. Table 4.1 provides a list of each IFLS math problem and the grade level by which students are expected to master that mathematics skill.

Table 4.1 Table grade-level mathematics skills

IFLS test question	Expected grade level to master	Difficulty co-efficient
2-digit subtraction: 49-23	1	-1.98
3-digit addition and subtraction: 267+112-189	2	-.965
One-digit addition and multiplication: (8+9)*3	3	-7.69
Order of operations: (412+213)/(243-118)	3	.021
2-digit division: 56/84	4	.918
Subtracting fractions: 1/3-1/6	4	1.07
Decimals: 0.76-0.4-0.23	4	-.190
Calculating percent (word problem): If 65% of citizens' smoke, and the current citizen population is 160 million, how many people do not smoke?	5	1.04
Calculating interest (word problem): Ali put 75,000 rupiah in his savings account. If he receives 5% interest a year, how much interest does Ali receive on his savings after one year?	5	.110

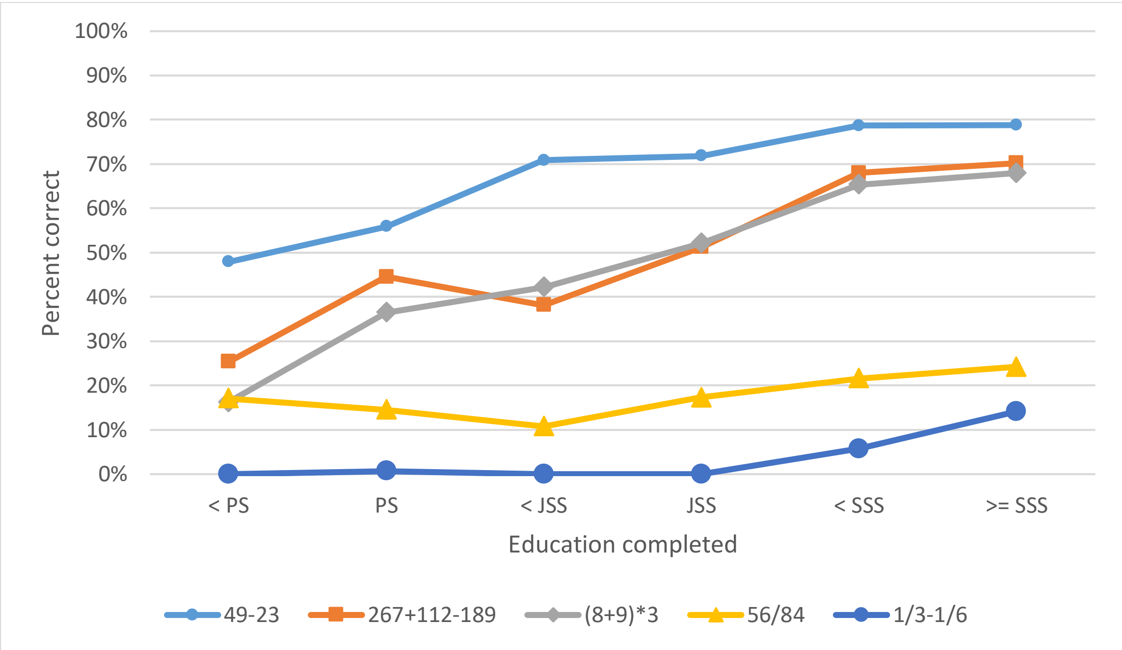
Source: Adapted from the Indonesian Minister of Education Regulation No 24/2016, Attachment 14, difficulty coefficent from Table A.3.4

When comparing the grade levels of the items with the relative difficulty levels found in the data (see Table A.3.4 in Appendix 3), we find that the difficulty levels roughly follow the grade pattern. The first three questions of the relatively easy version of the test have the lowest difficulty coefficients, as expected. However, the data show that Grade 4 fraction items were more difficult for the respondents than the Grade 5 percentage items. In addition, the Grade 4 decimal item is found to be easier than the Grade 3 item that requires the skill to divide numbers. This suggests that the items including fractions were most difficult to the respondents. Interestingly, the overlapping item has different difficulty coefficients in each version. Recall that a third of the respondents that answered both versions of the test gave different answers to these questions. Part of this difference can be explained by guessing.

By analysing IFLS data, we find a strong disconnect between what students are supposed to learn and what they are actually learning as they progress through the education system. Figure 4.1 shows descriptive results for five numeracy problems asked among respondents of the 2014 IFLS. We show results for the fraction who answered each problem correctly grouped by their completed education level. We examine results among 18- to 30-year-olds, a group of respondents who had the opportunity to complete senior secondary but are still young enough to reflect the current quality of the education system.

The easiest question involves subtracting 23 from 49, which is a simple type of subtraction since it does not involve a carryover. In addition, note that these are multiple choice tests, and in this particular problem, the answer choices are 25, 26, and 27. Therefore, the test-taker was only required to subtract three from nine. The education curriculum expects that all first graders can correctly answer this question. After correcting for guessing, we find only around 55% of respondents with only a primary school education can answer this question. This means that 45% of people whose highest level of education is Grade 6 do not have the basic skill of 9-3 and 4-2 stacked side-by-side. Performance on this question only improves by 15 percentage points for individuals with a junior secondary degree and 25 percentage points for individuals with a senior secondary degree, relative to the primary school graduates. Twenty percent of senior secondary graduates spent twelve years learning mathematics without being able to correctly answer an arithmetic question taught in the first grade. Learning levels are lower for more challenging questions. Only 15% of senior secondary graduates can correctly subtract fractions (1/3 minus 1/6), the hardest item in this test but a skill expected to be mastered by Grade 4. No individual with less than some senior secondary education correctly answered this question.

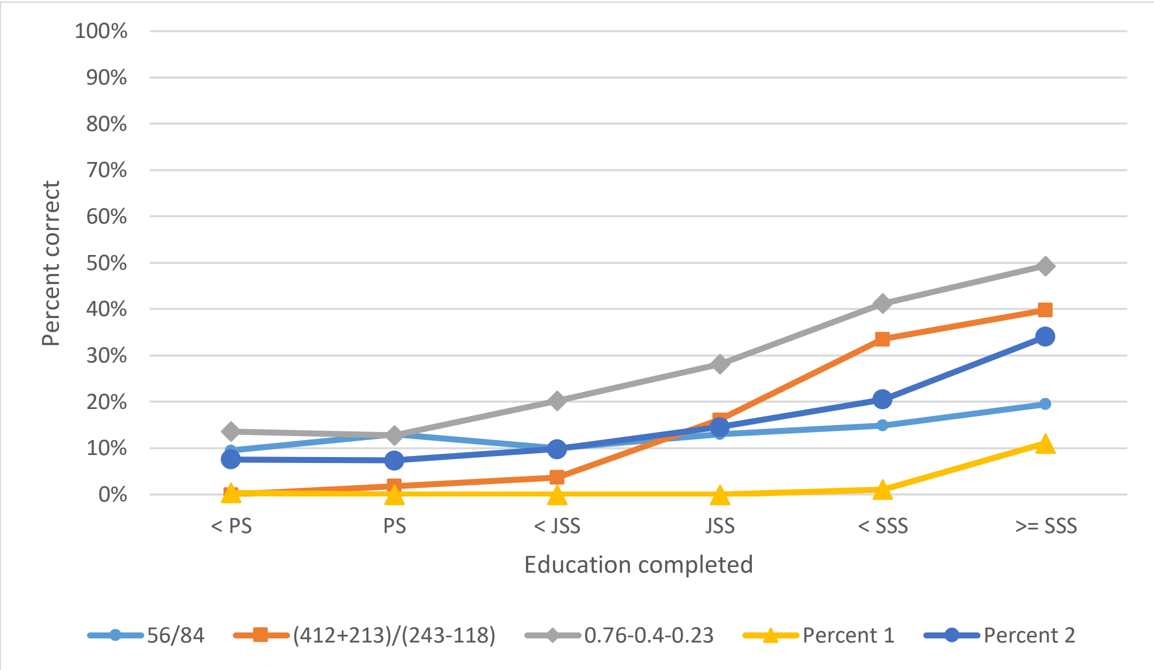
Figure 4.1 Learning by education level completed (easier item-level)



Source: IFLS 5, 2014
Note: Results show the percent who answered each question correctly among 18- to 30-year-olds. Only respondents in this age group who answered the easier item-level questions in IFLS 4 had the opportunity to answer these questions for IFLS 5. Fifty one percent of the weighted sample answered both the easier and the harder item-level test questions shown in Figure 4.2. Results are corrected for guessing as described in section 2. Students who graduated high school, including those who went onto to postsecondary education, are included in the >=SSS group. Within that group, 67% completed high school and did not continue their education, while 15% was reported at least graduated from college.

Respondents who answered the harder set of math questions also performed poorly (Figure 4.2). Interestingly, in comparison to the easier questions, we generally see larger learning gains between completing JSS and SSS, meaning that there is a learning benefit from SSS for relatively more difficult math questions. Yet, the absolute levels of learning are still low and even the majority of senior secondary graduates are unlikely to be able to correctly answer math problems covering skills they were supposed to have learned in PS. For example, even though 0.76 - 0.4 - 0.23 is a subtraction problem that students in fourth grade should have mastered, only around half of individuals graduating from SSS or higher answered this question correctly; and this was apparently the easiest question based on the fraction of people who could answer this question correctly. Furthermore, only 20% of individuals who went to SSS correctly answered the long division problem (e.g., 56 ÷ 84).

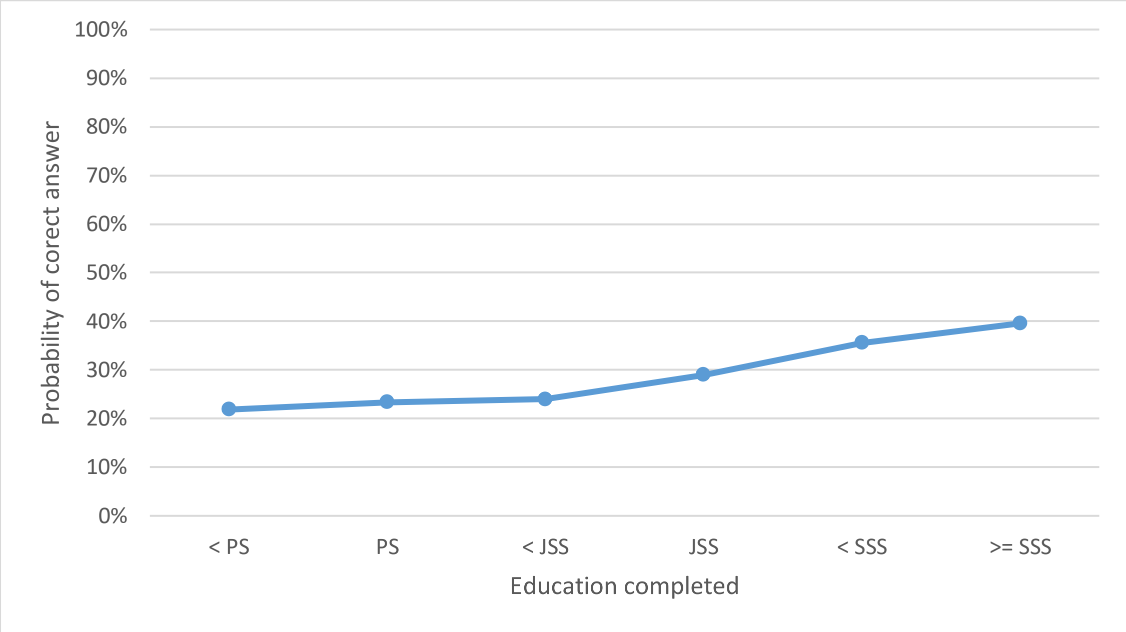
Figure 4.2 Learning by education level completed (harder item-level)



Source: IFLS 5, 2014
Note: Results show the percent who answered each question correctly among 18- to 30-year-olds. Results are corrected for guessing as described in section 2. Students who graduated high school, including those who went onto to postsecondary education, are included in the >=SSS group. Percent 1 = If 65% of citizens smoke, and the current citizen population is 160 million, how many people do not smoke? Percent 2 = Ali put 75,000 rupiah in his savings account. If he receives 5% interest a year, how much interest does Ali receive on his savings after one year?

As described in section 2, we use item response theory (IRT) to develop a numeracy construct that incorporates responses to all item-level math questions and adjusts for question difficulty. Figure 4.3 presents the IRT results for 18- to 30-year-olds, which shows that learning is occurring as students obtain higher levels of education, but the slope is gradual and the absolute levels of learning are low. SSS graduates are predicted to answer math problems right only about 40% of the time—problems they were supposed to have mastered in PS.

Figure 4.3 Learning by education level completed (IRT)

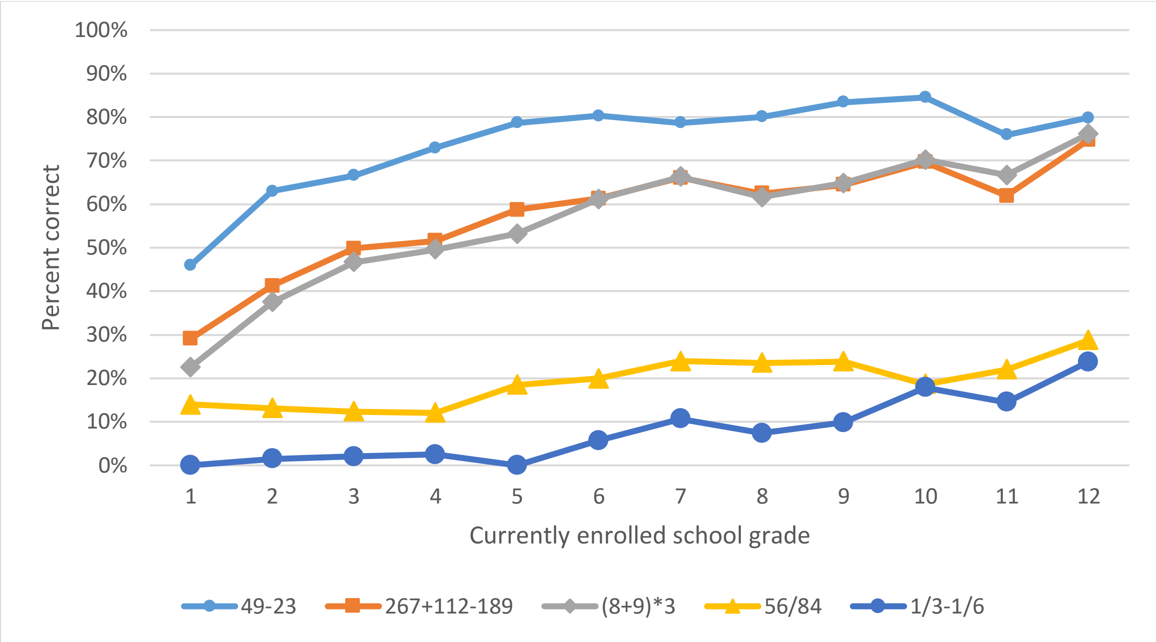


Source: IFLS 5, 2014
Note: Results show the mean probability of answering a math question correctly among 18- to 30-year-olds. Results are adjusted for guessing as described in section 2. Students who graduated high school, including those who went onto to postsecondary education, are included in the >=SSS group.

We also investigate how learning varied across genders, regions, and wealth quintiles, controlling for years of educational attainment (Appendix 1 Table A.5:). For regions, we look at differences between residents of Java-Bali compared to the rest of Indonesia. We find that girls learn slightly more than boys (by 2 percentage points on average) and residents on the islands of Java-Bali learn more than elsewhere in Indonesia by 4.3 percentage points. The point estimates, while statistically significant, are small, illustrating that learning levels are low across gender and region sub-groups. Interestingly, the wealthiest quintile does not significantly outperform the poorest quintile. Given prior studies documenting how learning gaps are associated with household income (Al-Samarrai and Cerdan-Infantes, 2013), it is surprising that the difference in learning in Indonesia between wealthy and poor groups with the same educational attainment are not statistically significant.

Indonesia has seen massive investments in the education system since many 18- to 30-year-olds were in school—the Indonesian Constitution was amended in 2002 to require the government to allocate 20% of public spending to education—so we also want to examine learning levels of current students for the 2014 cohort. Figures Figure 4.4 and Figure 4.5 show how learning is progressing among currently enrolled students for item-level math test questions.

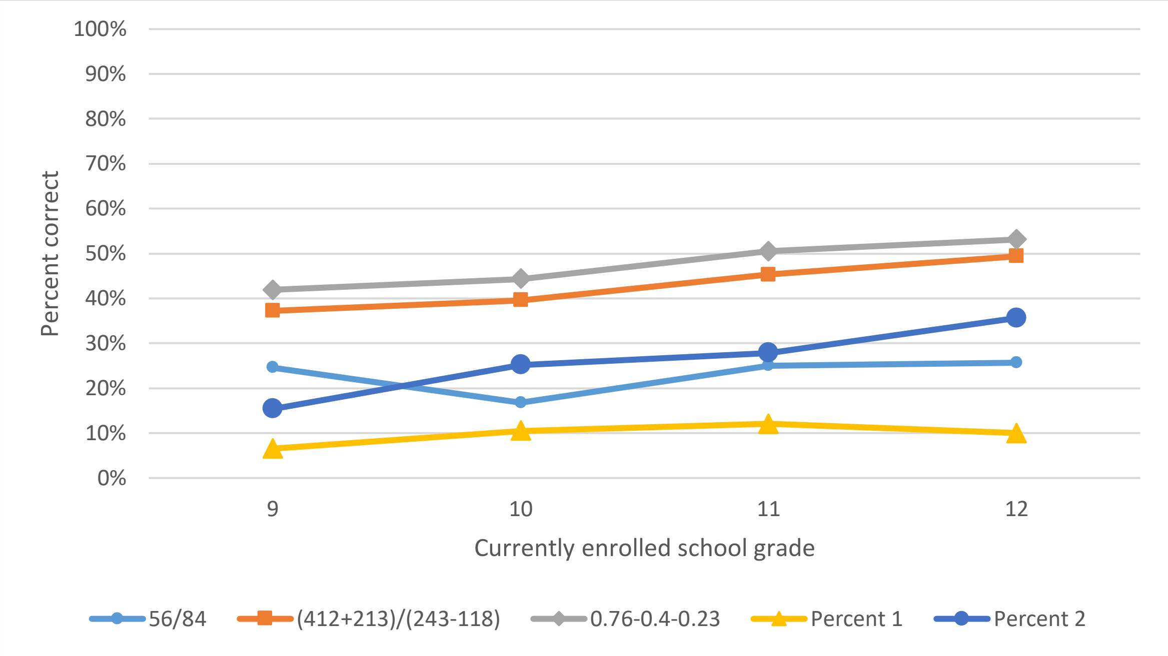
Figure 4.4 Learning by grade level (easier item-level)



Source: IFLS 5, 2014
Note: Results show the percent who answered each question correct among currently enrolled students for the current grade level in 2014. Results are corrected for guessing as described in section 2.

Since the math questions cover the educational curriculum from Grades 1 through 5 (Table 4.1), children in Grade 1, for example, would only be expected to answer the easiest question correctly, the 2-digit subtraction problem (49-23). Conversely, we would expect that all students from Grade 6 onwards would be able to answer all of these math questions correctly. What we find is that the slopes for each question are relatively flat—showing little learning as students' progress through the education system and that mathematical ability is not related to skills students are expected to master in each grade. There are two aspects to this observation. One is that a high fraction (half or more) of individuals who completed third grade answer the three simplest items correctly. This means that even in early grades students are doing two-digit subtraction, three-digit addition and subtraction, and some basic multiplication. The other is that there is very little improvement between first and 12th grade. Nearly 46% of first graders can answer 49-23. This figure increases to only 80% of 12th graders.

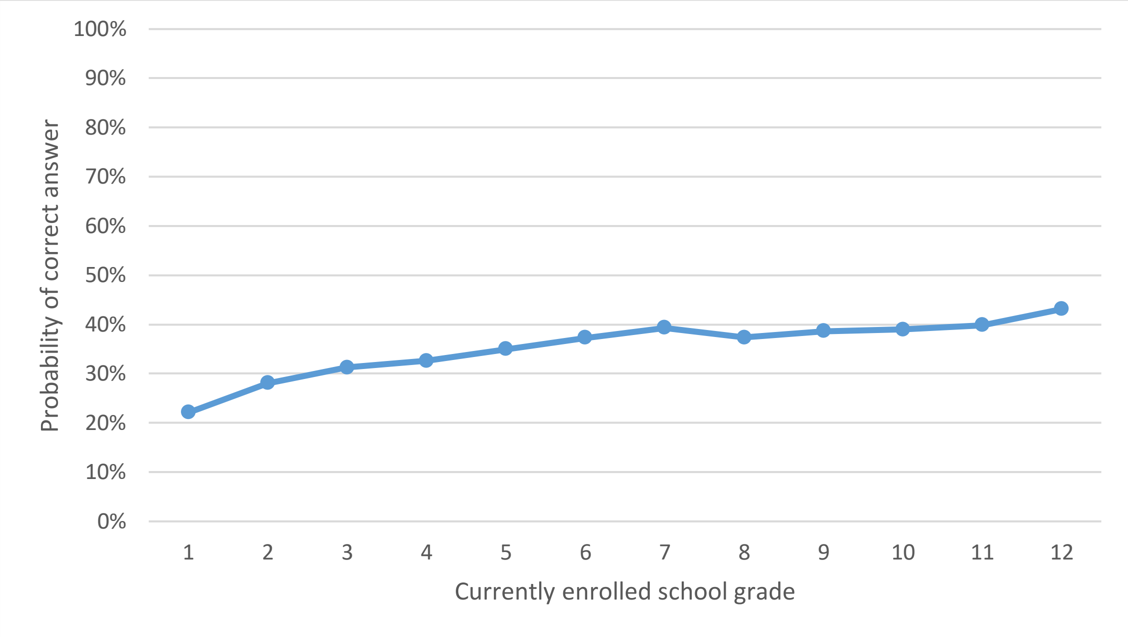
Figure 4.5 Learning by grade level (harder item-level)



Source: Analysis of IFLS 5 data from 2014
Note: Results show the percent who answered each question correct among currently enrolled students. Results are presented beginning with students who enrolled in ninth grade as harder item-level questions were only asked among an older age group (15 years and older). Results are corrected for guessing as described in section 2. Percent 1 = If 65% of citizens smoke, and the current citizen population is 160 million, how many people do not smoke? Percent 2 = Ali put 75,000 rupiah in his savings account. If he receives 5% interest a year, how much interest does Ali receive on his savings after one year?

For the more difficult questions, as with the easier set of questions, there is little learning progression as students' progress through higher grades (Figure 4.5). There are only small changes in learning between ninth grade and 12th grade. Absolute levels of learning are also low given the difficulty of the questions. Only 10% of 12th graders can correctly calculate a percent (percent 1 question). Our single measure of learning using IRT shows similar results (Figure 4.6). First graders have a 22% probability of answering a math question correctly. That number rises to only 43% for 12th graders. Currently, a student's entire primary and secondary educational experience only increases their average numeracy by 21 percentage points.

Figure 4.6 Learning by grade level (IRT)

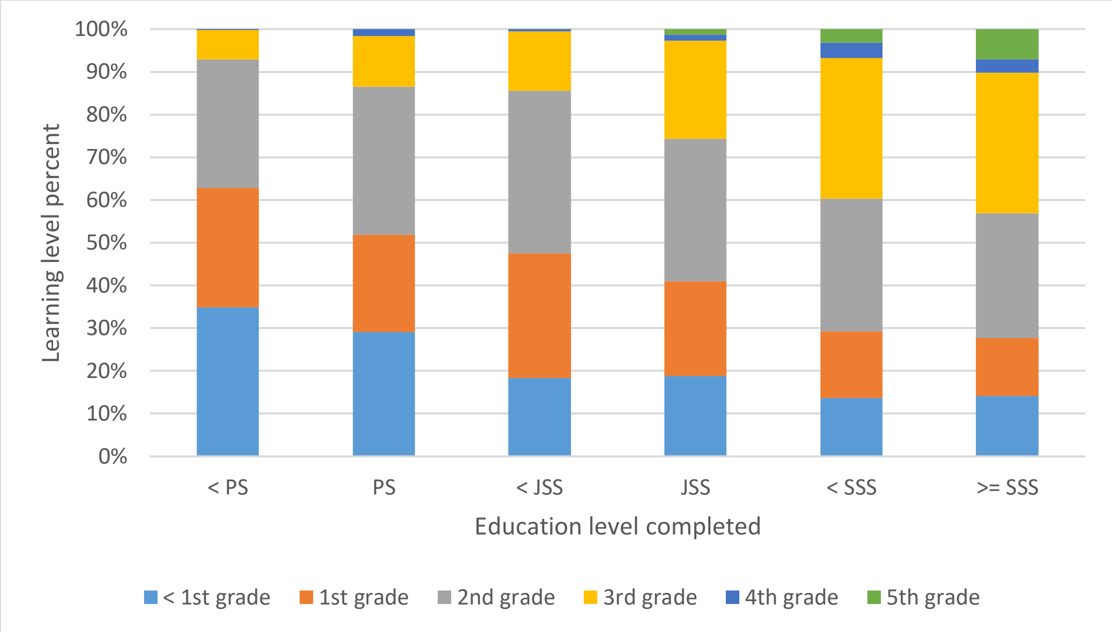


Source: IFLS 5, 2014
Note: Results show the mean probability of answering a math question correctly among currently enrolled students by current grade level. Results are adjusted for guessing as described in section 2.

Subgroup results for currently enrolled students (Appendix 1 Table A.6:) are comparable to what we found for 18- to 30-year-olds. Girls outperform boys by 4 percentage points on average, students outside of Java-Bali perform 2.6 percentage points worse, and the wealthiest quintile of students outperforms the poorest quintile of students by 2.6 percentage points. This gap in learning among wealth quintiles is a larger margin than what we found for older respondents, suggesting that the learning gap associated with household income may be growing over time.

The findings above show that a relatively high proportion of students enrolled in early grades were able to correctly answer higher grade math questions. For instance, even after adjusting for guessing, 23% of currently enrolled first graders correctly answered a third grade problem involving order of operations and subtraction: (8+9)*3 (Figure 4.4). We therefore wanted to investigate cumulative learning to identify the grade level of respondent skills. We return to our sample of 18- to 30-year-olds but restrict it to respondents who answered both the easier and harder item-level questions, which is 51% of our original sample. We define grade level learning for a respondent's ability to answer all questions correctly for skills taught in that grade and all preceding grades. For example, a third-grade learning level means that a respondent answered the first four questions correctly in Table 4.1 covering Grades 1 through 3. If a respondent answered the Grade 1 question incorrectly, but other questions correctly, we consider this respondent to have a learning level of less than first grade, since correctly answering the questions related to higher grades could have been due to guessing (and we cannot correct for guessing among individual respondents). Figure 4.7 reports our results by education level completed.

Figure 4.7 Cumulative learning levels by education level completed



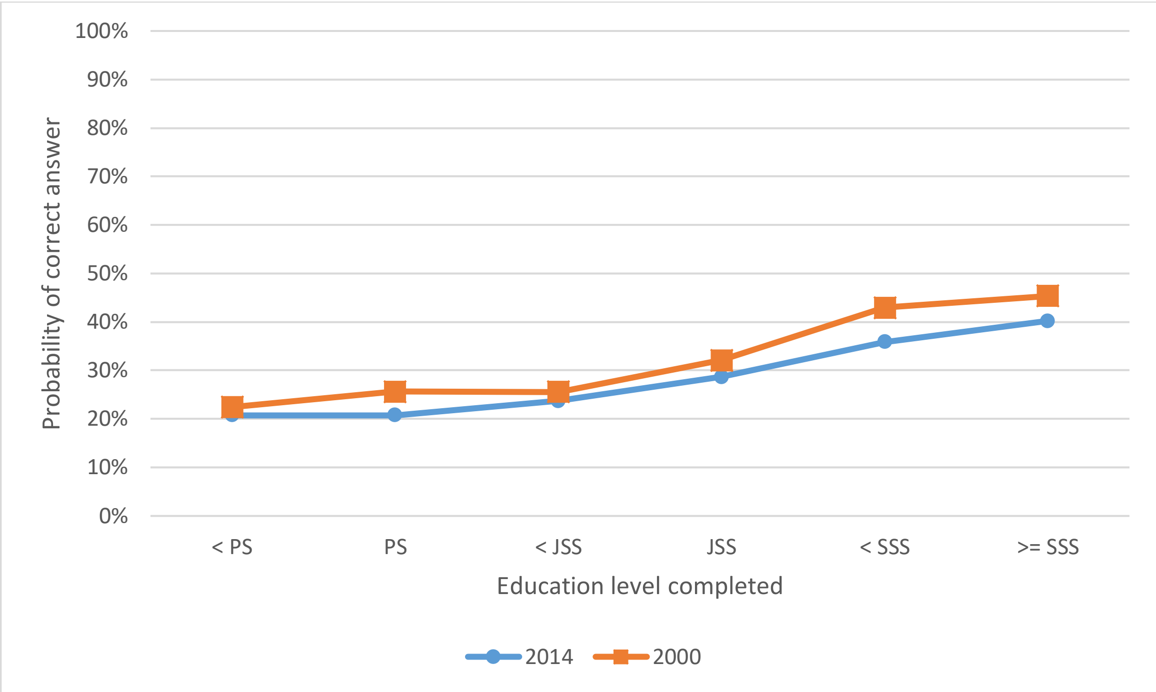
Source: IFLS 5, 2014
Note: Results show the percent of respondents who have achieved grade level learning by education level completed among 18- to 30-year-olds who answered both the easier and harder item-level questions. Students who graduated high school, including those who went onto to postsecondary education, are included in the >=SSS group. We define grade level learning for a respondent's ability to answer all questions correctly for skills taught in that grade and all preceding grades.

While a larger share of respondents who have completed higher education levels exhibit higher learning levels (learning level Grades 3 through 5), the magnitudes of those differences are small. Absolute learning levels are low. Only 7.1% of high school graduates have a fifth-grade mathematics ability, meaning they answered all test questions correctly for skills that should have been taught by fifth grade. While at least half of respondents who graduated from primary school or higher can complete the easier math questions (learning level Grades 1 and 2), only a very small share of respondents were able to grasp Grade 4 and 5 math skills. Overall, the average high school graduate can only complete math problems at the Grade 2 curriculum level (average math grade level is 2.2). That shows only a slight improvement among the average student who only completed primary school, whose math grade level is 1.4.

Recall that the item difficulty coefficients follow the grade pattern up to Grade 3, but the fourth-grade fraction items had higher difficulty coefficients than the fifth-grade percentage items (see Appendix 3: Psychometrics of Numeracy Score). Because the figure assumes that a respondent can only have fifth-grade numeracy skills if he or she also answered the fourth-grade items correctly, we check the percentage of respondents that answered the third-grade items and the fifth-grade items correctly. Leaving out the more difficult fraction items does not change the interpretation of the figure, as only 10.4% of respondents that completed at least SSS correctly answered the third- and fifth-grade items.

Changes in learning over time. Because IFLS asks the same questions to the same set of households over time, it allows us to observe changes in learning across years. Figure 4.8 presents learning results for 18- to 24-year-olds for the 2014 and 2000 samples. (IFLS 3 did not ask numeracy questions to respondents older than 24, so we adjust the age range in the IFLS 5 sample to be the same as in IFLS 3). Learning for the 2014 sample is slightly lower than for the 2000 sample, particularly for those with a senior secondary education. This difference is statistically significant[7]. The slope of the 2000 sample line is similar to the 2014 sample line; relatively flat with limited learning.

Figure 4.8 Learning over time by education level completed (IRT)



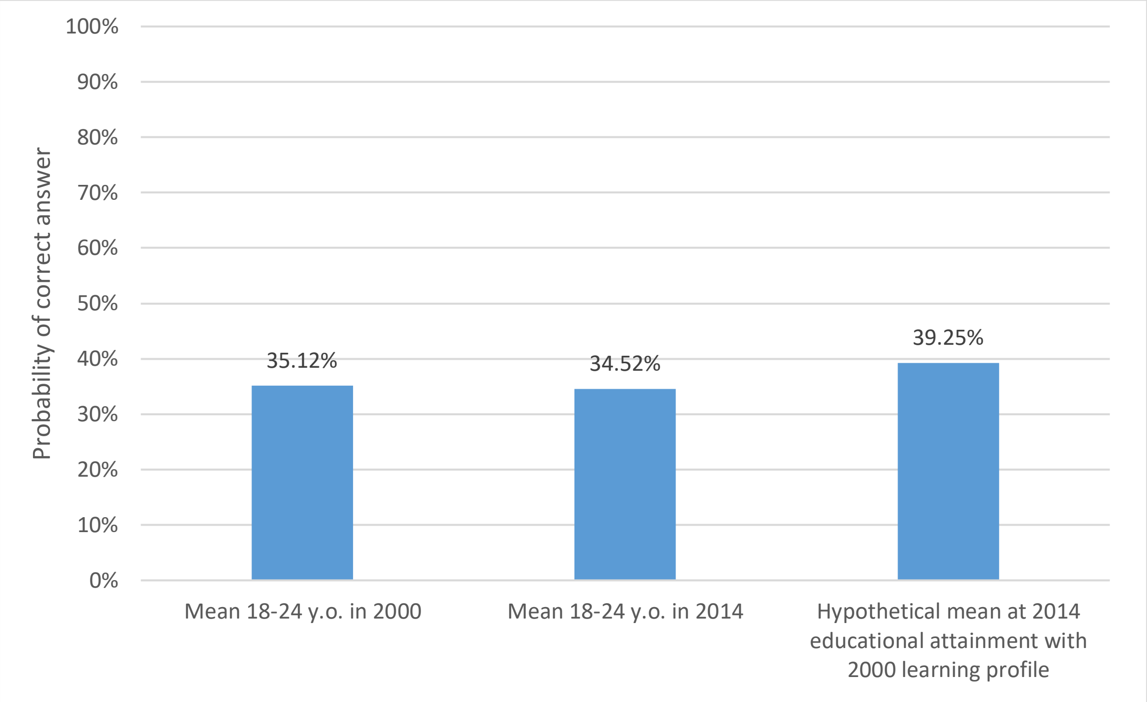
Source: IFLS 3 and IFLS 5
Note: Results show the mean probability of answering a math question correctly among 18- to 24-year-olds. IFLS 3 did not ask numeracy questions to older respondents. Results are adjusted for guessing as described in section 2.

As discussed in Figure 3.1, SSS enrolment increased during this time by 24 percentage points, rising from 47% in 2000 to 71% in 2014, and we would expect the changing enrolment pattern to impact overall learning levels in two ways. On the one hand, the learning profiles show that higher education levels are associated with more learning. Therefore, rising enrolment in secondary education should increase overall learning. On the other hand, a large increase in SSS enrolment changes the composition of the SSS student body. If we assume that the decision to enrol in SSS is correlated with ability (i.e., lower performing students on average drop-out of school), then encouraging increased enrolment is likely to lower the average ability of a student in SSS. We would expect then that the new student body should have lower learning on average, which is exactly what we find in Figure 4.8. Comparing the average IRT score between 2000 and 2014 would give a biased view of changes in learning over time because of the substantial rise in secondary education enrolment. Our dataset allows us to generate the hypothetical mean of learning (defined here as the IRT score) at 2014 enrolment rates using the learning profile of the 2000 sample. Figure 4.9 compares the average IRT score in 2000 and 2014 with this hypothetical measure of learning. If enrolment and other developments had not affected the learning profile, the increased enrolment would have resulted in an increase in the mean numeracy score of 4

[7] As a robustness check for this finding, we checked whether this results is driven by differential item functioning between the years. Table A.3.4 in Appendix 3 shows that this is not the case.

percentage points, with the average probability of correctly answering a math question increasing from 35% to 39% among 18- to 24-year-olds. This is equal to 0.18 standard deviation of the 2000 mean numeracy score. However, because increased secondary enrolment lowered the learning profile, the actual mean IRT score in 2014 was 4 percentage points below the hypothetical one. The findings show that only increasing enrolment will not result in substantial increases in average learning. Moreover, they show that the large investments in the education system between 2000 and 2014 have been ineffective at improving average numeracy.

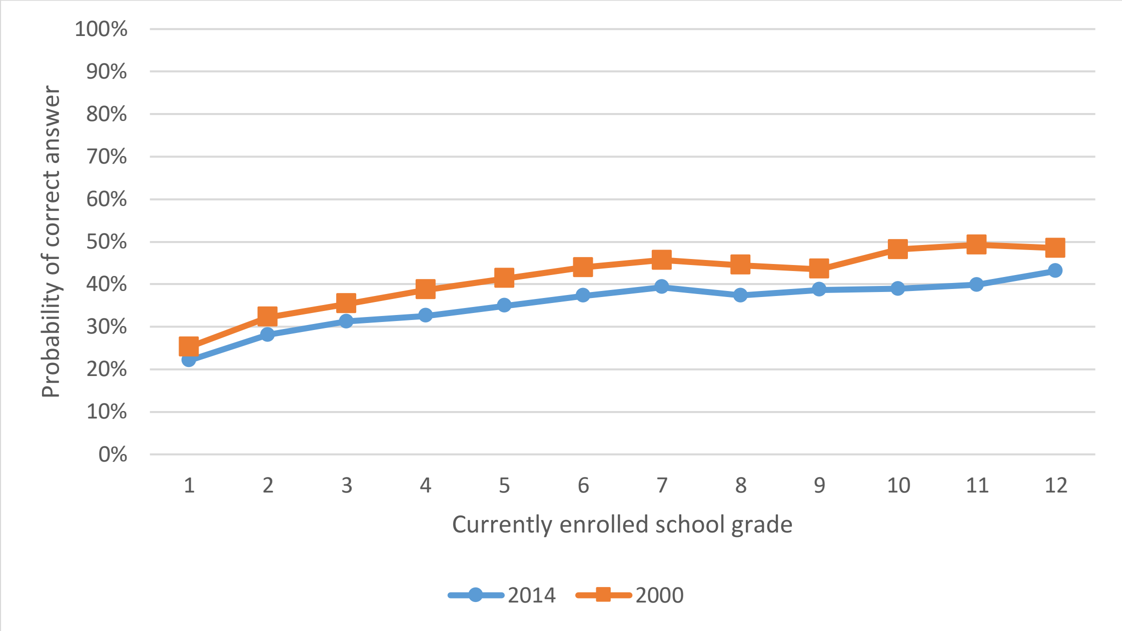
Figure 4.9 Decomposition of learning over time



Source: IFLS 3 and IFLS 5
Note: Results show the mean probability of answering a math question correctly among our actual sample of 18- to 24-year-olds in 2000 and 2014 as well as a hypothetical sample of 18- to 24-year-olds in 2014 who have the numeracy skills according to the 2000 learning profile. Results are adjusted for guessing as described in section 2.

We also find that there are statistically significant decreases in learning over time for each sample of currently enrolled students, with the year 2000 sample performing the best and the 2014 sample performing the worst (Figure 4.10). Same as Figure 4.8 above, part of the decrease in learning in JSS and SSS could be due to greater access to education as students who would not previously be enrolled in school are now in school (as described in Figure 3.1). We find that the average enrolled student in junior secondary school came from a household 0.23 standard deviation above the mean asset index in 2000, while the enrolled students in 2014 came from households only 0.004 standard deviation above the mean (p=0.000). These numbers are 0.25 and 0.05 standard deviation respectively for students enrolled in senior secondary school (p=0.000). Enrolment rate of PS has been near universal since 2000, and thus does not explain the slightly lower learning between the cohorts of students in Grades 1 through 6. Similar to Figure 4.8, the slope of the line for each sample is remarkably similar, presenting a flat learning curve for 12 years of schooling.

Figure 4.10 Learning over time by grade level completed (IRT)

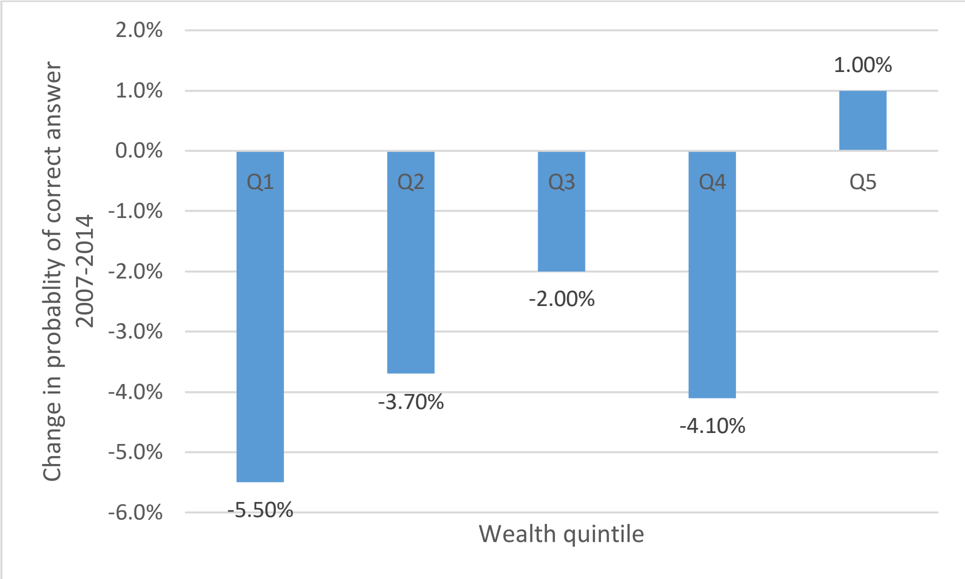


Source: Analysis of IFLS 3 and IFLS 5 data
Note: Results show the mean probability of answering a math question correctly among currently enrolled students by current grade level. Results are adjusted for guessing as described in section 2.

Finally, we take advantage of the panel data to examine changes in learning among the same respondents in the 2007 survey and the 2014 survey. We examine two groups of respondents: 7- to 9-year-olds from the 2007 survey (who are 14 to 16 years olds in the 2014 survey); and 10- to 12-year-olds in the 2007 survey (who are 17 to 19 years old in the 2014 survey). We then divide each group into quintiles based on their asset index in 2007 as defined in Section 2. The poorest households are in quintile 1 and the richest households are in quintile 5. We examine the mean IRT scores within each group for 2007 and 2014 (Figure A.4.1). We find that, in general, learning gains are flat across income groups. For the 7- to 9-year-olds, the wealthiest quintile experienced the largest gains, jumping 10.5 percentage points, and the poorest quintile fares the worst, jumping 3.7 percentage points between 2007 and 2014.

Among the 10- to 12-year-olds, there is actually a small *decrease* in performance across all but the richest quintile between 2007 and 2014 (Figure 4.11). This decrease is the largest among the poorest quintile (5.5 percentage points). Note that this analysis does not control for enrolment, part of the 10 to 12-year-olds in 2007 were out of school in 2014. When we only look at children in school in 2014, there is still a drop in the numeracy score of 0.5 percentage points for the poorest wealth quintile, while the score increased for the richer quintiles. This means that the numeracy skills of the respondents worsen quickly after dropping out of school. While there may be small differences in learning between the wealthiest and poorest children, the larger point is that learning levels remain low over time even as children progress through the education system and that this affects all children across the household wealth distribution.

Figure 4.11 Learning over time by wealth quintile as children aged from 10-12 years old to 17-19 years old between 2007 and 2014



Source: Figure A.4.1

Robustness checks. We conduct several tests to assess the robustness of our findings to different imputation specifications. Our primary results are presented using imputations of random guessing for (partial) missing cases. In Appendix 2 Figure A.1, we present the IRT results for learning in 2014 by school level using various imputation approaches in addition to showing our primary approach as in Figure 4.3. We show IRT results without imputation for item-level questions (no imputation). This method still generates an IRT value for each respondent but only uses non-missing and non-imputed data. We also show IRT results using only respondents without missing data for item-level questions (no imputation—complete tests). Finally, we show IRT results if we impute 1 for all item-level missing responses (upper bound imputation) and if we impute 0 for all item-level missing responses (lower bound imputation). We also examine using different imputation methods for our learning by grade level results (Figure 4.6). We show these imputation findings in Appendix 2 Figure A.2. Overall, we find that results from our primary approach are similar to results without conducting any imputation.

In addition, we compare our results to other recent survey data. The IFLS was conducted at the household level, in a low-stakes environment where the test results did not have any impact on a respondent's academic career or job market prospects. One concern is that such a test structure may result in low effort by respondents. To confirm this, we compare our findings to similar math questions asked as part of the 2011 BERMUTU (Better Education through Reformed Management and Universal Teacher Upgrading Project) survey. BERMUTU was a project implemented by the Ministry of Education and Culture, funded in-part through a World Bank loan. It included a study on teacher certification that covered a near-representative sample of 360 schools across twenty districts of Indonesia (de Ree et al., 2017). These tests were also administered in a low-stakes environment, but in school instead of at the household-level, which might have made respondents take the test more seriously. Overall, learning outcomes from the BERMUTU survey are comparable to the IFLS results, with the IFLS results showing higher learning levels (Appendix 1 Table A.8). For example, 41% of second graders from IFLS 5 correctly answered a 3-digit addition and subtraction problem, compared to 30% in the BERMUTU sample for 3-digit addition and 19% for 3-digit subtraction.

Our findings of low learning levels are also in-line with recent international assessments. Results from the 2015 PISA scores rank Indonesia 64th out of 72 countries. Less than one-third of 15-year-olds achieve a baseline proficiency in mathematics (OECD, 2016). This compares to an average of 77% of students attaining this level of proficiency across OECD countries. To achieve the low international benchmark on the 2015 TIMMS test, fourth-grade students must demonstrate basic mathematical knowledge, including adding and subtracting whole numbers, one-digit multiplication, solving simple word problems and simple fractions, and completion of simple bar graphs and tables. Approximately 27% of fourth graders in Indonesia could not meet this low benchmark. Another 50% met this low benchmark, but nothing more advanced on the assessment (Mullis et al., 2016).

5. Characteristics of Individuals Who Demonstrate Very Low Numeracy Levels

In this section, we examine the characteristics of individuals who have very low numeracy. We define an individual to have low numeracy if the probability that s/he could correctly answer the Grade 2 mathematics problem in IFLS is less than 50% as predicted by IRT. In terms of data, we combine IFLS 5 and IFLS East to get the latest data covering most of Indonesia. As mentioned in section 2, the former was collected in 2014 and covers western and central Indonesia (where most Indonesians live), while the latter was collected in 2012, focusing on eastern Indonesia provinces. We include IFLS East in this analysis because residents of those provinces are disproportionally disadvantaged in terms of socioeconomic status, school access, and familiarity with Indonesian, the main—in most cases only—language of instruction. We focus on school-age respondents, ranging from 7 to 18 years old.

The data consist of 10,992 respondents. We find that 28.4% have low numeracy as defined in the previous paragraph. Examining the proportion of individuals with low numeracy by major island groups (Java-Bali, Sumatra, Eastern Indonesia, and Kalimantan-Sulawesi), we find that the low numeracy rate in Kalimantan-Sulawesi is 25.5%, which is not statistically different from the low numeracy rate in Java-Bali. Meanwhile, the proportion of individuals with low numeracy is significantly higher in Sumatra at 28.4% and Eastern Indonesia at 37.6% compared to the other major island groups. Since primary level education is close to universal in Indonesia, low numeracy in Sumatra and Eastern Indonesia could be due to background conditions or lower quality education, rather than lack of access.

To examine whether there are certain characteristics associated with low numeracy, we estimate a limited dependent variable model (*probit*) of the probability of having low numeracy on individual, household, and regional-level characteristics. These are the same characteristics we use in the previous section. We could not include more school input variables, as the IFLS does not have much information on schools. The results are shown in Table 5.1. The first set of estimates (columns 1 and 2) uses the whole sample, while the second set of estimates limits the sample to those with at least three years of education, respondents who are expected to have already mastered the Grade 2 mathematics curriculum. We observe that even in the sample that is expected to master the Grade 2 curriculum, 24.4% still fail the Grade 2 problem.

We find that individuals who are older and male are more likely to have lower numeracy scores, although the magnitude of the difference between males and females is small.

Controlling for urban/rural residence, those living in Eastern Indonesia have a significantly higher probability of having low numeracy by about 9.4 to 10.8 percentage points compared to Kalimantan-Sulawesi, while residents of Sumatra have a higher probability to have low numeracy by 4.0 to 4.2 percentage points. We find no correlation between household wealth and the probability of having low numeracy in column 1; however, when we limit the sample to only those with more than two years of schooling, individuals living in the richest 20% of households have a lower probability of having low numeracy by 3.2 percentage points. This relatively muted correlation between wealth and low numeracy is similar to the finding in section 4 on the small differences in overall numeracy between wealth quintiles.

Table 5.1 Correlates of low numeracy

	All sample		**Sample >2 years of schooling**	
	Estimates	Average marginal effects	Estimates	Average marginal effects
	(1)	(2)	(3)	(4)
Age (years)	0.019*	0.006	0.055***	0.017
	(0.011)		(0.014)	
Male (Yes = 1)	0.142***	0.047	0.140***	0.043
	(0.026)		(0.031)	
Live in an urban area (Yes = 1)	-0.078***	-0.026	-0.100***	-0.031
	(0.029)		(0.035)	
Completed years of schooling (Reference: no schooling (Col 1); 3 years of schooling (Col 3))				
1	-0.255***	-0.079		
	(0.062)			
2	-0.460***	-0.135		
	(0.063)			
3	-0.521***	-0.149		
	(0.067)			
4	-0.563***	-0.159	-0.077	-0.023
	(0.072)		(0.059)	
5	-0.850***	-0.218	-0.402***	-0.110
	(0.078)		(0.065)	
6	-0.849***	-0.214	-0.444***	-0.119
	(0.088)		(0.076)	
7	-0.785***	-0.203	-0.401***	-0.109
	(0.093)		(0.081)	
8	-0.897***	-0.221	-0.542***	-0.140
	(0.101)		(0.092)	
9	-0.804***	-0.205	-0.495***	-0.129
	(0.111)		(0.104)	
10	-0.929***	-0.220	-0.638***	-0.155
	(0.123)		(0.119)	
11	-1.115***	-0.241	-0.853***	-0.187

	All sample		Sample >2 years of schooling	
	(0.137)		(0.135)	
12	-1.092***	-0.236	-0.848***	-0.183
	(0.150)		(0.149)	
Mother's completed years of schooling	-0.025***	-0.008	-0.021***	-0.007
	(0.004)		(0.005)	
Father's completed years of schooling	-0.009**	-0.003	-0.009*	-0.003
	(0.004)		(0.005)	
Household wealth quintile (Reference: poorest quintile)				
2	-0.007	-0.002	-0.005	-0.002
	(0.042)		(0.049)	
3	-0.002	-0.001	-0.040	-0.012
	(0.040)		(0.048)	
4	-0.029	-0.009	-0.062	-0.019
	(0.043)		(0.051)	
5 (richest)	-0.040	-0.013	-0.107**	-0.032
	(0.043)		(0.052)	
Region of residence (Reference: Kalimantan/Sulawesi)				
Java-Bali	0.035	0.012	0.058	0.018
	(0.042)		(0.050)	
Sumatra	0.118**	0.040	0.133**	0.042
	(0.047)		(0.057)	
Eastern Indonesia	0.269***	0.094	0.329***	0.108
	(0.046)		(0.056)	
Constant	-0.002		-0.894***	
	(0.115)		(0.165)	
Observed probability	0.284		0.244	
N	10,992		8,005	
Log-likelihood	6237.919		4286.646	
Note: *** 1%; ** 5%; * 10% significance; estimates are from a *probit* estimation, where the dependent variable is low numeracy skills = 1; robust standard errors in parentheses.				

Source: Analysis of IFLS 5 and IFLS East data

On education levels, we find that higher education attainment reduces the probability of having low numeracy. An individual with one year of schooling has an 8 percentage point lower probability of having low numeracy compared to an individual with no schooling. All the effects are statistically significant, with increasing point estimates for each year of completed education. However, it appears that the highest benefit lies in the first five years of education. Individuals completing five or twelve years of education do not have much of a difference in the probability of having low numeracy. The results also indicate that if an individual could not master Grade 2 mathematics curriculum by Grade 5, s/he is unlikely to be able to master the skill despite attaining higher levels of education. Indeed, we find that 22% of individuals with five or more years of schooling have low numeracy.

Finally, we see a small but statistically significant and negative correlation between mother’s education attainment and her child’s probability of having low numeracy. This indicates that mother's level of education benefits children's numeracy, albeit only very slightly after other socioeconomic characteristics are controlled for. We find an even smaller negative correlation between low numeracy and father's education attainment.

The results in this section show that the largest correlate of numeracy is years of schooling. In comparison, the correlation between age, sex, urban residence, parental education, and household quintiles are relatively small. There also appears to be differences between major islands across Indonesia, with Eastern Indonesia and Sumatra having a significantly higher probability of low numeracy. We also find, however, that while more schooling, especially at Grade 5 or higher, reduces the probability for an individual to have low numeracy by a large amount, we find that 22% of individuals with more than five years of education still have low numeracy. This indicates that individuals with such low numeracy are still promoted through the Indonesian education system.

6. Discussion and Conclusion

Indonesia has succeeded in near universal attainment in PS. Although attainment rates are lower for JSS and SSS, the government has significantly increased access to both school levels over the past fourteen years, increasing SSS enrolment by 24 percentage points between 2000 and 2014. Further, over 93% of students who start a school level, including PS, JSS, and SSS, end up completing it.

Despite the positive progress on school enrolment, little was known about learning levels in Indonesia and how gains in educational access and attainment translate into student learning. By analysing three rounds of IFLS survey data, for the first time in Indonesia we develop learning profiles of children and young adults as they progress through the education system and profiles to identify how learning changed over the past fifteen years. We find that higher levels of school enrolment and attainment have translated into zero improvement in basic numeracy. Even graduates of SSS struggle to correctly answer math problems representing foundational skills such as order of operations, fractions, and 1-digit multiplication. Progressing through twelve years of the education system results in only a limited improvement in math knowledge. Students in Grade 1 have a 22% probability of answering a math question correctly. That rises to only 43% for 12th graders so that a student's entire primary and secondary educational experience only increases their average numeracy by 21 percentage points. While there are differences within subgroups, such as girls slightly outperforming boys, the magnitudes of these differences are small and absolute levels of learning across groups remain relatively low.

We also curiously find that learning levels slightly decreased over time. Average learning levels for the 2014 IFLS sample are lower than for both the 2007 and 2000 samples. Those differences are small in magnitude so that absolute learning levels remain low for all samples. The 2014 sample of SSS graduates have a lower probability of correctly answering math problems compared to 2000 sample by 6 percentage points. The lack of learning improvement between samples is surprising given the changes in the education system that occurred between 2000 and 2014. This includes implementing decentralisation in 2001 to allow districts more flexibility with introducing innovative education policies and adjusting policy to reflect local context; the 2002 amendment to the Constitution that required 20% of the budget be devoted to education expenditures—resulting in a threefold increase in real education budget; and the 2005 teacher certification policy as a way to improve teacher quality. The net effects of these education reforms have not seemed to improve numeracy among children in school or recent graduates, even if enrolment in JSS and SSS has increased substantially.

However, finding no improvement or even a slight decline in learning is perhaps less surprising given that many of these policies were not directly targeted at improving the types of numeracy questions analysed in this paper. For example, districts could use greater education policy autonomy to achieve goals that are not necessarily aligned with improving student learning, such as satisfying certain constituent demands for job opportunities within the school system. Negative externalities to higher enrolment, such as larger class sizes, a greater range of student ability within each classroom (making it harder for teachers to teach), or an increase in behavioural disruptions could result in lower average test scores. An impact evaluation of the 2005 teacher certification policy showed that the policy reform had no impact on student learning (de Ree et al., 2017). If the 2002 budget requirements on education expenditures were largely focused on inputs, there is mounting evidence around the world that an increase in inputs are not associated with improvements in learning (Pritchett, 2013).

When looking at learning outcomes by gender, wealth, and region, we find that no particular subgroup of children is driving these low performance results. However, we identified characteristics of children who are struggling to grasp basic mathematical concepts. These children are more likely to live in Eastern Indonesia, in rural areas, and be older and male. Increased years of schooling does improve student learning—completing PS reduces the likelihood that a child will have low numeracy by 21 percentage points. However, there are few marginal gains of schooling beyond that. Completing SSS only reduces the likelihood that a child will have low numeracy by 1 additional percentage point compared to completing primary school. Interestingly, in all of our analyses we find hardly any learning gaps between wealthy and poor students, controlling for years of education.

The IFLS data has two important limitations for the learning profiles. First, we can only assess learning with a numeracy test, and learning profiles for other skills such as reading could be different. Second, the IFLS contains only a small number of test items that form the numeracy score. Adding items would improve the internal consistency of the score, and we would be able to distinguish the ability levels of individuals in more detail. Despite the limitations, the findings are in-line with larger international educational assessments of Indonesia, such as PISA and TIMMS that rank Indonesia as having low levels of student learning. Indonesia is succeeding with increasing educational access but struggling with improving school quality. While provincial and district governments have, and continue to, experiment with education policy changes, they have yet to identify policies that significantly improve student learning. Further research is needed to rigorously evaluate which policies improve student outcomes and if those policies can be replicated on a larger scale.

References

Afkar, R., de Ree, J. and Khairina, N. (forthcoming). *Who Learns What in Basic Education? Evidence from Indonesia*. Jakarta: World Bank.

Ali, M., Kos, J., Lietz, P., Nugroho, D., Furqon, A. and Emilia, E. (2011). *Quality of education in Madrasah: main study*. Washington, D.C.: World Bank.

Al-Samarrai, S. and Cerdan-Infantes, P. (2013) "Where Did All the Money Go? Financing Basic Education in Indonesia," in Suryadarma, D. and Jones, G. W. (eds) Education in Indonesia. ISEAS–Yusof Ishak Institute, pp. 109–138.

Banerji, R., Bhattacharjea, S. and Wadhwa, W. (2013). The Annual Status of Education Report (ASER). *Research in Comparative and International Education*, 8(3), pp.387-396.

Crouch, L. and Gustafsson, M. (2018). *Worldwide Inequality and Poverty in Cognitive Results: Cross-sectional Evidence and Time-based Trends*. RISE-WP-1 8 /0 19. [online] RISE Programme. Available at: https://www.riseprogramme.org/sites/www.riseprogramme.org/files/publications/RISE_WP-019_Crouch-Gustafsson.pdf [Accessed 8 Jun. 2018].

de Ree, J., Muralidharan, K., Pradhan, M. and Rogers, F. (2017). Double for Nothing? Experimental Evidence on the Impact of an Unconditional Teacher Salary Increase on Student Performance in Indonesia. *The Quarterly Journal of Economics*, Volume 133(2), pp.993-1039.

Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological methods*, *23*(1), 138.

Filmer, D. and Pritchett L. (2001). Estimating Wealth Effects without Expenditure Data or Tears: An Application to Educational Enrollments in the States of India. *Demography,* 38(1), pp. 115-132.

Frankenberg, E., Karoly, L., Gertler, P., Achmad, S., Agung, I., Hatmadji, S. and Sudharto, P. (1995). *The 1993 Indonesian Family Life Survey: Overview and Field Report*. Santa Monica, CA: RAND Corporation.

Hanushek, E.A., Lavy, V. and Hitomi, K. (2008) Do Students Care about School Quality? Determinants of Dropout Behavior in Developing Countries. *Journal of Human Capital*, 2(1), pp. 69-105.

Jones, S., Schipper, Y., Ruto, S. and Rajani, R. (2014). Can Your Child Read and Count? Measuring Learning Outcomes in East Africa. *Journal of African Economies*, 23(5), pp. 643-672.

Kaffenberger, M. and Pritchett, L. (2017). *More School or More Learning? Evidence from Learning Profiles from the Financial Inclusion Insights Data*. RISE -WP-17/012. [online] RISE Programme. Available at: https://www.riseprogramme.org/sites/www.riseprogramme.org/files/publications/RISE_WP-012_Pritchett%20Kaffenberger.pdf [Accessed 8 Jun. 2018].

Mullis, I. V. S., Martin, M. O., Foy, P., & Hooper, M. (2016). *TIMSS 2015 International Results in Mathematics.* Retrieved from Boston College, TIMSS & PIRLS International Study Centre website: http://timssandpirls.bc.edu/timss2015/international-results/

OECD (2016). *PISA 2015 Results (Volume I): Excellence and Equity in Education*. PISA. Paris: OECD Publishing, p.192.

Pritchett, L. (2013). *Rebirth of Education*. Center for Global Development.

Pritchett, L. and Beatty, A. (2015). Slow down, you're going too fast: Matching curricula to student skill levels. *International Journal of Educational Development*, 40, pp.276-288.

Pritchett, L. and Sandefur, J. (2017). *Girls' Schooling and Women's Literacy: Schooling Targets Alone Won't Reach Learning Goals*. CGD Policy Paper 104. [online] Washington D.C.: Center for Global Development. Available at: https://www.cgdev.org/sites/default/files/girls-schooling-womens-literacy-targets-alone-reach-learning-goals.pdf [Accessed 8 Jun. 2018].

Rokx, C., Giles, J., Satriawan, E., Marzoeki, P., Harimurti, P. and Yavuz, E. (2010). *New insights into the provision of health services in Indonesia: a health workforce study*. Washington, D.C.: The World Bank.

Rolleston, C. (2014). Learning profiles and the 'skills gap' in four developing countries: a comparative analysis of schooling and skills development. *Oxford Review of Education*, 40(1), pp.132-150.

Rolleston, C. and James, Z. (2015). After access: Divergent learning profiles in Vietnam and India. *PROSPECTS*, 45(3), pp.285-303.

Sikoki, B., Witoelar, F., Strauss, J., Meijer, E. and Suriastini, W. (2013). *IFLS East User's Guide and Field Report*. Yogyakarta: SurveyMETER.

Singh, A. (2017). *Learning More With Every Year: School Year Productivity and International Learning Divergence*.

Spaull, N. and Kotze, J. (2015). Starting behind and staying behind in South Africa. *International Journal of Educational Development*, 41, pp.13-24.

Stata: Release 15. (2017). College Station, TX: StataCorp LLC.

Strauss, J., Beegle, K., Sikoki, B., Dwiyanto, A., Herawati, Y. and Witoelar, F. (2004). *The Third Wave of the Indonesia Family Life Survey: Overview and Field Report*. WR144/1-NIA/NICHD.

Strauss, J., Witoelar, F. and Sikoki, B. (2016). *The Fifth Wave of the Indonesia Family Life Survey: Overview and Field Report*. WR-1143/1-NIA/NICHD. Santa Monica, CA: RAND Corporation.

Strauss, J., Witoelar, F., Sikoki, B. and Wattie, A. (2009). *The Fourth Wave of the Indonesia Family Life Survey: Overview and Field Report*. WR-675/1-NIA/NICHD.

Wicaksono, T. and Witoelar, F. (2018). Early Educational Experience and Later Education Outcomes: Evidence from School Transition in Indonesia. *Bulletin of Indonesian Economic Studies*, pp.1-66.

Appendix 1: Tables

Table A.1: Fraction correct with and without imputations—currently enrolled in 2004

	NOT IMPUTED					IMPUTED				
Grade /Item	***EK13***	***EK14***	***EK15***	***EK16***	***EK17***	***EK13***	***EK14***	***EK15***	***EK16***	***EK17***
1	0.65	0.42	0.46	0.24	0.00	0.46	0.29	0.23	0.14	0.00
SD	0.64	0.73	0.72	0.66	0.58	0.72	0.75	0.75	0.64	0.56
N	329	283	245	207	206	457	457	457	457	457
2	0.69	0.48	0.48	0.21	0.01	0.63	0.41	0.37	0.13	0.01
SD	0.61	0.71	0.71	0.66	0.58	0.65	0.73	0.74	0.64	0.59
N	904	839	794	611	615	997	997	997	997	997
3	0.70	0.54	0.51	0.14	0.02	0.67	0.50	0.47	0.12	0.02
SD	0.60	0.69	0.71	0.64	0.59	0.62	0.71	0.72	0.63	0.59
N	1012	988	958	759	772	1053	1053	1053	1053	1053
4	0.75	0.53	0.52	0.14	0.03	0.73	0.52	0.50	0.12	0.02
SD	0.56	0.70	0.70	0.64	0.60	0.58	0.70	0.71	0.63	0.59
N	995	981	971	832	860	1013	1013	1013	1013	1013
5	0.80	0.61	0.55	0.20	0.01	0.79	0.59	0.53	0.18	0.00
SD	0.51	0.66	0.69	0.65	0.58	0.52	0.67	0.70	0.65	0.58
N	961	956	949	889	899	973	973	973	973	973
6	0.81	0.62	0.63	0.22	0.07	0.80	0.61	0.61	0.20	0.06
SD	0.50	0.66	0.65	0.66	0.61	0.51	0.66	0.66	0.65	0.61
N	950	945	943	903	902	956	956	956	956	956
7	0.79	0.67	0.68	0.25	0.11	0.79	0.66	0.66	0.24	0.11
SD	0.52	0.62	0.62	0.66	0.63	0.52	0.63	0.63	0.66	0.63
N	854	851	847	832	837	862	862	862	862	862
8	0.80	0.63	0.62	0.24	0.08	0.80	0.62	0.62	0.24	0.07
SD	0.51	0.65	0.65	0.66	0.62	0.51	0.65	0.65	0.66	0.61
N	802	798	799	788	795	805	805	805	805	805
9	0.84	0.65	0.65	0.24	0.10	0.83	0.64	0.65	0.24	0.10
SD	0.46	0.64	0.63	0.66	0.63	0.47	0.64	0.64	0.66	0.62
N	686	683	685	677	679	691	691	691	691	691
10	0.85	0.70	0.71	0.20	0.17	0.85	0.70	0.70	0.19	0.18
SD	0.45	0.60	0.59	0.65	0.65	0.46	0.60	0.60	0.65	0.65
N	467	467	465	454	455	470	470	470	470	470
11	0.76	0.62	0.67	0.22	0.15	0.76	0.62	0.67	0.22	0.15
SD	0.55	0.65	0.63	0.66	0.64	0.55	0.65	0.62	0.66	0.64
N	386	384	385	377	381	386	386	386	386	386
12	0.80	0.75	0.77	0.29	0.25	0.80	0.75	0.76	0.29	0.24
SD	0.51	0.56	0.54	0.67	0.66	0.51	0.56	0.55	0.67	0.66
N	413	413	412	405	407	414	414	414	414	414

Source: IFLS 5
Note: Questions are specified in Table 2.1. Results presented for respondents currently enrolled in specified grade. Means are corrected for guessing.

Table A.2: Fraction correct with and without imputations—above 18 in 2014

	NOT IMPUTED										IMPUTED									
Level/Item	***EK13***	***EK14***	***EK15***	***EK16***	***EK17***	***EK18***	***EK19***	***EK20***	***EK21***	***EK22***	***EK13***	***EK14***	***EK15***	***EK16***	***EK17***	***EK18***	***EK19***	***EK20***	***EK21***	***EK22***
Less than primary	0.52	0.30	0.15	0.18	0.00	0.12	0.00	0.16	0.00	0.10	0.48	0.25	0.16	0.17	0.00	0.10	0.00	0.14	0.00	0.08
SD	0.70	0.75	0.75	0.65	0.55	0.63	0.54	0.64	0.55	0.63	0.72	0.75	0.75	0.65	0.55	0.62	0.57	0.64	0.58	0.62
N	115	112	112	111	111	360	358	354	355	357	141	141	141	141	141	507	507	507	507	507
Primary	0.59	0.48	0.42	0.16	0.00	0.16	0.00	0.14	0.00	0.08	0.56	0.45	0.36	0.15	0.01	0.13	0.02	0.13	0.00	0.07
SD	0.67	0.72	0.73	0.65	0.57	0.64	0.58	0.64	0.54	0.62	0.68	0.73	0.74	0.64	0.58	0.64	0.59	0.63	0.54	0.61
N	322	321	313	300	299	877	882	874	874	880	334	334	334	334	334	1006	1006	1006	1006	1006
Less than junior secondary	0.74	0.41	0.44	0.11	0.00	0.13	0.04	0.22	0.00	0.11	0.71	0.38	0.42	0.11	0.00	0.10	0.04	0.20	0.00	0.10
SD	0.57	0.73	0.73	0.63	0.50	0.64	0.60	0.66	0.50	0.63	0.60	0.74	0.73	0.63	0.49	0.63	0.60	0.65	0.50	0.62
N	163	163	162	161	162	357	358	358	355	357	169	169	169	169	169	395	395	395	395	395
Junior secondary	0.74	0.53	0.55	0.20	0.00	0.14	0.18	0.31	0.00	0.15	0.72	0.51	0.52	0.17	0.00	0.13	0.16	0.28	0.00	0.15
SD	0.57	0.70	0.69	0.65	0.55	0.64	0.65	0.67	0.55	0.64	0.59	0.70	0.70	0.65	0.55	0.64	0.64	0.66	0.55	0.64
N	710	701	699	674	678	1520	1544	1532	1516	1524	725	725	725	725	725	1665	1665	1665	1665	1665
Less than senior secondary	0.80	0.68	0.67	0.22	0.06	0.15	0.35	0.43	0.01	0.22	0.79	0.68	0.65	0.22	0.06	0.15	0.34	0.41	0.01	0.21
SD	0.51	0.61	0.62	0.66	0.61	0.64	0.67	0.66	0.58	0.66	0.52	0.62	0.63	0.66	0.61	0.64	0.67	0.66	0.58	0.65
N	320	321	320	310	313	500	512	506	502	509	324	324	324	324	324	529	529	529	529	529
Senior secondary or more	0.80	0.71	0.69	0.25	0.14	0.20	0.41	0.52	0.11	0.36	0.79	0.70	0.68	0.24	0.14	0.19	0.40	0.49	0.11	0.34
SD	0.51	0.59	0.61	0.66	0.64	0.65	0.66	0.64	0.63	0.67	0.52	0.60	0.61	0.66	0.64	0.65	0.66	0.65	0.63	0.67
N	2884	2877	2868	2817	2834	4989	5027	5002	4984	5003	2913	2913	2913	2913	2913	5221	5221	5221	5221	5221

Source: IFLS 5
Note: Questions are specified in Table 2.1. Results presented for respondents 18 years old and above by highest completed school level. Means are corrected for guessing.

Table A.3: Subgroup results for school type

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	Regular-Public			Regular-Private			Madrasah-Public			Madrasah-Private		
Male	-0.005			0.001			-0.004			0.009		
	(0.010)			(0.006)			(0.004)			(0.007)		
Other than Java/Bali		0.088***			-0.028***			0.004			-0.063***	
		(0.022)			(0.011)			(0.008)			(0.016)	
Wealth quintile 2			0.033*			-0.004			-0.002			-0.026**
			(0.019)			(0.011)			(0.008)			(0.013)
Wealth quintile 3			0.031			0.009			-0.007			-0.033**
			(0.019)			(0.011)			(0.008)			(0.013)
Wealth quintile 4			-0.002			0.026**			0.002			-0.026*
			(0.021)			(0.013)			(0.009)			(0.013)
Wealth quintile 5 (richest)			0.016			0.016			-0.004			-0.028**
			(0.020)			(0.012)			(0.007)			(0.013)
JSS	-0.156***	-0.153***	-0.156***	0.029***	0.028***	0.028***	0.057***	0.057***	0.057***	0.071***	0.068***	0.071***
	(0.016)	(0.016)	(0.016)	(0.009)	(0.009)	(0.009)	(0.008)	(0.008)	(0.008)	(0.013)	(0.013)	(0.013)
SSS	-0.100***	-0.101***	-0.100***	0.089***	0.089***	0.090***	0.040***	0.040***	0.040***	-0.029**	-0.028**	-0.030***
	(0.019)	(0.019)	(0.019)	(0.014)	(0.014)	(0.014)	(0.009)	(0.009)	(0.009)	(0.011)	(0.011)	(0.011)
Constant	0.831***	0.800***	0.812***	0.068***	0.078***	0.059***	0.018***	0.014***	0.018**	0.083***	0.107***	0.111***

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
	(0.016)	(0.019)	(0.020)	(0.007)	(0.008)	(0.008)	(0.004)	(0.005)	(0.007)	(0.012)	(0.015)	(0.015)
Observations	10007	10007	9991	10007	10007	9991	10007	10007	9991	10007	10007	9991

Note: Standard errors in parentheses and corrected for clustering at the EA level. Results for respondents currently enrolled.

* p < .10, ** p < .05, *** p < .01

Source: Analysis of IFLS 5 data

Table A.4: Subgroup results for educational attainment over time

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
	Completed at least primary school			**Completed at least junior secondary school**			**Completed at least senior secondary school**		
	2014	**2007**	**2000**	**2014**	**2007**	**2000**	**2014**	**2007**	**2000**
Male	-0.008*	-0.000	0.015**	-0.003	0.021**	0.057***	0.043***	0.050***	0.050***
	(0.005)	(0.006)	(0.006)	(0.010)	(0.009)	(0.012)	(0.013)	(0.011)	(0.012)
Constant	0.959***	0.938***	0.909***	0.807***	0.739***	0.633***	0.590***	0.528***	0.481***
	(0.005)	(0.006)	(0.008)	(0.015)	(0.014)	(0.015)	(0.017)	(0.016)	(0.016)
Observations	9912	10515	8758	9912	10515	8758	9912	10515	8758

Note: Standard errors in parentheses and corrected for clustering at the EA level. Results for respondents between 20 and 30 years old.
* p < .10, ** p < .05, *** p < .01

Source: Analysis of IFLS 3, 4, 5 data

Table A.5: Subgroup results for learning for 18–30 year olds

	(1)	(2)	(3)	(4)
	IRT 2014	**IRT 2014**	**IRT 2014**	**IRT 2000-2014**
Male	-0.023***			
	(0.005)			
Other than Java/Bali		-0.043***		
		(0.007)		
Wealth quintile 2			0.004	
			(0.010)	
Wealth quintile 3			0.012	
			(0.009)	
Wealth quintile 4			0.014	
			(0.009)	
Wealth quintile 5 (richest)			0.013	
			(0.009)	
2007				0.009*
				(0.005)
2014				-0.047***
				(0.005)
Constant	0.231***	0.237***	0.213***	0.222***
	(0.012)	(0.013)	(0.013)	(0.007)
Highest grade completed fixed effects	Yes	Yes	Yes	Yes
Observations	9348	9348	8186	14572

Note: The sample for the regression in the last column is restricted to individuals between 18 and 24 years old. Standard errors in parentheses and corrected for clustering at the EA level.
** p < .10, ** p < .05, *** p < .01*

Source: Analysis of IFLS 3, 4, 5 data

Table A.6: Subgroup results for currently enrolled students

	(1) IRT 2014	(2) IRT 2014	(3) IRT 2014	(4) IRT 2000-2014
Male	-0.040*** (0.005)			
Other than Java/Bali		-0.026*** (0.008)		
Wealth quintile 2			0.001 (0.008)	
Wealth quintile 3			0.005 (0.009)	
Wealth quintile 4			0.012 (0.009)	
Wealth quintile 5 (richest)			0.026*** (0.009)	
2007				-0.016*** (0.005)
2014				-0.061*** (0.005)
Constant	0.245*** (0.012)	0.470*** (0.009)	0.459*** (0.011)	0.531*** (0.007)
Highest grade completed fixed effects	Yes	Yes	Yes	Yes
Observations	10468	10468	8954	26614

Note: Standard errors in parentheses and corrected for clustering at the EA level.
** p < .10, ** p < .05, *** p < .01*

Source: Analysis of IFLS 3, 4, 5 data

Table A.7: Comparable questions between IFLS and BERMUTU

Variable	IFLS	BERMUTU	BERMUTU	BERMUTU
ek13	49-23	27-23 or 34-24 (Grade 1)	87-25 (Grade 2)	
ek14	267+112-189	235+157 or 289+189 (Grade 2, a)	452-235 or 492-368 (Grade 2, b)	21+38-41 (Grade 3)
ek15	(8+9)*3	35*6 or 52*9 (Grade 3)		
ek16	56/84	-		
ek17	1/3-1/6	2/8+3/8 or 5/12+3/12 (Grade 3)	5/12+3/12 (Grade 4, a)	1/4+3/10 or 3/5+2/7 (Grade 4, b)

Source: Analysis of IFLS 4, 5 data and BERMUTU 2011 data

Table A.8: Comparison of numeracy results between IFLS and BERMUTU

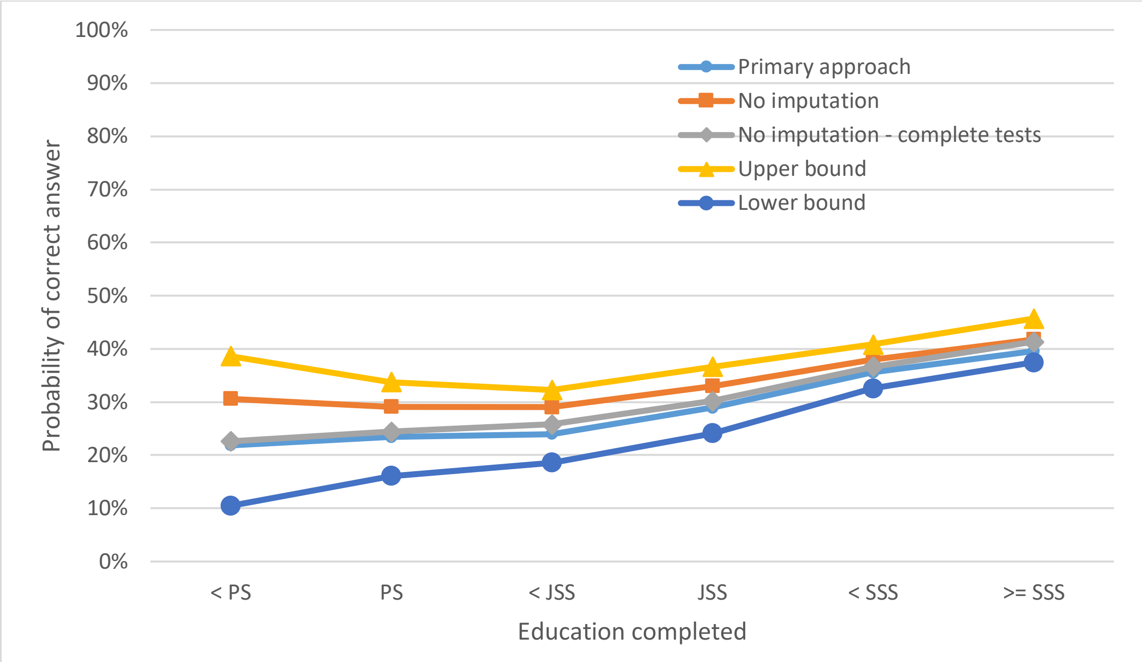
	BERMUTU 2011			IFLS 2014			IFLS 2007		
Variable	**Mean**	**Std. Dev.**	**Obs**	**Mean**	**Std. Dev.**	**Obs**	**Mean**	**Std. Dev.**	**Obs**
ek13 Grade 1	0.22	0.75	6,840	0.46	0.72	457	0.65	0.64	620
ek13 Grade2	0.61	0.66	6,868	0.63	0.65	997	0.77	0.54	970
ek14 Grade 2 (a)	0.30	0.75	6,868	0.41	0.73	997	0.51	0.70	970
ek14 Grade 2 (b)	0.19	0.75	6,868	0.41	0.73	997	0.51	0.70	970
ek14 Grade 3	0.63	0.60	6,850	0.50	0.71	1053	0.57	0.68	851
ek15 Grade 3	0.31	0.67	6,850	0.47	0.72	1053	0.43	0.73	851
ek17 Grade 3	0.47	0.65	6,850	0.02	0.59	1053	0.04	0.60	851
ek17 Grade 4 (a)	0.81	0.46	6,847	0.02	0.59	1013	0.00	0.57	801
ek17 Grade 4 (b)	0.00	0.57	6,847	0.02	0.59	1013	0.00	0.57	801

Note: means are corrected for guessing. The results are shown for students currently enrolled in the grades listed in the first column.

Source: Analysis of IFLS 4, 5 data and BERMUTU 2011 data

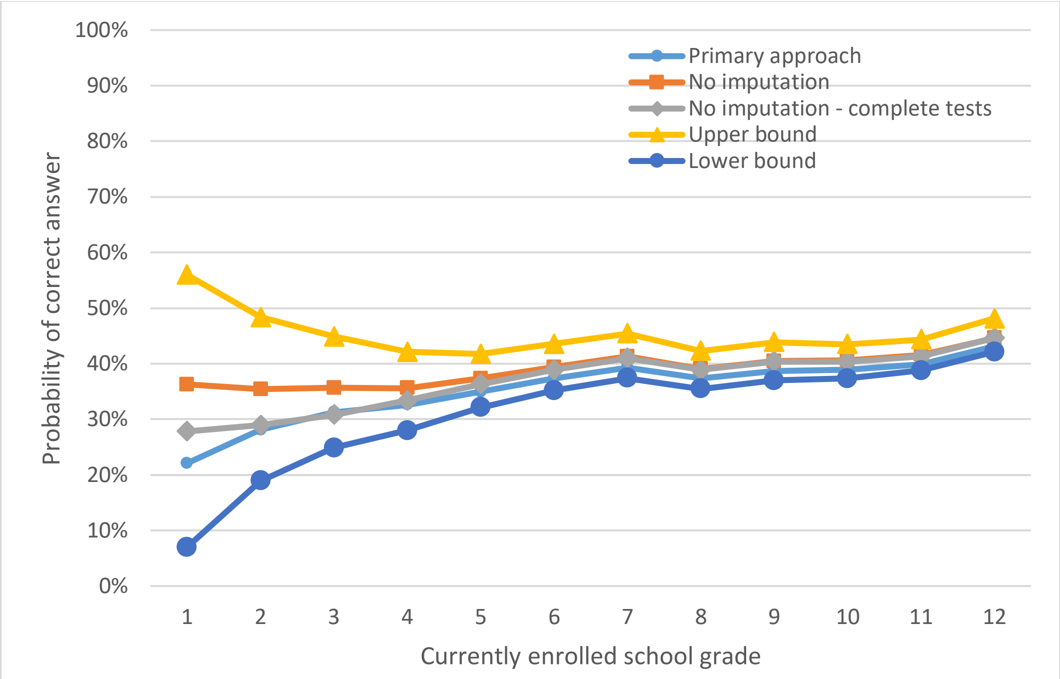
Appendix 2: Figures

Figure A.1: Learning by education level completed (imputation approaches)



Source: IFLS 5, 2014**Note**: Results show the mean probability of answering a math question correctly among 18 to 30 year-olds. Results are adjusted for guessing as described in section 2. Students who graduated high school, including those who went onto to postsecondary education, are included in the >=SSS group. The primary approach uses random guessing imputation for item-level missing values and is shown in Figure 4.3. No imputation shows IRT results without imputation for item-level questions (this method still generates an IRT value for each respondent but only uses non-missing and non-imputed data to do so). No imputation—complete tests—shows IRT results using only respondents without missing data for item-level questions. Upper bound imputation shows results if all item-level missing responses received an imputation of 1. Lower bound imputation shows results if all item-level missing responses received an imputation of 0.

Figure A.2: Learning by grade level (imputation approaches)



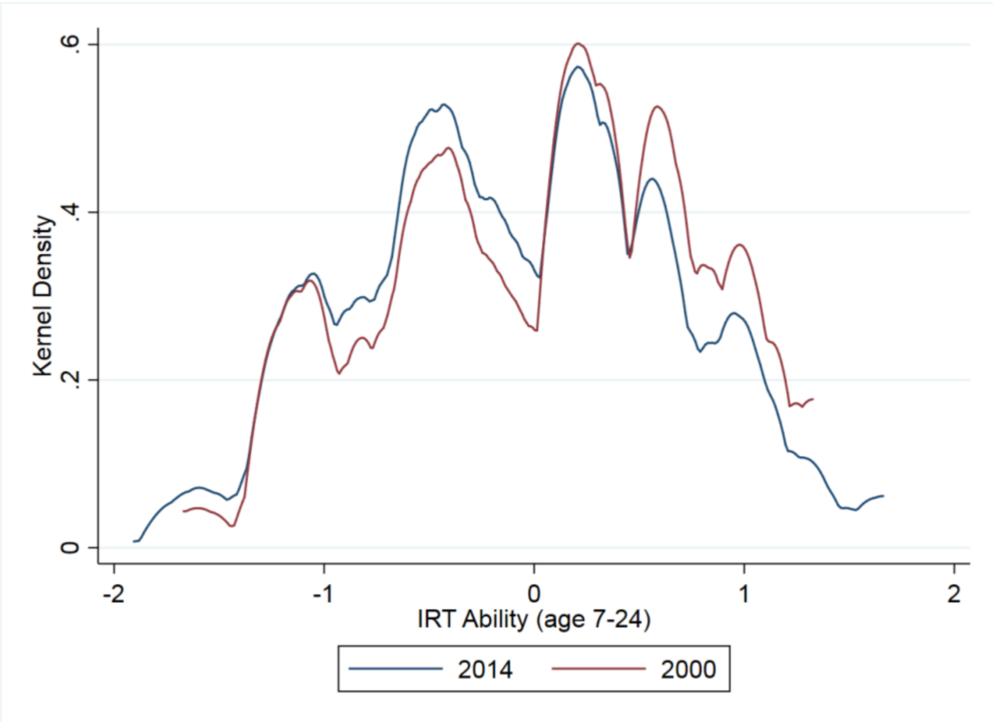
Source: IFLS 5, 2014
Note: Results show the mean probability of answering a math question correctly among currently enrolled students by current grade level. Results are adjusted for guessing as described in section 2. The primary approach uses random guessing imputation for item-level missing values and is shown in Figure 4.6. No imputation shows IRT results without imputation for item-level questions (this method still generates an IRT value for each respondent but only uses non-missing and non-imputed data to do so). No imputation – complete tests -- shows IRT results using only respondents without missing data for item-level questions. Upper bound imputation shows results if all item-level missing responses received an imputation of 1. Lower bound imputation shows results if all item-level missing responses received an imputation of 0.

Appendix 3: Psychometrics of Numeracy Score[1]

This appendix presents the psychometric properties of the test items. We indicate the items using the variable name. The questions behind these items can be found in Table 2.1.

An important limitation of our analysis is the small number of items in the test. Therefore, the test is restricted in the extent to which it can distinguish different levels of ability. Figure A.3.1 shows the distribution of the ability score for 2000 and 2014, which has peaks around a score of -1, -0.5, 0.5, 0.75 and 1. Nevertheless, the test is able to capture ability levels ranging from -1.9 to 1.7. The rest of this appendix tests the quality of our instrument, keeping this limitation in mind.

Figure A.3.1: Kernel Density plot of numeracy score by year for 7 to 24 year-old respondents



Source: Analysis of IFLS 3, 4, 5 and East data

Validity

Factor analysis

The factor analysis shows that the items are almost unidimensional. Only ek16 and ek18 have the highest correlation with the second factor. Note that these items indicate the same question (56/84), but they were part of two different versions. However, the correlation of

[1] Edwards, M. C., Houts, C. R., & Cai, L. (2018). A diagnostic procedure to detect departures from local independence in item response theory models. *Psychological methods*, *23*(1), 138.

these items with the first factor is relatively high as well. We conclude that the items measure the same thing, namely numeracy skills.

Table A.3.1: Factor analysis of numeracy items

Variable	Factor1	Factor2	Factor3	Uniqueness
ek13	**0.2575**	-0.1298	0.1881	0.8815
ek14	**0.4297**	-0.225	0.1714	0.7353
ek15	**0.4713**	-0.2277	0.1296	0.7092
ek16	0.3435	**0.4496**	0.1266	0.6638
ek17	**0.4215**	0.094	-0.1612	0.7875
ek18	0.3433	**0.4575**	0.1062	0.6615
ek19	**0.5464**	-0.1459	-0.0337	0.679
ek20	**0.4904**	-0.1338	-0.0894	0.7336
ek21	**0.3502**	0.0608	-0.1755	0.8428
ek22	**0.4291**	0.0241	-0.1666	0.7875

Source: Analysis of IFLS 3, 4, 5 and East data

Infit and outfit to the Rasch model

We also check the validity of the score by estimating the infit and outfit mean squares in the table below. When these values are between 0.5 and 1.5, the items are productive for the measurement[1]. The items have a good outfit index, namely 0.86–1.32. This means that the numeracy score contains little noise. The good infit index shows that the items do not follow unexpected patterns (0.88–1.16).

[1] https://www.winsteps.com/winman/misfitdiagnosis.htm

Table A.3.2: Infit and outfit of numeracy items

			INFIT		OUTFIT	
ITEM	MEASURE	MODEL S.E.	MNSQ	ZSTD	MNSQ	ZSTD
ek13	-2.44	0.02	1.04	3.6	1.32	9.9
ek14	-1.39	0.02	0.88	-9.9	0.86	-9.5
ek15	-1.14	0.01	0.86	-9.9	0.81	-9.9
ek16	0.92	0.01	1.08	9.9	1.27	9.9
ek17	1.46	0.02	1.04	4.8	1.32	9.9
ek18	0.77	0.02	1.16	9.9	1.24	9.9
ek19	0.23	0.01	0.88	-9.9	0.86	-9.9
ek20	-0.17	0.02	0.91	-9.9	0.89	-9.9
ek21	1.46	0.02	1.07	8.2	1.17	9.9
ek22	0.3	0.01	0.99	-1.9	1	-0.3
MEAN	0	0.02	0.99	-0.5	1.07	1
S.D.	1.23	0	0.1	8.3	0.2	9.3

Source: Analysis of IFLS 3, 4, 5 and East data

Reliability

Cronbach's alpha

Cronbach's alpha measures the internal consistency of the score items. The results say that the item-test correlations are good (above 0.25), but the alpha of the test scale is relatively low as we prefer an alpha between 0.75 and 0.95. Generally, an alpha of at least 0.7 is considered acceptable[1], and the alpha of this numeracy score is slightly below this threshold. This could indicate that the test needs more items to achieve high consistency.

Table A.3.3: Cronbach's alpha of numeracy items

Item	Observations	Sign	item-test correlation	item-rest correlation	average interitem covariance	alpha
ek13	32776	+	0.4589	0.2301	0.0396912	0.6561
ek14	32776	+	0.5793	0.3200	0.0358025	0.6372
ek15	32776	+	0.5951	0.3256	0.0350671	0.634
ek16	32776	+	0.4946	0.1771	0.0381294	0.6608
ek17	32776	+	0.5091	0.2084	0.0371917	0.6509
ek18	27221	+	0.4762	0.2056	0.0368608	0.6516
ek19	27221	+	0.6259	0.3887	0.0319867	0.6137
ek20	27221	+	0.6140	0.3737	0.0327994	0.6205
ek21	27221	+	0.5107	0.2695	0.0359332	0.6411
ek22	27221	+	0.5675	0.3135	0.0339778	0.6304
Test scale					**0.0357208**	**0.6639**

Source: Analysis of IFLS 3, 4, 5 and East data

[1] https://stats.idre.ucla.edu/spss/faq/what-does-cronbachs-alpha-mean/

Item Response Theory assumptions and output

The table below shows the IRT 2pl (item response theory two-parameter model) results for the difficulty and discrimination power coefficients, ordered by difficulty level. The item characteristics curves in the graph below visualise these results. The higher the discrimination power coefficient, the steeper the curve and the higher the difficulty coefficient, the more the curve is placed to the right. The difficulty coefficients show that the difficulty levels of the items as found in the data partly follow the curriculum. The pattern shows that division is a skill that the sample struggled with. Questions ek13 to ek20 are about addition and subtraction without any division, fraction or percentage, while questions ek19 to ek17 all include at least one division (see Table 2.1). The Grade 3 division question is also found in the data as the easiest one of that category. However, the percentage questions from Grade 5 seem to be relatively easy compared to the fraction questions from Grade 4.

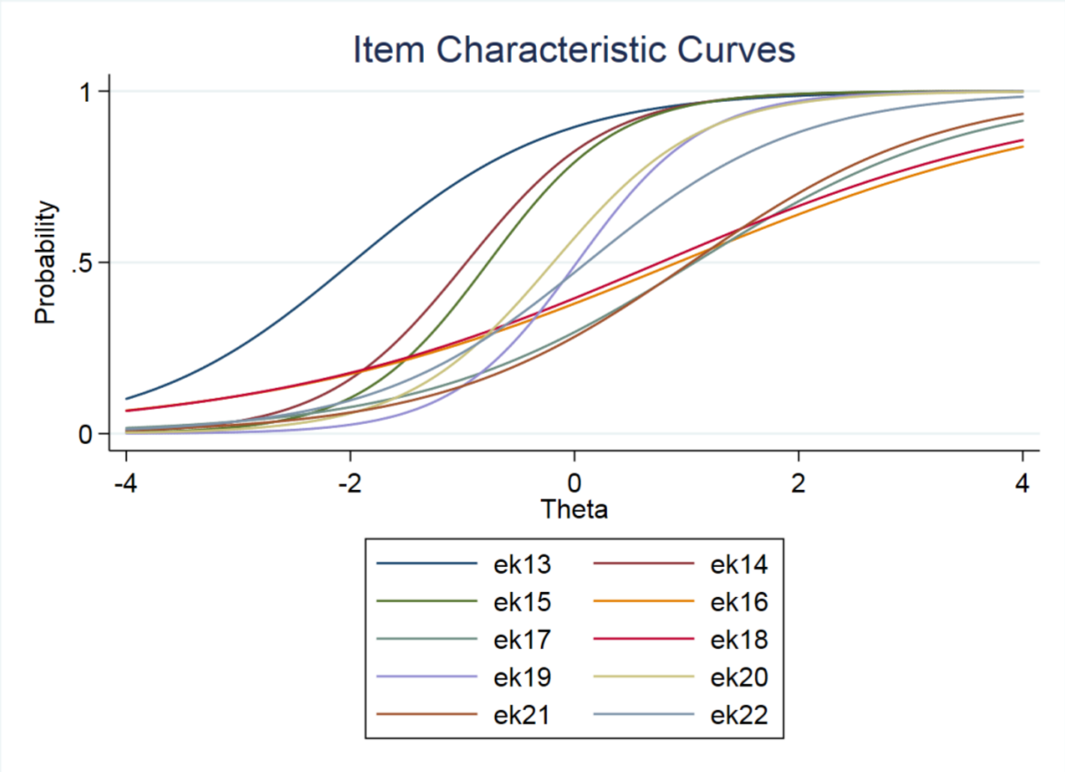
The relatively difficult questions have lower discrimination power scores, shown in the graph by straighter lines. This means that these questions are limited in distinguishing between low and high levels of ability (theta), but they still add information to the analysis since the coefficients are significantly different from zero. Therefore, in order to maximise the information we use from the test, we leave these items in the IRT analysis.

Table A.3.4: IRT difficulty and discrimination coefficients

Item	Grade	Difficulty Coefficient	Robust s.e.	p-value
ek13	1	-1.985871	0.0612384	0.000
ek14	2	-0.9650155	0.0307568	0.000
ek15	3	-0.7691838	0.0279077	0.000
ek20	4	-0.190352	0.0264657	0.000
ek19	3	0.0209208	0.0278497	0.453
ek22	5	0.1102797	0.0372446	0.003
ek18	4	0.7656642	0.0589496	0.000
ek16	4	0.9187338	0.0637493	0.000
ek21	5	1.040347	0.0564288	0.000
ek17	4	1.072711	0.0663508	0.000
Item	Grade	Discrimination Coefficient	Robust s.e.	p-value
ek13	1	1.08125	0.0339137	0.000
ek14	2	1.604345	0.0461243	0.000
ek15	3	1.743684	0.0490808	0.000
ek20	4	1.524966	0.0439502	0.000
ek19	3	1.789066	0.0511086	0.000
ek22	5	1.054839	0.0359237	0.000
ek18	4	0.5543347	0.028319	0.000
ek16	4	0.5334883	0.0272455	0.000
ek21	5	0.8940243	0.0371736	0.000
ek17	4	0.8054867	0.0356571	0.000

Source: Analysis of IFLS 3, 4, 5 and East data

Table A.3.5: IRT item characteristic curves



Source: Analysis of IFLS 3, 4, 5 and East data

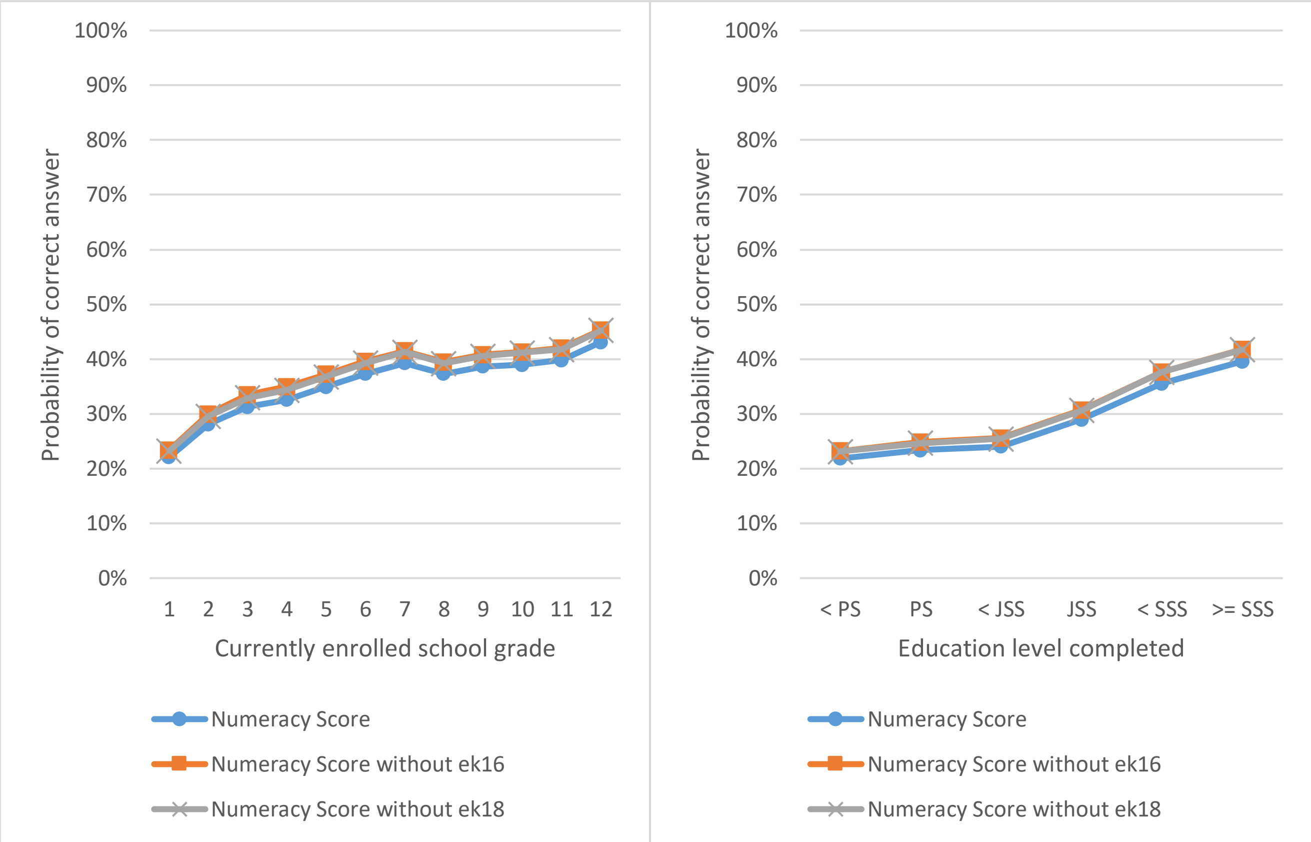
These IRT results are based on three core assumptions:

1. Unidimensionality: A single latent individual-specific trait determines performance on the test
2. No differential item functioning: Item characteristics are person-invariant, such that the ability (theta) measures the same level of numeracy skills across individuals with different characteristics
3. (Conditional) local independence: Conditional on ability, item responses are independent

If items function differently across subgroups, violating the second assumption, or if responses of an individual are dependent across items for reasons unrelated to ability, violating the third assumption, this means that there is another underlying trait besides ability. We argue that these assumptions hold in our analysis.

The factor analysis in Table A.2.1 showed that our data complies with the first assumption, although the correlation between ek16 and ek18 could be an issue for the conditional local independence assumption. Therefore, we do a robustness check by removing ek16 and ek18 sequentially from the analysis. This is a commonly used method for resolving local dependence. As described by Edwards, Houts and Cai (2018), removing one of the locally dependent items would cause the slope (or the discrimination coefficient) of the other item to substantially decrease. In our case, removing ek18 decreases the discrimination coefficient of ek16 from 0.508 to 0.507, and the discrimination coefficient of ek18 changes from 0.296 to 0.295 when removing ek16. Moreover, when plotting the numeracy scores without ek16 or ek18 in Figure A.2.3, the results are very similar to the score that includes both of these items. The scores are slightly higher, but removing the items does not alter the slope of the learning profiles. Because of the small changes in the discrimination coefficients and the numeracy score, we conclude that local dependence between ek16 and ek18 is not an issue in our analysis.

Table A.3.6: Numeracy score when removing ek16 or ek18 for currently enrolled students (left) and 18 to 30 year old individuals (right) in 2014



To test for compliance with the second assumption, we present a logistic Differential Item Functioning analysis. In the paper, we compare learning by gender, wealth, region and year. We test if items function similarly across these groups, or in other words, if the item characteristic curves are the same. The results are presented in Figure A.3.5. Overall, the item characteristic curves overlap for most items. We only find substantial differences between the item characteristic curves for EK16 and EK17 by wealth quintile and for EK17 by year. An important finding in our analysis is that learning in 2014 was lower than learning in 2000. We want to make sure that the differential item functioning in item ek17 does not drive this finding. Therefore, we check if excluding EK17 from the IRT analysis changes the results. Figure A.3.4 presents the numeracy score by year when excluding EK17 and shows that the differential item functioning does not drive the decrease in learning over time.

Table A.3.7: Numeracy score for 18 to 24 year olds in 2000 and 2014 excluding EK17

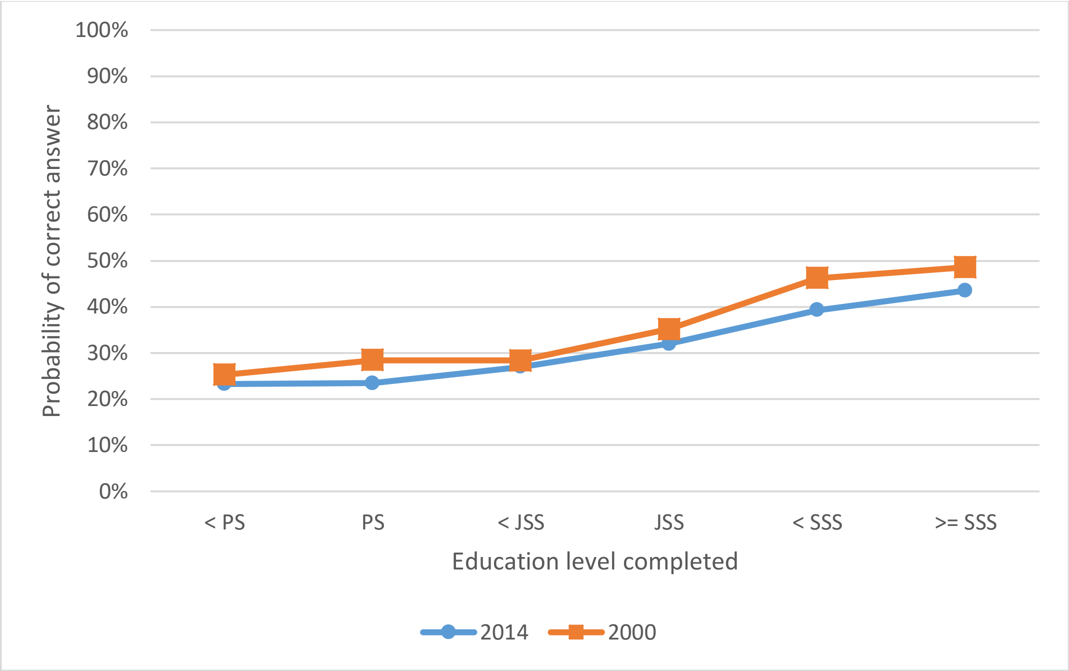
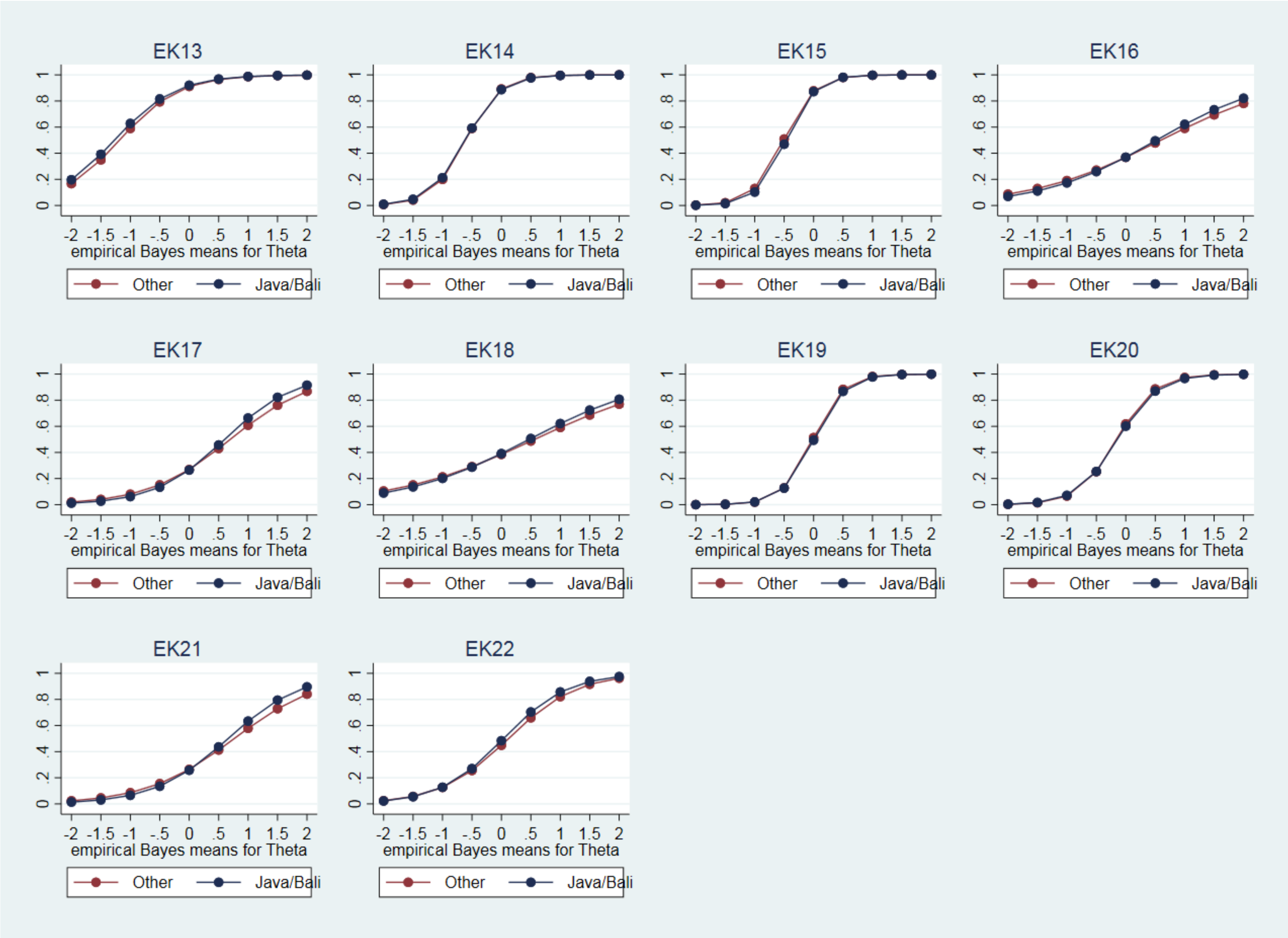


Table A.3.8: Differential Item Functioning graphs by gender, wealth, region and year


EK13
EK14
EK15
EK16
EK17
EK18
EK19
EK20
EK21
EK22
0 .2 .4 .6 .8 1
-2 -1.5 -1 -.5 0 .5 1 1.5 2
empirical Bayes means for Theta
Poor
Rich


EK13
EK14
EK15
EK16
EK17
EK18
EK19
EK20
EK21
EK22
0 .2 .4 .6 .8 1
-2 -1.5 -1 -.5 0 .5 1 1.5 2
empirical Bayes means for Theta
Other
Java/Bali

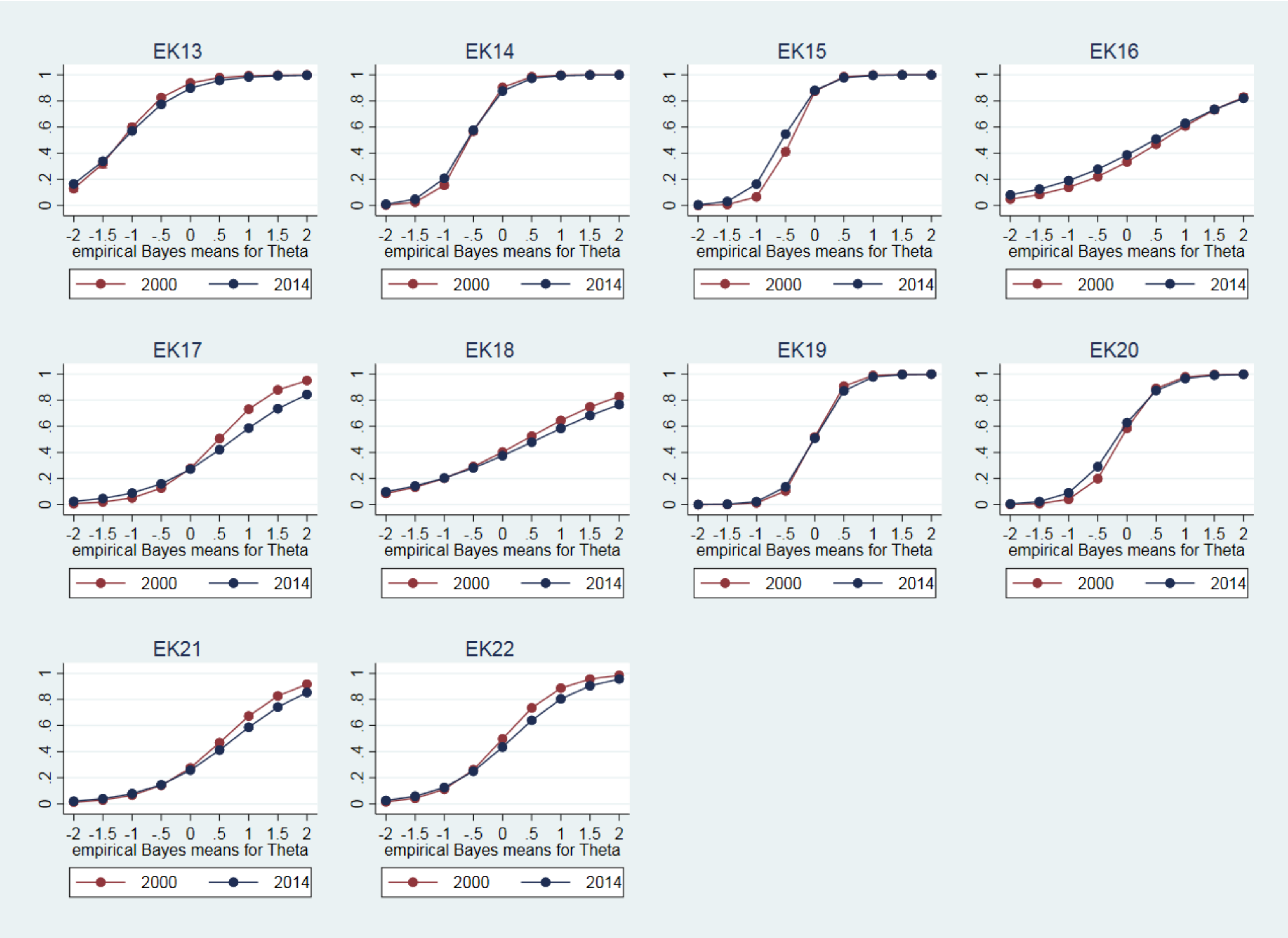
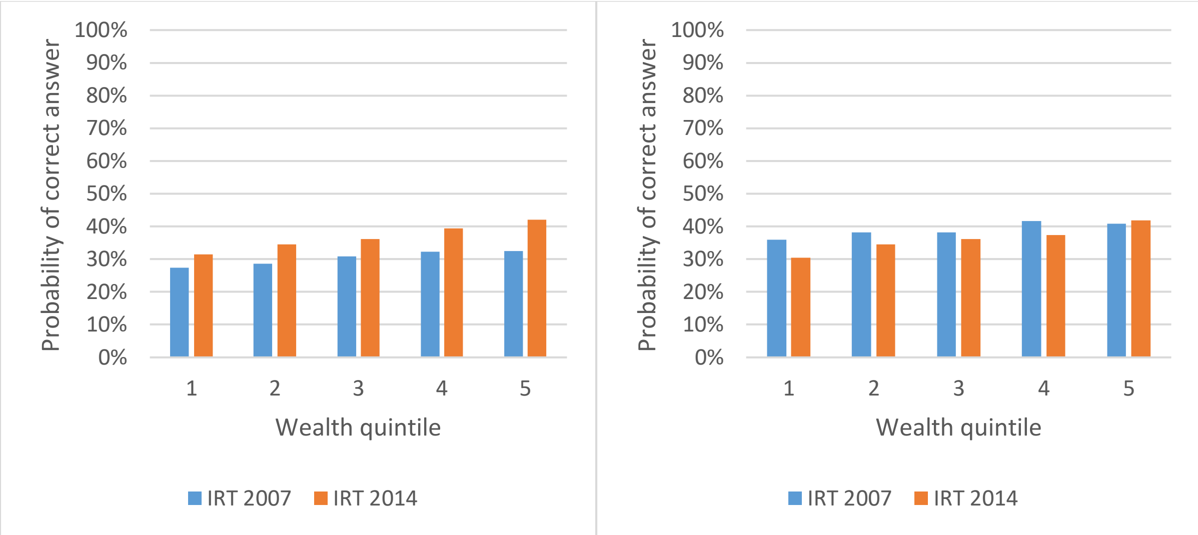

EK13
EK14
EK15
EK16
EK17
EK18
EK19
EK20
EK21
EK22
0 .2 .4 .6 .8 1
-2 -1.5 -1 -.5 0 .5 1 1.5 2
empirical Bayes means for Theta
2000
2014

Figure A.0.1 Learning over time for 7- to 9-year-olds (a) and 10- to 12-year-olds (b) by wealth quintile



Source: Analysis of IFLS4 and IFLS 5 data
Note: Results show the mean probability of answering a math question correctly among each group of respondents by wealth quintile. The quintiles are based on the asset index in 2007 described in section 2. The frame A sample (left side) comprises 7- to 9-year-olds in 2007 and the frame B sample (right side) comprises 10- to 12-year-olds in 2007.