Working Paper 179

Young Lives

Predictors of Mathematics and Literacy Skills at 15 Years Old in Ethiopia, India, Peru and Vietnam

A Longitudinal Analysis

Colin Tredoux and Andrew Dawes

www.younglives.org.uk

AUGUST 2018

Predictors of Mathematics and Literacy Skills at 15 Years Old in Ethiopia, India, Peru and Vietnam: A Longitudinal Analysis

Colin Tredoux and Andrew Dawes

ISBN 978-1-912485-04-8

A catalogue record for this publication is available from the British Library.

Core-funded by

UKaid
from the British people

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

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

Contents

The authors

Colin Tredoux is Professor of Psychology at the University of Cape Town, South Africa, and Chaire d'Attractivité at the Université de Toulouse, Mirail, France. He researches a variety of topics in social and cognitive psychology, especially in intergroup contact and segregation. He has a keen interest in quantitative data analysis and consults on this widely.

Andrew Dawes is Associate Professor Emeritus in Psychology at the University of Cape Town and a Research Associate with Young Lives. His primary research interest is the development of children affected by poverty, and in particular, the effectiveness of interventions to improve early learning outcomes in preschool children growing up in disadvantaged circumstances. Other research interests include the development of indicators for measuring children's well-being, and the prevention of maltreatment and violence to young children. He has extensive experience in translating research to policy.

Acknowledgements

The authors are very grateful for the support of the Young Lives team throughout this complex exercise, and in particular Liza Benny, Kristine Briones, Grace Chang, Santiago Cueto, Paul Dornan, Patricia Espinoza, Marta Favara, Zoe James, Juan Leon, Gabriella Munares, Jack Rossiter, Caine Rolleston, Anne Solon, and Martin Woodhead. We thank the Technical Advisory Group on this project and the Young Lives International Advisory Board for their comments on earlier presentations.

Summary

This working paper reports on longitudinal analyses of predictors of growth in receptive vocabulary and mathematics abilities by age 15 conducted using data from the Younger Cohort in the four Young Lives countries. Theoretical models of plausible influences on these outcomes were constructed. Latent Growth Modelling undertaken on these two key aspects of development provides unusually powerful cross-country evidence that household economic well-being and proximal relationships in the early childhood environment are particularly important, either placing young children at risk for developmental hazards such as growth stunting or conferring advantage for future development.

The models indicate that children whose mothers are rendered psychologically vulnerable by the stresses of poverty are at risk of poor physical growth in the early years, and that this impacts language, cognition and numeracy skills across childhood. Early disadvantages in poorer children are compounded by challenges of combining household responsibilities with schooling from middle childhood through adolescence. Those children with better-educated mothers are more likely to attend preschool programmes, which contribute to improved learning outcomes. While not directly measured, it is likely that those who made most gains attended programmes that were better provisioned. While there is variation across the countries, the high degree of similarity suggests good external validity of the general findings.

About Young Lives

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

1. Introduction

How, and at what points in the life course do poverty and other influences shape the development of literacy and numeracy abilities in Young Lives children from early childhood to the age of 15? This working paper seeks to address this question. The longitudinal design and large sample sizes in each of the four Young Lives countries permit us to explore this question using Latent Growth Modelling (LGM). Duncan and Duncan (2009) express the contribution of this approach to the analysis of longitudinal data as follows:

> 'Building upon traditional longitudinal models, methodologists have extended the latent variable framework to accommodate longitudinal models that include multivariate or higher-order specifications, multiple populations, the accelerated collection of longitudinal data, multilevel or hierarchical structures, and complex relations including mediation, moderation, and reciprocal causation.' (2009: 2)

They go on to note that as an approach to modelling developmental processes, LGM permits description of both individual trajectories over time (e.g. in mathematics test performance), as well as differences in trajectories between individuals that are occasioned by variation in their backgrounds and in other influences to which they may have been exposed (predictors of the trajectories).

Our approach follows Collins (2006: 505), who maintains that:

> 'ideal longitudinal research is characterized by the seamless integration of three elements: (a) a well-articulated theoretical model of change observed using (b) a temporal design that affords a clear and detailed view of the process, with the resulting data analyzed by means of (c) a statistical model that is an operationalization of the theoretical model.'

1.1. Theoretical models of change

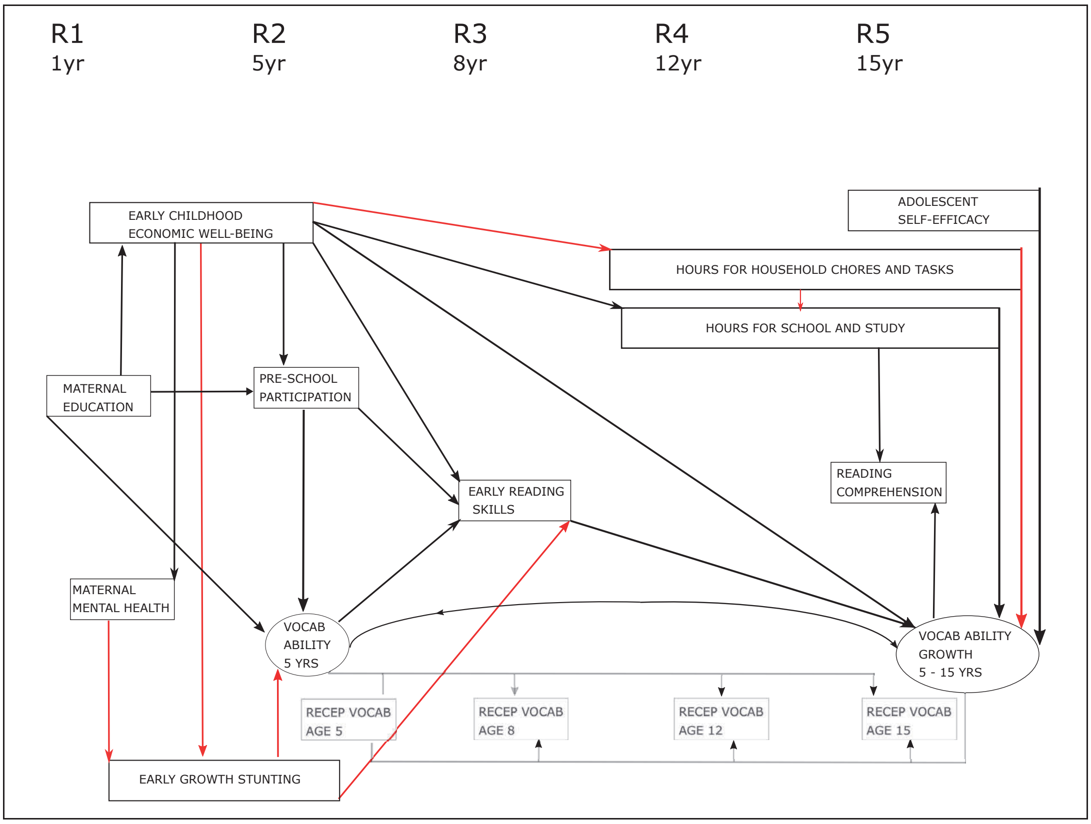
In constructing the theoretical models to be tested separately on children in each of the four countries, we have drawn on the literature presented in the rationale that follows. In the interests of brevity we have not cited all of the many Young Lives papers that address the issues of interest, but have rather mentioned those of particular relevance, and have no doubt missed some. We note the variables used in the modelling as we proceed (see Table 1a). Those tested and discarded are listed in Table 1b.

The outcomes of interest at 15 years old, just prior to the final phase of high school, are growth in receptive vocabulary from age 5 to 15 years, reading comprehension at 15, and growth in numeracy skills from 8 to 15 years. Competence in these areas is associated with later educational performance and improvement in life chances once schooling has been completed. Early mathematics skills are strongly predictive of later school success – more so than language ability, and those who perform well in early grades continue to stay ahead (Behrman et al. 2006; Siegler et al. 2012; Duncan et al. 2007). In developing countries, each additional year of education increases a person's earnings by about 11 per cent (Psacharopoulos and Patrinos 2004). Understanding sources of influence on these skills has implications for policy and supportive interventions.

The theoretical models we developed are similar for all the countries. Figure 1 shows the model used for *Receptive vocabulary growth and Reading comprehension* for India, Ethiopia and Vietnam. The model of *Mathematics growth* for these countries is presented in Figure 2.

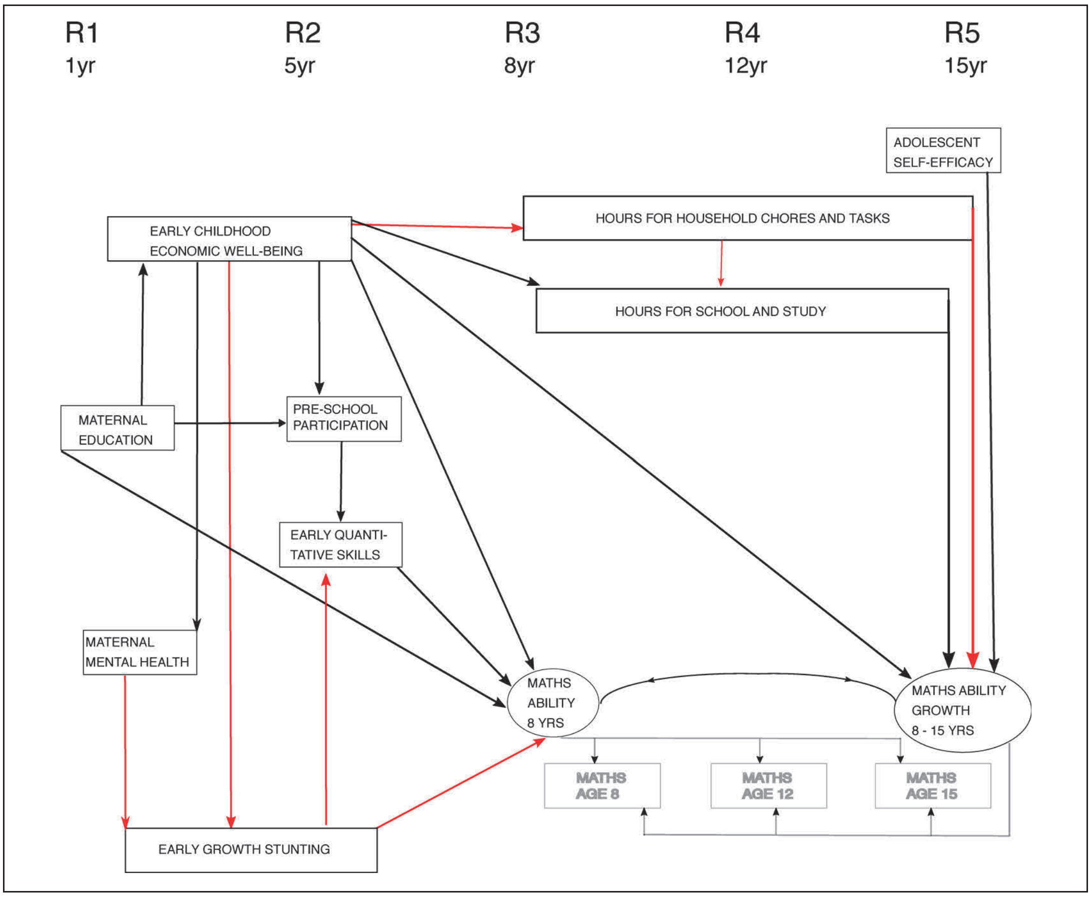
Peru is an exception as it has certain variables that were not available for other countries, and the theoretical models for Peru are presented when it is considered.

Figure 1. *General theoretical model of conditional development of receptive vocabulary*



This model applies to Ethiopia, India, and Vietnam. It is also the base for the more complex model for Peru presented when that country is discussed. Black lines represent predicted positive directional effects, and red lines negative directional effects. Change in receptive vocabulary over time is represented as a Latent Growth Curve (with latent intercept and slope variables, and their covariance [the double-headed arrow]), constituted by tests administered at four time points (ages). Predictor variables are hypothesised to have a particular chronological precedence, as shown in the figure. Ellipsoid and circle shapes represent latent variables, and rectangular shapes represent manifest variables. Empirical findings for receptive vocabulary are presented in Section 2.1.

Figure 2. *General theoretical model of conditional development of mathematics ability*



As with receptive vocabulary, this model covers Ethiopia, India, and Vietnam and is the base for the Peru model. Conventions for red and black lines are as described above. Change in mathematics performance over time is represented as a Latent Growth Curve (with latent intercept and slope variables, and their covariance [the double-headed arrow]), constituted by tests administered at three time points (ages). Predictor variables are hypothesised to have a particular chronological precedence, as shown in the figure. Ellipsoid and circle shapes represent latent variables, and rectangular shapes represent manifest variables. Empirical findings for mathematics are presented in Section 2.2.

1.1.1. Rationale

In what follows we provide a rationale for the inclusion of predictor variables during three developmental periods: early and middle childhood, and adolescence. Inclusion was determined by availability in the Young Lives datasets and suitability for analysis.

Early childhood

For this analysis early childhood is understood as commencing in conception and ending at 5 years of age. Socio-economic well-being in early childhood is associated with a wide array of developmental outcomes (McLoyd 1998), and research from around the world has consistently found relationships between household wealth in childhood, school achievement, and the development of language and mathematical skills in adolescence and early adulthood (e.g. Gibb, Fergusson and Horwood 2012; Duncan, Ziol-Guest and Kalil 2010; Duncan et al. 2007; Ekber and Çelik 2009; White 1982). Evidence from Young Lives has shown that measures of material wealth and living standards, such as consumption per capita, are significantly associated with a child's skill acquisition and subjective well-being (e.g. Dercon and Krishnan 2009). Cueto et al. (2014) find that household wealth when the child is 12 months old is related to mathematics achievement 10 years later.

We use the Young Lives wealth index (Briones 2017) aggregated at Rounds 1 and 2 of the study (when the children were on average 12 and 60 months old respectively), as our measure of *Early childhood economic well-being.* In the model for *Vocabulary growth* (Figure 1), *Early childhood economic well-being* was expected to predict *Maternal mental health, Early growth stunting*, *Preschool participation, Receptive vocabulary* at 5 years and *Growth* by 15 years*, Early reading skills* at 8 years old, the time children are required to devote to household responsibilities (*Hours for household chores and tasks*) and the time they have available for education (*Time for school and studies*). The influence of *Early childhood economic well-being* was expected to be similar for *Mathematics growth* (Figure 2).

Young children spend most of their time within the context of the home where household members – particularly the main caregivers (commonly mothers) – have particular influences on their development. A long-recognised contributor to children's educational outcomes is caregiver education (e.g. Mercy and Steelman 1982). Young Lives research supports these observations, and also shows that better-educated parents are more likely to send their children to school, to report higher aspirations for their children, and to invest more in their education (Dercon and Krishnan 2009). We used *Maternal education* (in years) as a predictor variable (Table 1) recorded at Round 1, as preliminary analyses indicated that there was little evidence of change in mothers' levels of education over study rounds. This variable was expected to predict participation in preschool and both receptive vocabulary at 5 years and mathematics ability at 8 years. *Early childhood economic well-being* was also expected to influence both abilities from early childhood through early adolescence.

Maternal psychological well-being is associated with better care of young children and can be compromised by poverty and its associated challenges (Lund et al. 2010). Ongoing psychological distress reduces the capacity to provide adequate care and stimulation and is associated with increased likelihood of growth stunting and other negative outcomes (Wachs and Rahman 2013). Young Lives administered the World Health Organization SRQ20 screening tool for Common Mental Disorder (CMD) risk (1994) at Round 1 of the study (and beyond). Following Wachs and Rachman, we were interested in early influences of maternal psychological distress on young children. We also note Harpham and colleagues' (2005) report of a relationship between high rates of maternal CMD and poor child nutritional status in two Young Lives countries: India and Vietnam. SRQ20 scores were used as the measure of *Maternal mental health* (higher scores indicating less risk, after reversing the direction of the original scale) when the child was 12 months old. Scores on the SRQ20 were expected to predict *Early growth stunting*. We expected the effects of *Maternal mental health* on early *Receptive vocabulary and quantitative skills* at 5 years, and *Mathematics ability* at 8 years old, to be mediated by the child's growth status.

Growth status in early childhood has a powerful influence on early neurological development and is associated with cognitive growth and with competencies such as mathematics and language in early childhood and in later years (e.g. Grantham-McGregor et al. 2007; Duc and Behrman 2017). Using Young Lives data, Crookston and colleagues (2013) found that children who were stunted at 12 months were more likely to be over-age for school grade at 8 years old, and to have lower mathematics, vocabulary and reading scores than those whose growth was not compromised. Recovery from stunting has been observed in Young Lives children and has been associated with improved cognitive outcomes (Georgiadis et al. 2017). We did devise an approach that would capture recovery and faltering through to 15 years of age. However, on testing, it added considerable complexity and was difficult to interpret. We therefore opted for an early childhood indicator. For modelling, we used height-for-age z-scores (HAZ) at 12 months and 5 years old and used WHO reference norms (HAZ) as normal, stunted, or severely stunted at both points to derive an indicator of *Early growth stunting* (see Table 1 for the approach scoring; note that the categorisation of what is inherently a continuous variable is more common in this literature than leaving it in a continuous form). We are aware of the controversy in this area regarding the appropriate metric for tracking changes in growth status over time. For example, Lundeen and colleagues (2014) have compared HAZ scores with a score calculated from the difference in the child's height in centimetres from the 50th percentile of the reference population. They observe that 'mean HAZs increased, even though height deficits relative to the reference median also increased. These 2 metrics may result in different interpretations of the potential for and the impact of catch-up growth in height' (Lundeen et al. 2014: 821). We have not used the difference measure here.

Sustainable Development Goal 4.2 charges countries to provide access to quality preschool programmes for children from 3 years old. A considerable body of research conducted in both high-income countries and the developing regions (including Young Lives analyses) attests to the value of quality preschool experiences in the development of cognitive and psychosocial skills (e.g. Woldehanna and Araya 2017; Woodhead et al. 2017; Rebello-Britto, Engle and Super 2013; Nores and Barnett 2010). Young Lives collected preschool participation data (labelled 'attendance' in most papers), and while data were collected on months of preschool enrolment, data on daily attendance (a measure of programme exposure) were not measured. Programme quality was not systematically measured, although small-scale qualitative studies suggested particularly poor quality for the poorest children (Woodhead et al. 2009). However, type of provision was recorded as a proxy measure for quality (see below). These differ by country. As evident in Table 1, and for each country, categories of preschool programme were given progressively higher scores, from non-participation to private centre-based services (with private as a proxy for better quality). The scales were validated against the children's receptive vocabulary and mathematics test scores prior to inclusion (controlling for household wealth). This exercise showed that no participation was associated with lower scores, and private provision with higher scores. It is important to note, however, that private provision may vary considerably (Woodhead et al. 2009). Our variable *Preschool participation* was expected to predict *Early quantitative skills* (a mediator of *Mathematics ability* at 8 years), and to predict *Receptive vocabulary ability* (5 years), and *Early reading skills* (EGRA test at 8 years).

Children's early vocabulary, expressive language abilities, their comprehension and quantitative skills just prior to entering school are important predictors of later language and mathematics abilities, as well as school progress (Fernald et al. 2009).

From age 5 to 15 years, we measured growth in receptive vocabulary (our variable name is *Vocab ability*). In Peru, the Peabody Picture Vocabulary Test Spanish version (PPVT-R) (Dunn et al. 1986) was used for Castellano speakers and translated for Quechua children. In the other countries PPVT-III (Dunn and Dunn 1997) was translated into local languages and used in Rounds 2 (age 8) and 3 (age 12). However, Cueto and Leon (2012) report that translation was not always standardised, raising questions regarding the reliability of scores and validity of the test. They undertook psychometric analysis of the Round 3 PPVT-111 data in each country, other than Peru. Based on these analyses, revised and considerably shortened measures of receptive vocabulary (based on the words in PPVT-111 used in earlier rounds) were devised for Rounds 4 and 5. These took into account item difficulty in the main languages of each country. Leon and Singh (2017) have conducted psychometric analyses using item response theory (IRT) to derive transformed scores for tests of receptive vocabulary administered in Rounds 2 through 4 for India, Vietnam and Ethiopia. Unfortunately, this process had not been completed for Round 5 at the time of our analysis, which meant that their contribution could not be used; it is likely that using IRT-corrected scores would have made comparisons between participants, or comparisons within participants over time, more robust (Bortolotti et al. 2012).

It is essential to note that for modelling, we harmonised the stimulus words used in PPVT administration in Rounds 2 and 3 with the shorter list in the revised receptive vocabulary tests administered in Rounds 4 and 5. This made it possible to use the identical set of words for analysis across all the rounds for these three countries. We recognise a limitation that this procedure cannot take account of the order in which words were presented in Rounds 2 and 3, which was not identical to that for the later rounds. Our measure for all countries except Peru is a test of receptive vocabulary and is no longer the original PPVT. Languages used in the LGM are restricted to those on which IRT analyses were performed on Round 3 data (Table 1). Details may be found in Cueto and Leon (2012).[1]

At the end of the early childhood period (age 5) we included a measure of children's *Early quantitative skills* using a subscale of the Cognitive Developmental Assessment (CDA) as proxy for cognitive development and as a predictor of early performance in *Mathematics ability* at 8 years old. The CDA was developed by the International Association for the Evaluation of Educational Achievement (Cueto et al. 2009).

Middle childhood

We see this period as including children aged 6 years through 12. Schooling, which commences at 6 or 7 years old depending on the country, contributes significantly to competence in language and mathematics skills over time (Krutikova, Rolleston and Aurino 2014). Rolleston and James (2011: 3) note that:

> 'The development of young people's cognitive and non-cognitive skills occurs through a complex set of mechanisms among which school education is one. School plays a particularly important role in the development of 'basic' cognitive skills in the form of literacy and numeracy, and these skills form a foundation for the development of more complex cognitive skills such as information technology skills and problem-solving abilities.'

While it would have been optimal for this analysis to include measures of school effects from the Young Lives school surveys in the model, this was not possible as not all children who participated in the household surveys participated in those conducted in the schools. Restricting the analyses to a sample of matched household and school survey children would

1 For several reasons we did not separate analyses of receptive vocabulary on the basis of home language in any country. In some countries, taking home language differences into account would have resulted in very small groups, which would not have been suitable for this modelling exercise. The models reported here also do not examine differences between the various home languages in each country. By age 15 the range of languages in which the children preferred to answer the test had reduced significantly. To maintain consistency over the age points, the majority language of testing at 15 was used. For example, we observed that in Peru, while the test was taken in both Spanish and indigenous languages at earlier ages, by age of 15, 95 per cent took the test in Spanish, so children who responded in that language at every round were chosen for the analyses.

have substantially reduced power in most of the countries and compromised the representativeness of the samples. From Round 3 when they were 8 years old, Younger Cohort data do include measures of caregiver and child ratings of school quality and experience. Responses to these questions were explored but not included due to the high frequency of non-responses by caregivers and children.[2] We also doubt whether such questions provide valid measures of school quality.

Children from more affluent homes arrive in school with higher levels of literacy (including a greater range of vocabulary). They are also more proficient in tests of early reading in the early grades (Hecht et al. 2000; White 1982). At age 8, we assess *Early reading skills* using the Early Grade Reading Assessment (EGRA). The test assesses basic building blocks of reading acquisition, including emergent literacy, decoding and confirmation fluency (e.g. word, sound and letter recognition) understanding sentences and paragraphs, and listening with comprehension. EGRA was developed with the support of the World Bank and has been used in many developing countries (Gove and Wetterberg 2011). We expected performance on EGRA to predict growth in receptive vocabulary (*Vocab ability*) through middle childhood to 15 years old.

Longitudinal studies have shown that mathematics ability in Grade 1 is a strong predictor of later achievement in this ability (Duncan 2007) and at age 15 specifically (Watts et al. 2015). The latter study also finds that reading proficiency in Grade 1 is associated with the mathematics performance at 15 years old, and that children's mathematics self-concept and placement in advanced classes mediated their performance. They hypothesise that children who do well early on in mathematics are likely to develop a sense of competency in the area and that this influences their later learning and achievement.

Mathematics ability was measured on tests administered at 8, 12 and 15 years old on Young Lives mathematics tests (our variable is *Maths ability*). The tests contain items of increasing difficulty at each age point drawn from PISA and TIMMS tests (Cueto and Leon 2012) and are answered in the child's preferred language. Young Lives analyses have frequently extracted a set of common items across rounds to compare performance over rounds. We did not use this approach as we noticed ceiling effects on the common item set at ages 12 and 15. Instead, we used total raw scores transformed to standard normal deviates at each of the three age points. This has the effect of fixing the mean level of mathematics performance at zero for each age point, but since we modelled growth with random intercepts and random slopes, across participants, we were nevertheless able to model individual variability in performance over time points. Also, our predictor variables allowed us to model mean performance across groups conditionally, over time. Predictors of mathematics ability have been addressed above.

Young Lives provides an opportunity to examine the role of children's time use and its relationship to educational achievement during middle childhood (Watts et al. 2015), and research using the study data has shown that as poor and rural children grow older, the likelihood of their involvement in chores and work around the house increases. While studies report the pride taken by children in assisting their families, there are trade-offs. Studies have

2 Missing data constitutes a known threat to conclusions drawn in longitudinal studies. All our models used Full Information Maximum Likelihood (FIML), which is a recognised method for dealing with missing data analytically (Deng and Peng 2013). Although it does not constitute data imputation, FIML produces similar results to those obtained with imputation in many circumstances (Collins, Shafer, and Kam 2001). It is also important to note that the attrition rate in Young Lives was particularly low for the Younger Cohort (4.9 per cent, on average, across the countries, over the full time period).

reported tensions in demands on children's time for schooling and for family responsibilities (including paid work), with school attendance and performance being compromised when work demands on children are high (Pankhurst, Crivello and Tiumelissan 2016; Woldehanna and Gebremedhin 2015; Dawes et al. 2012; Orkin 2012). Unsurprisingly, the amount of time spent in school and learning at home has been found to influence Young Lives children's literacy and numeracy (e.g. Glewwe, Chen, and Katare 2015).

We constructed a predictor, *Hours for household chores and tasks* (see Table 1), to measure the impact of these demands on the child's study time (*Hours for school and study)*. The latter variable was computed from the child's report of hours spent in school and doing studies at home. In both cases, time use was averaged across 8, 12 and 15 years of age (Table 1), and used to predict on *Vocabulary* and *Mathematics growth*, and *Reading comprehension* at 15 years old.

Two plausible predictors of our outcomes that were excluded are parent and child educational and occupational aspirations. We excluded occupational aspirations as these were consistently limited to (probably unrealistic) occupational categories throughout the study rounds, and also because classification of occupational categories (e.g. high/low status) would have proved a challenge. Educational aspirations can be defined as the desire to achieve a certain level of education. Research in developed countries such as the United States shows that aspirations and expectations for the future are associated with socio-economic status, parent aspirations, peers and school achievements (Guerrero et al. 2016; Wigfield and Eccles 2000). The relationship between aspirations and outcomes such as school completion is stronger in adolescence as more realistic self-appraisals of academic abilities and future prospects emerge (Gottfredson 1981). Educational aspirations have been examined in Young Lives countries. For example, in Ethiopia, Favara (2017) found parent and child education aspirations in earlier phases of childhood to be related to grade attainment (controlling for cognitive skills and other likely influences). However, in the same country, Tafere (2017) observes that despite the odds against them, 78 per cent of 15-year-old Ethiopian boys and 70 per cent of girls still aspired to a university education. Parents' aspirations were similarly high. However, less than a quarter had completed school by age 19. In Peru, Guerrero et al. (2016) describe Young Lives findings of similarly high aspirations for parents and their children, as well as disappointing outcomes. In their study of aspirations in the Indian sample, Serneels and Dercon (2014) report an effect of maternal aspirations on children's grade achievement and performance on mathematics test scores at age 15 (controlling for a number of plausible influences). Interestingly, the effects were greater for children from more deprived backgrounds. Despite their evident role in predicting educational outcomes, aspirations were excluded from our analysis (a) because these were uniformly high and lacked variation (particularly prior to age 15), and (b) because of multicollinearity across rounds which would result (among other problems) in difficulties with interpretation of the data.

Adolescence

Growth in receptive vocabulary by 15 years was assessed using the measures described above and was used as a predictor of a test of reading comprehension (in each main language) at 15 years. The Young Lives reading comprehension test comprises sets of reading tasks that are designed to measure literacy at basic, intermediate and advanced levels. Items are drawn from the OECD Programme for International Student Assessment (PISA), the UNESCO Literacy Assessment and Monitoring Programme (LAMP), and the

Young Lives Peru school survey (personal communication J. Leon, 15 January 2018). We used total scores as IRT Rasch scores were not available for Round 5 data at the time of the study.

We included measures of psychosocial functioning at age 15 that are likely to influence school performance and hence the outcomes of interest. Aspects of psychosocial functioning are assessed at earlier points but we chose to observe their influence in mid adolescence, a sensitive period of changes in psychological well-being and sense of self (Marsh, Parada and Ayotte 2004). In all countries we included children's responses to Young Lives agency items as these had been used to assess self-efficacy in a number of relevant Young Lives' reports. These items have been used in different ways and for different purposes. For example, Dercon and Krishnan (2009) used three agency items and found a relationship between poverty and self-efficacy (agency), sense of inclusion and self-esteem in the Older Cohort at 12 years old. Dercon and Sanchez (2013) used the same five agency items used in the current study to explore relationships between children's growth status in middle childhood and psychosocial functioning at 12 years of age. Finally, Favara and Sanchez (2017) used the five agency items as a self-efficacy index to explore its relationship with risk behaviour. We also included the measure of self-esteem which had been used in Young Lives studies.

We found that the measure of self-efficacy had low internal consistency in our sample, once we investigated it empirically (coefficient omega was typically lower than 0.5, which is in line with what Favara and Sanchez (2017) found), and decided against using it. Although the measure of self-esteem had a slightly better level of internal consistency, it did not prove predictive of outcome variables in our models (Table 1b).

Favara and Sanchez (2017) note that several studies using Young Lives Peru data find negative relationships between risk behaviours and both psychosocial well-being and mathematics and language abilities. Not all the countries measured the same risk behaviours, and the most comprehensive set is available for Peru. Very few children in the Vietnamese sample endorsed these items. For Peru (see Table 1), we constructed two variables: (a) *Early crime and violence* and b) *Early substance use* (alcohol, smoking and illicit drugs). For Peru, we were also able to include a measure of *Mental wellness* using a reversal of the anxiety/depression subscale of the Strengths and Difficulties Questionnaire (Goodman 1997). We expected the two risk behaviour variables to be negatively associated with *Mathematics ability* and *Receptive vocabulary growth*. We also expected *Mental wellness* to be positively associated with both variables.

In sum, plausible predictors of growth in receptive vocabulary and in mathematics achievement by 15 years, and reading comprehension at the same point, have been included in models for each county. A longer list of variables was tested prior to finalisation. Modelling was limited to variables that showed sufficient variation, and to those with sufficient responses to warrant inclusion. Those used in final models are presented in Table 1a. Those considered but excluded following preliminary analyses are presented in Table 1b. As discussed, lack of indicators of school quality is a significant gap in our set of predictors and must await further study using the school survey samples.

Table 1a. *Variables included in the models for predicting reading comprehension at 15 years old, as well as for conditioning latent growth in mathematics skills between 8 and 15 years old*

Predictor variables
All countries:
Early childhood economic well-being (Rounds 1 and 2 averaged): Young Lives wealth index (Briones 2017). The index comprises consumer durables, access to services, and housing quality on a continuous scale of wealth. Higher values reflect higher household wealth.
All countries:
Maternal education (Round 1): total years of education (school and post-school).
All countries:
Hours for school and study (Rounds 3, 4 and 5): total hours per day the child reports as being spent at school plus time (in hours) reported as studying after school per day (averaged over Rounds 3, 4 and 5 per child).
Hours for household chores and tasks (Rounds 3, 4 and 5): total hours per day the child reports as being spent on (a) caring for others in the household, (b) doing household chores (e.g. cleaning, fetching water) and (c) household tasks (e.g. assisting on the family plot). Total hours for all three are averaged over Rounds 3, 4 and 5 per child.
All countries:
Preschool participation (Round 2 for all countries, except Ethiopia (in Round 3)): country-specific scaling by preschool provision type was used, with higher scores awarded to what is likely to be better provision based on other Young Lives findings cited above.
Peru: No preschool = 0; Public PRONOEI = 01; Public Centro Educativo Inicial (CEI) Público (Public Jardine) = 02; Private Centre-based CEI = 3.
Ethiopia: No preschool = 0; Public/community-based = 01; Private = 02.
Vietnam: No preschool = 0; Public/community-based = 01; Private = 02.
India: No preschool = 0; Public/community-based = 01; Private = 02.
All countries:
Maternal mental health (Round 1): Self-Reporting Questionnaire (SRQ20) (World Health Organisation 1994) (Round 1). A high score indicates risk of common mental disorder (anxiety and depression). In the model the scale was reversed for ease of interpretation. The SRQ20 has established reliability and validity at acceptable levels in many developing countries, including Vietnam, India, Peru, and Ethiopia (Bennett et al. 2016; Harpham et al. 2005; Tuan, Harpham, and Huong 2004; Harpham, Reichenheim, and Oser 2003).
All countries:
Early growth stunting (Round 1 at age 1 and Round 2 at age 5): each child is given a score based on their growth status at each age, as follows: Normal height for age =0; Stunted = 1; Severely stunted = 2. Scores at age 1 and age 5 were aggregated, thus the early growth stunting score had a range 0-4.[3]
Peru only:
Substance use and anti-social behaviour (Round 5): following exploratory factor analysis and consideration of question origin and content, two factors were constructed from Young Lives self-administered items at age 15. Factor analysis suggested that the substance use factor could be treated as unidimensional and had acceptable to good internal consistency. However, the anti-social behaviour scale was less clear in this respect and had marginally acceptable internal consistency.
Substance use internal consistency Omega = 0.72; Anti-social behaviour internal consistency Omega = 0.60.[4]
All countries:
Early quantitative skills (Round 3): Cognitive Development Assessment (CDA) - Quantity Concepts (15 items). For psychometry, see Cueto et al. (2009).
Peru only:
Emotional well-being (Round 5): five Young Lives ‘emotional stability’ items, from the strengths and difficulties questionnaire (Goodman 1997).
Internal consistency: based on the Older Cohort at 19 years, Yorke and Ogando Portela (2018) report Cronbach’s Alpha coefficients of 0.675 (Ethiopia), 0,663 (India) and 0.661 (Vietnam) for the measure. Peru (this sample): Omega = 0.74, Alpha = 0.74.

3 Stunting: HAZ > 2 standard deviations below the population median; Severe stunting: HAZ > 3 standard deviations below the population median.

4 Items are listed in Appendix B.

Chronologically latest outcome variable
All countries:
Reading comprehension (Round 5): Reading comprehension test raw score at 15 years old. The test has five sets of items that assess basic to advanced literacy. Items are drawn from publicly available international tests. The number of items varies by country. Psychometric properties were established using both classical test theory (CTT) and item response theory (IRT) for Round 4. Round 5 tests were based on piloted items in each country to establish random responses (guessing), item difficulty, and item discrimination. Items with sound properties were selected. Psychometry on Round 5 had not been completed at the time of this analysis.

Latent growth variables
Latent growth in receptive vocabulary:
Measured at ages 5 (Round 2), 8 (Round 3), 12 (Round 4) and 15 (Round 5),using the child's total score in the same language at each of these age points. Minor languages were excluded as explained above.
Peru: Peabody Picture Vocabulary Test (PPVT-R) Castellano and translated into Quechua.
India, Vietnam and Ethiopia: receptive vocabulary was measured using translated versions of PPVT111 that were adapted to ensure correct order of item difficulty (and shortened) following IRT analyses in Round 3.
At the time of this analysis, psychometric analyses and Rasch scores were not available for Round 5 (for earlier rounds see Leon and Singh 2017; Cueto et al. 2009; Cueto and Leon 2012). Psychometry for Rounds 4 and 5 was not available at the time of this study. Due to adaptation, the test is not strictly equivalent to the PPVT.
Raw score totals were used in analyses. Languages of administration were as follows: India – Telugu and English; Vietnam –Vietnamese; Ethiopia – Oromifa, Tigrinya, Amharic.
Latent growth in mathematics ability:
Mathematics ability was measured at 8 (Round 3), 12 (Round 4) and 15 years (Round 5). Tests increase in difficulty at each age point. As the tests were not equivalent across time points, total scores at each point were transformed to standard normal deviates (z-scores) for comparative purposes. For Young Lives psychometric analyses, see: Cueto et al. 2009; Cueto and Leon 2012. Psychometry for Rounds 4 and 5 was not available as the time of this study.

Notes: The latent growth curve in each case was constituted by a latent variable representing the intercept, and a latent variable representing slope/change. Both were modelled as varying randomly across individual children.

Because we attempt to equate the variables and compare their relative importance, we report only standardised coefficients, and hence these are not interpretable in terms of the unscaled variables described in this table.

Child age was controlled in both models by entering it as an exogenous variable with a directional effect on the latent variable representing the intercept component of the latent growth curve. We do not show child age in any of the model diagrams, as it is a control rather than substantive element.

Table 1b lists variables that were proposed for inclusion but were excluded following preliminary analyses.

Table 1b. *Variables excluded from the models*

Variable	Reason for exclusion
Shocks (total) Rounds 1-4)	Usually highly correlated with the wealth index, and low incidence across the samples.
Social protection programmes (Round 2) Peru: Juntos; Vietnam: Molisa; Ethiopia: N/A (too few beneficiaries); India: NREGS public works programme.	Confounded with the wealth index and had no relationship to the other variables in the theoretical models.
Child ethnicity	Would have added considerable complexity to the models, and in most cases would have resulted in low power cells or regions in the model. Better suited to specific follow-up analyses rather than global models.
Caregiver dependency burden: number of children cared for at Round 1.	Restricted range (small number of children for most mothers/caregivers).
School type (private\not) (Round 3)	Variation across the countries in terms of local practices dissuaded us from using the variable, for example, it was not available for Vietnam in Round 3 and we wished the models to be as consistent as possible.
Adolescent self-esteem (Round 5)	Eliminated during preliminary model fitting; no predictive value.
Adolescent self-efficacy (Round 5): five Young Lives agency items were used.	Eliminated during preliminary model fitting; no predictive value. A further issue was poor and unacceptable internal consistency (Omega and Alpha coefficients) for all the countries. Our finding is similar to that reported by Yorke and Ogando Portela (2018), and Dercon and Krishnan (2009). In contrast, Dercon and Sanchez (2013) report Alpha = 0.89. on items pooled across the countries at 12 years old.
Encouragement of child's reading (Round 4 Peru only)	Inter-correlated with hours at school and hours of studying after school.
Caregiver view on school quality (Rounds 3 and 4) (teachers miss class; class unattended; informed of progress; quality of school)	Too few records, thus reducing total sample available for modelling too much.

1.2. Modelling influences on growth in vocabulary and mathematics ability, and on reading comprehension at 15 years old

Latent Growth Modelling (LGM) has become an important method for the investigation of developmental change and for understanding the variables on which this change is conditional (Singer and Willett 2003). In the present analysis, LGM was used to model change in receptive vocabulary and mathematics test performance over ages 5 to 15 years and 8 to 15, respectively. LGM allowed us to model individual differences in outcomes at 5 years (receptive vocabulary) and 8 years (mathematics), differences in their trajectories over time to 15 years of age, and to explore whether these trajectories are conditional on other variables. LGMs for each country were informed by the theoretical model of change described above, and shown in Figures 1 and 2. Temporal orders of variables likely to influence trajectories during early (0-5 years old) and middle childhood (6-12 years old) and early adolescence (13-15 years old) were constructed, based on the evidence presented in the preceding section, and are also shown in Figures 1 and 2.

Initially, we considered testing fully latent conditional LGMs (i.e. with latent variables representing not only the outcome growth variable, but also the predictors), for improved consideration of measurement error. However, this proved impracticable. The models were very complex, with dozens of manifest variables, and examining latent variable structures of predictors for ill-fit proved very time-consuming, with increased risk of indeterminacy since many of the predictors we derived were not designed as latent variables in the first place, and thus could not be expected to have a clear factor structure. Hence, all our predictors were manifest variables.

We accessed Young Lives datasets for our analyses through the ukdataservice.ac.uk portal (http://dx.doi.org/10.5255/UKDA-SN-7483-2), and directly from Young Lives, particularly to obtain some data that were not in the public domain (e.g. maternal mental well-being) and data not yet archived (from Round 5, in particular). All data access was officially sanctioned by Young Lives. Data dictionaries and questionnaires were examined, and a preliminary set of outcome variables was identified for potential secondary analysis. Peru had the most comprehensive set.

The key longitudinal measures in models for all countries were change in receptive vocabulary over time (four measurements from age 5) and change in quantitative ability (maths at ages 8, 12 and 15). Each of the models was constructed with slightly different considerations in mind, and so are discussed separately.

To fit the statistical models, we used the package Lavaan (Rosseel 2012) within the R Package for Statistical Computing (R Core Team 2017). Models were fit with full information maximum likelihood estimation. We assessed overall fit of the model with standard SEM measures of fit – χ^2 against the saturated model, RMSEA (Root Mean Square Error of Approximation), SRMR (Standardised Root Mean Square Residual), and CFI (Comparative Fit Index), but report only the latter three, since the χ^2 tests are known to be proportional to sample size, and our sample size was large, thus inflating p-values. These are shown for each of the figures documenting the results of the LGM. The three measures of fit are widely accepted as being among the most appropriate for model evaluation (e.g. Hu and Bentler 1999; Hooper, Coughlan and Mullen 2008). There is some disagreement about the conventional values of relative fit for each measure, but based on our reading of the literature (e.g. Hu and Bentler 1999; Kenny 2018), a reasonable compromise for present circumstances seems to be as follows: for RMSEA, .05 and lower indicates very good fit, between .05 and .08 indicates moderate fit, and above .1 indicates poor fit; for SRMS values, less than .08 indicate reasonable fit, and values less than .05 very good fit; and for CFI values, above .90 indicate moderate fit, and values above .95 very good fit.

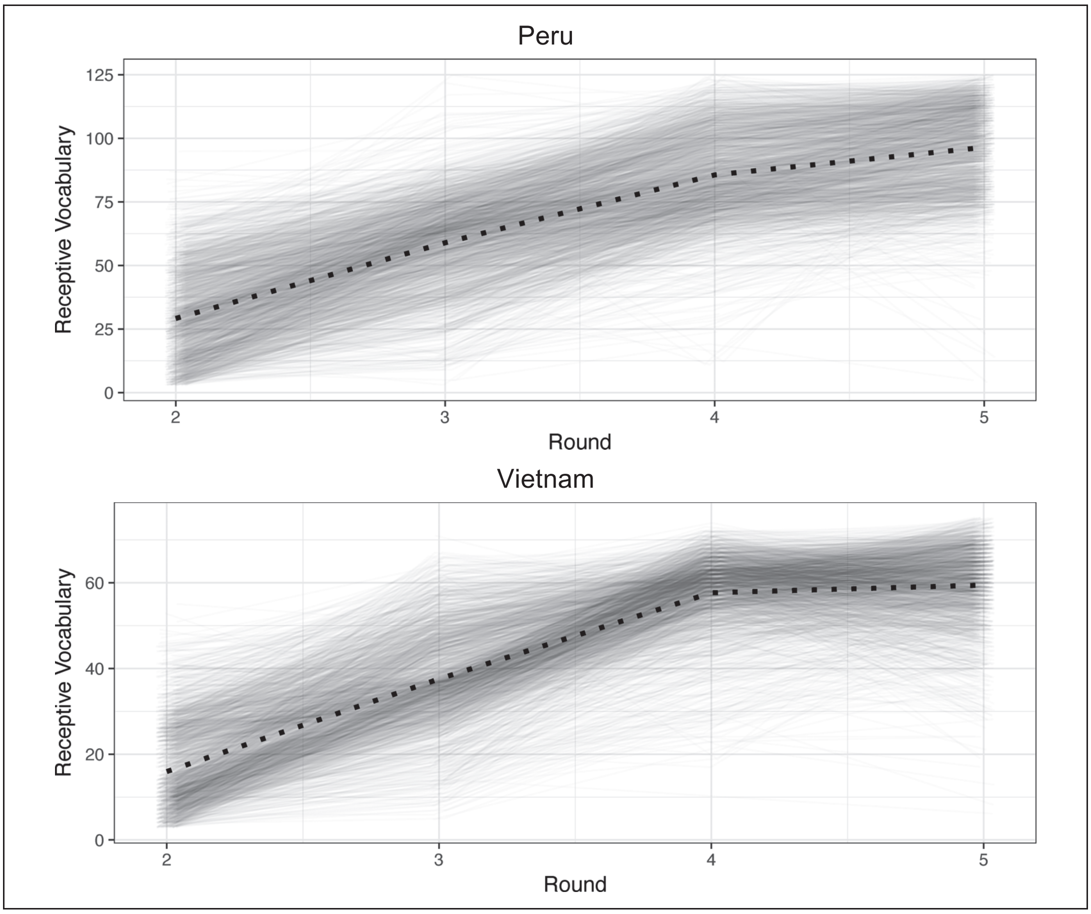
It is important to note that the large sample size in each country means that many of the statistical significance tests for individual paths in the vocabulary and mathematics models are highly powered. In the empirical findings for each country, we present all statistically significant paths with lines of thickness proportionate to the standardised coefficients (i.e. thicker lines indicate larger effect size) and we colour lines black and red to indicate positive and negative relationships, respectively. At the end of the presentation of findings for each of vocabulary and reading, and mathematics, we present a summary diagram that attempts to reflect common patterns for all four countries, showing paths that were consistently statistically significant across countries, and paths that were often but not always significant. The ability to apply the models to children in four very different countries speaking a variety of languages and enrolled in different education systems provides an unusual opportunity for validation and testing of models.

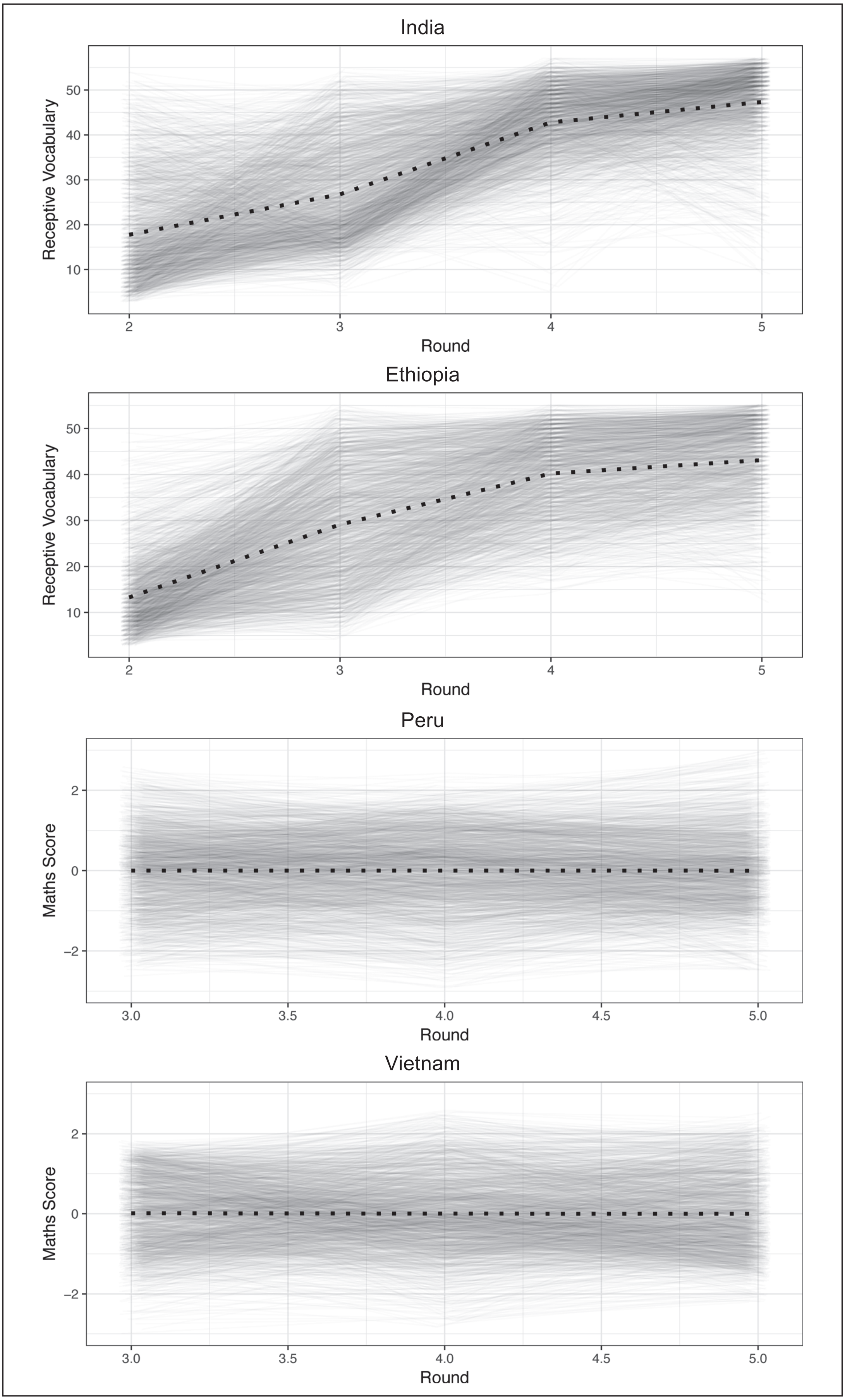
In the interest of brevity, we do not describe all the steps involved in modelling and rather present summaries of model fit and findings. To some extent, decisions were on occasion ad hoc, and in danger of capitalising on chance, but this is an issue we address elsewhere in this report, specifically pointing out that the opportunity to test models across multiple countries probably gave us some licence in this respect.

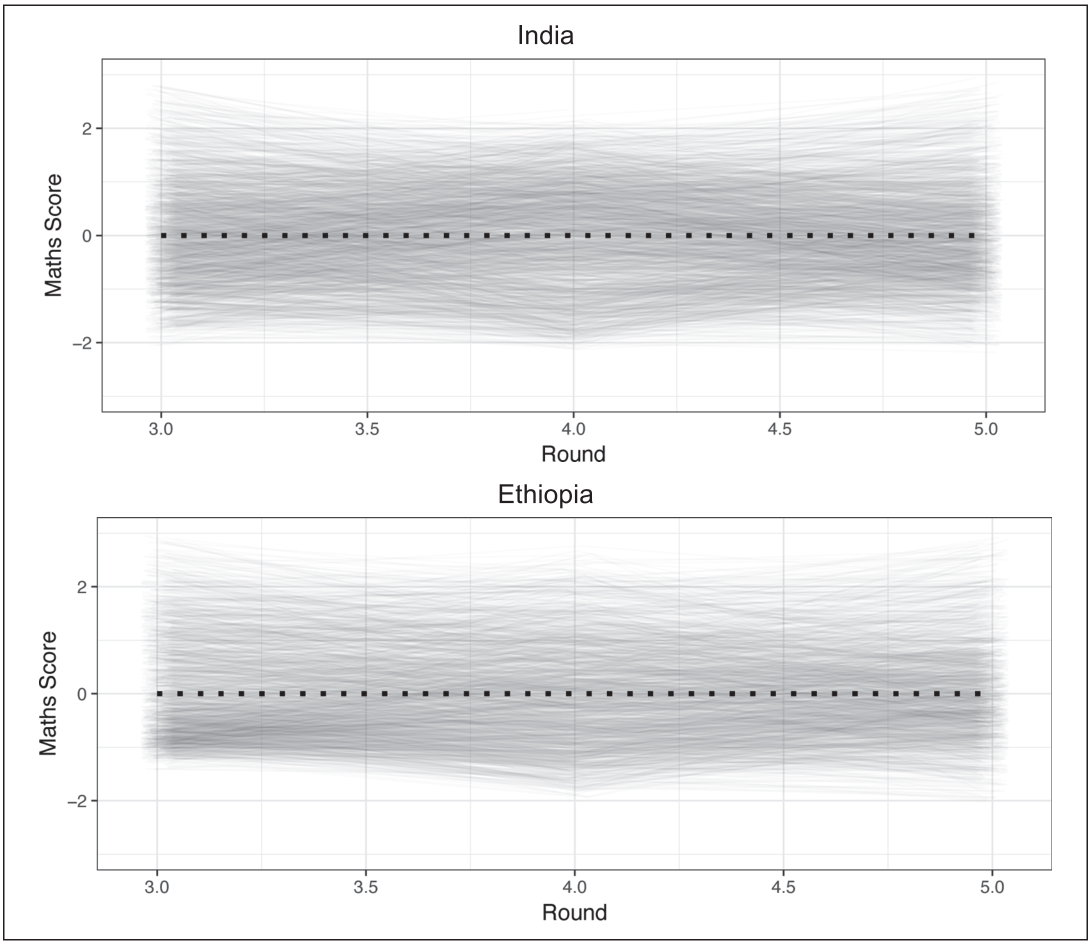
2. Modelling influences on receptive vocabulary and reading comprehension

In all the countries, we first modelled change in receptive vocabulary unconditionally, and after concluding that this could be done reasonably successfully with a linear model of change, moved on to consider the conditional models of change shown in Figures 1 and 2 (the number of time intervals at which outcome variables were measured were in addition too few on which to reliably assess alternative non-linear fit). As indicated earlier, our models of change declared random intercepts and slopes across individual children. Figures 3a to 3h, which are person-period plots of receptive vocabulary and mathematics, across the four countries, show the need for random intercepts in the modelling process, with some indication too of a need for random slopes: there is clearly considerable variation between individuals, even though there is also clearly group-level growth over time in receptive vocabulary. Note the flattening slope from age 12 to 15 for receptive vocabulary; this apparent non-linearity is partly a function of the shorter time period between Rounds 4 and 5.

Figures 3a to 3h. *Receptive vocabulary and mathematics scores, over time, per child, in four countries*




India
Receptive Vocabulary
10
20
30
40
50
2
3
4
5
Round
Ethiopia
Receptive Vocabulary
10
20
30
40
50
2
3
4
5
Round
Peru
Maths Score
−2
0
2
3.0
3.5
4.0
4.5
5.0
Round
Vietnam
Maths Score
−2
0
2
3.0
3.5
4.0
4.5
5.0
Round



Notes: (dashed line = mean; light solid lines = each child). The horizontal axes are not linear estimations of time (but round of study instead), and the apparent non-linearities of growth in receptive vocabulary are thus exaggerated, and ultimately small departures from linearity.

We then added additional covariates in conditional LGMs. These were entered according to a chronology of likely early impact. We entered maternal education as chronologically prior in the network of paths, followed by our indicator of early childhood economic well-being using the Young Lives wealth index aggregated over the first two rounds of the study (12 months through 5 years of age). We considered adding wealth index for later rounds as a time varying covariate to the LGMs, but the correlations across rounds were very high, and introduced too much collinearity into the model. This indicated that there was little variation in changes in wealth over time.

Maternal mental health when the child was an infant (Round 1) was entered as a predictor subsequent to the wealth index, as was child growth stunting. Preschool participation (age 3 to 5 years depending on the country) was scaled as noted in Table 1a and entered at Round 2. Variables measured at Rounds 1 and 2 were allowed to affect both the intercept, and the rate of change in the LGM. Predictors entered at subsequent rounds of measurement were only allowed to affect the rate of change, since they were measured after the round at which the intercept was determined. In middle childhood, these included our measures of early reading ability (the EGRA), opportunity hours for learning, type of school attended (public versus private), and whether the caregiver encouraged reading (Peru only).

In adolescence (Round 5 at 15 years old) the child’s sense of self-efficacy was entered as a predictor of growth in receptive vocabulary in all countries. For Peru only, measures of emotional well-being and involvement in risk behaviour were added. These were either not measured in the other countries or the response rates were too low. Although receptive vocabulary was the main outcome measure at Round 5 in the language LGM, reading comprehension was included as a single outcome (i.e. not longitudinal).

We should note that throughout the modelling exercise we started with the theoretical models, and attempted to fit these in the first instance, observing the various time-dependent relations, and conjectured regression relations. We then pruned non-significant paths, added auto-regressive structures for the vocabulary models, and on occasion allowed the error variances of some variables to correlate with each other (this was typically the case for the vocabulary models), in order to reduce large model residuals. This procedure might be considered a little ad hoc, and in danger of capitalising on chance, but we tested very similar models across four different countries, and that the models replicate fairly well across the countries is some control against the dangers of building models with significance tests. We also end our analysis by choosing single vocabulary and maths models that replicate well across all the countries (see Figures 16 and 17), which is an additional, if stringent, control (since there may well be real variation across countries, and requiring high levels of replication might err, effectively making a Type II error in the desire to control Type I errors).[5]

Final model fits for each country were assessed using standard measures in the SEM literature (RMSEA, sRMR, and CFI). Our fits were all in the moderately well-fitting range, for the most part, according to the criteria set down by Hu and Bentler (1999), but we should note that these measures are known to work less well in the case of LGMs specifically (Preacher et al. 2008). Significance tests on individual coefficient sizes (against null hypotheses of size = 0) in SEM and LGM are known to depend on the multivariate normal distribution of the predictor and outcome variables. There were high individual residual scores for several of the covariances we tried to fit, for all models we report here, casting some doubt on the model estimates and significance tests. However, we did have large sample sizes for each model (never lower than 1,400 cases), which provides some protection (the effect of distributional problems is much less severe with large sample sizes), and we also computed empirical bootstrap confidence intervals for each of the coefficients we estimated in the models, as an additional check.

Next, we present models and findings for receptive vocabulary growth and reading comprehension at 15 for each country in turn.

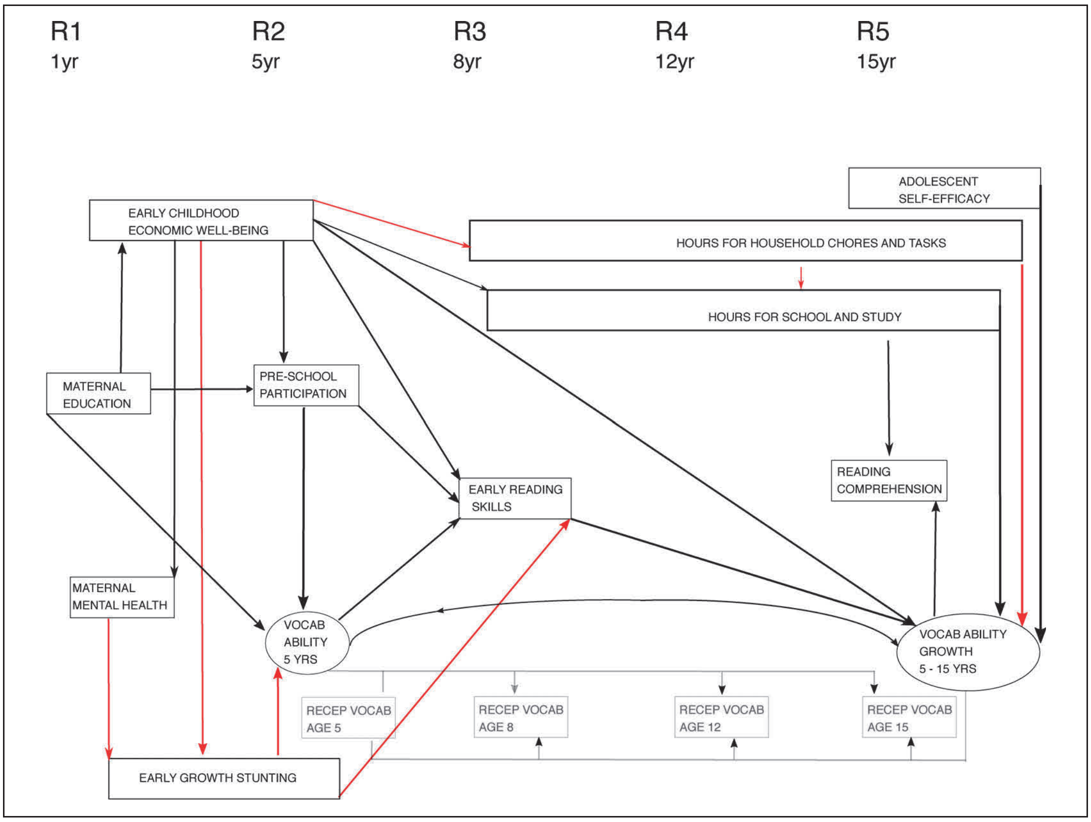
5 A Type I error occurs when the null hypothesis is falsely rejected, that is, statistical analysis leads us to conclude that there is a difference between two groups in an experiment when the opposite is true (there is no difference); a Type II error occurs when the null hypothesis is accepted, but should have been rejected (there is a difference).

3. Findings 1: Receptive vocabulary growth

3.1. Receptive vocabulary growth in India, Ethiopia and Vietnam

Figure 4 presents the temporal design for India, Ethiopia and Vietnam. Thereafter, we present findings for each of the countries. Peru is presented separately in Section 3.2.

Figure 4. *General theoretical model of conditional development of receptive vocabulary: India, Ethiopia and Vietnam*



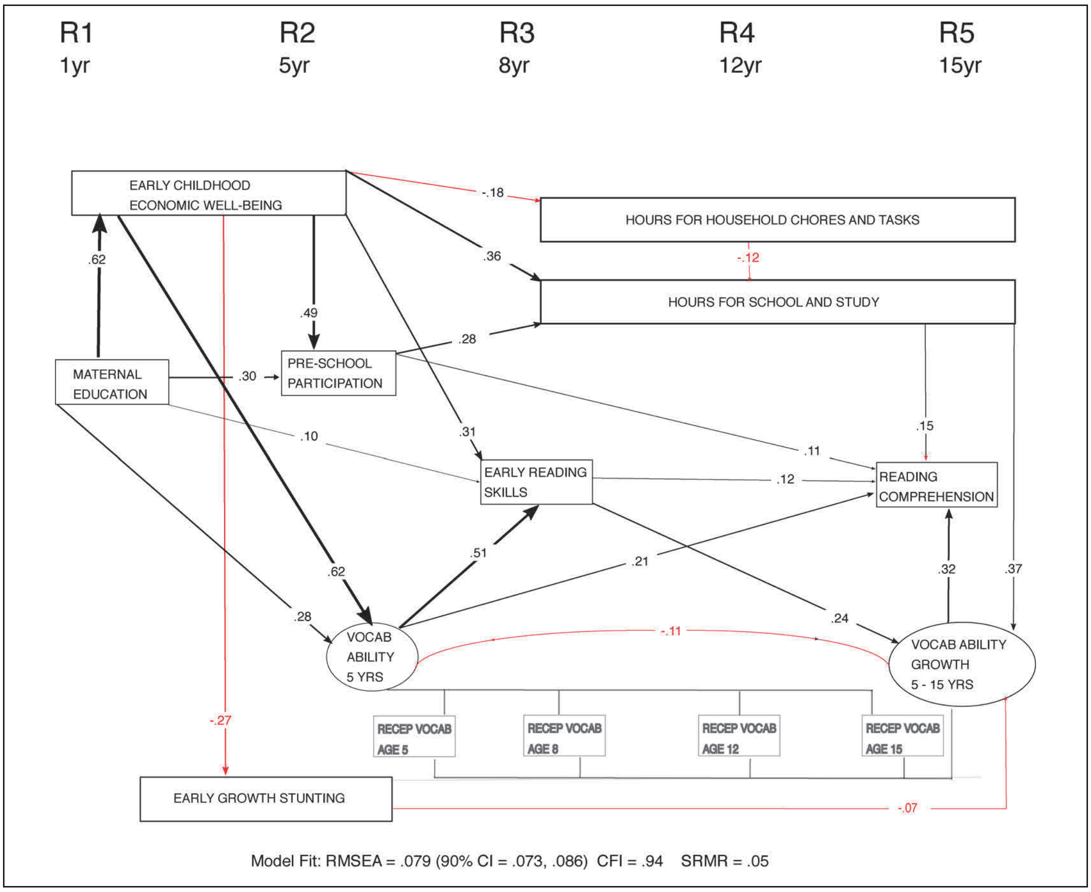
In the model, black lines represent predicted positive directional effects, and red lines negative directional effects. Vocabulary is a Latent Growth Curve (with latent intercept and slope variables, and their covariance [the double-headed arrow]), constituted by tests administered at four time points (ages). Predictor variables are hypothesised to have a particular chronological precedence, as shown in the figure. Ellipsoid shapes represent latent variables, and rectangular shapes represent manifest variables.

In all the figures that follow, the same convention for colour of lines is followed (black positive; red negative) is used. Line thickness reflects strength of relationship. Arrows indicate direction. Ellipsoid shapes represent latent variables, and rectangular shapes represent manifest variables.

3.1.1. Ethiopia

Figure 5 shows the fitted model for the Ethiopian sample.

Figure 5. *Empirical model of conditional development of receptive vocabulary in Ethiopia*



Notes: All coefficients are statistically significant (at a maximum $p < .05$, except for the covariance between latent slope and intercept, which is not significant, but shown here as it is a model component), and standardised. N = 1421. Further model details, including bootstrapped confidence intervals, are in Appendix A1.

The overall level of model fit, as assessed with the three chosen measures, is moderately good (but note, as explained earlier, there is disagreement on what degree of model fit one should consider as being acceptable for LGM models); CFI and SRMR are both just below levels considered very good fit (.94 and .053 respectively), whereas RMSEA (.079) is just below a level considered moderately good. Examining the path coefficients, one can see a ‘mesh’ of strong effects emanating from maternal education, and early childhood economic well-being, in particular. Vocabulary ability at 5 years (the intercept term in the latent variable) is strongly related to economic well-being, and to maternal education (inspection of the bootstrapped confidence intervals in Appendix A will show that they are non-overlapping, suggesting that economic well-being is more potent). Preschool participation appears to mediate the effect of economic well-being and maternal education on reading comprehension at age 15, but does not itself appear to lead to improved receptive vocabulary at age 5, or on the degree of change in receptive vocabulary between 5 and 15 years old.[6] Early reading skills, on the other hand, do appear to serve this function: that is, both economic well-being and maternal education have a positive relation with early reading, and better early reading skills are associated with greater growth in receptive vocabulary and better reading comprehension at age 15. However, early reading skills are also related to level of receptive vocabulary at age 5, so the path is complex. An additional early stage predictor, growth stunting at ages 1 and 5, is weakly and negatively related to growth in receptive vocabulary, suggesting greater levels of stunting lead to a lower rate of increase.

Interesting later stage predictors concern the time spent by children getting involved in household chores and tasks, and in studying at school and at home. These variables are related in an interesting set of paths to economic well-being, preschool participation and quality, and to each other. Children who are less well-off economically spend more time doing household chores and tasks, and the more they do this, the less time they appear to spend studying. Tracing the path further, children who spend more time studying show greater growth in receptive vocabulary, and better reading comprehension at age 15.

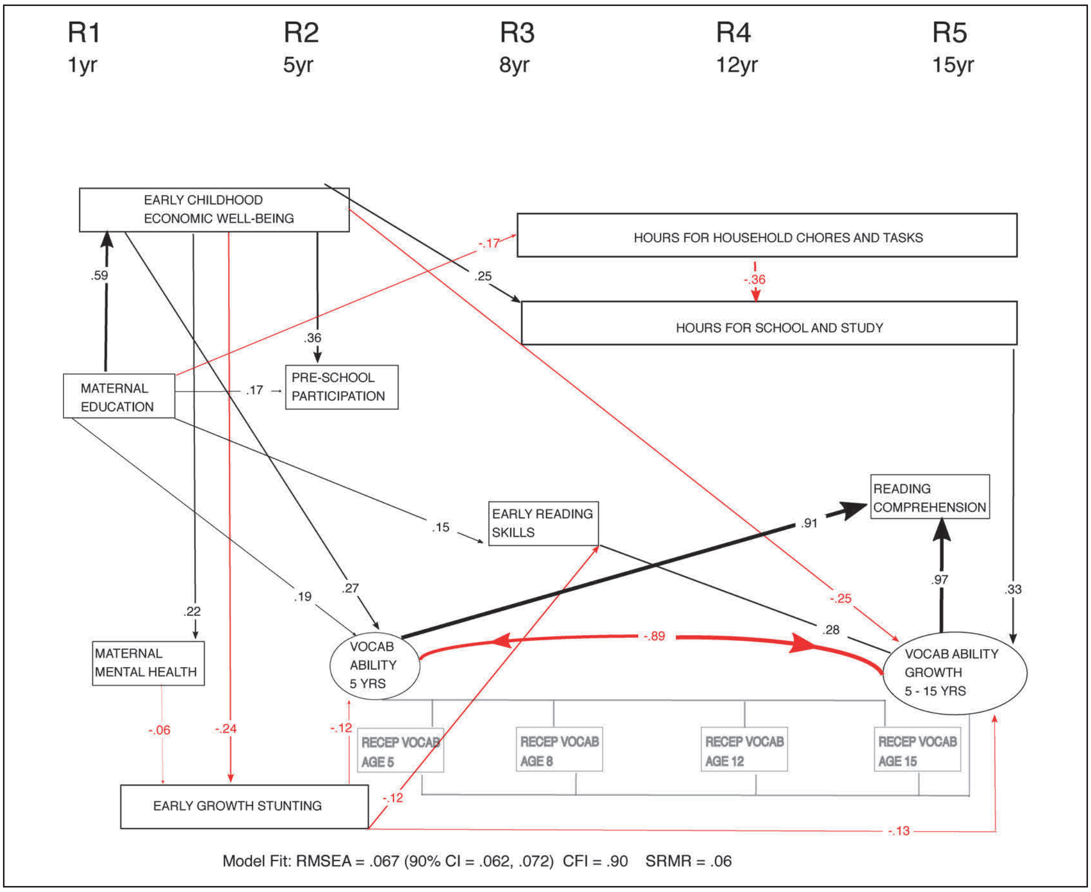
Despite the complexity of the model, the overall picture appears clear: economic well-being and maternal education are the starting points in a mesh of paths that lead through preschool participation to higher levels of receptive vocabulary early on, and to better early reading skills, and finally to greater growth in receptive vocabulary, and better reading comprehension at age 15. A separate set of paths leads from greater economic well-being to less time spent on household chores and tasks, and more time spent studying, and through there, to greater growth in receptive vocabulary and better reading comprehension at age 15. Early growth stunting, an important phenomenon in poor and developing countries, is related in the expected direction to economic well-being, and does seem to be related (albeit weakly) to growth in receptive vocabulary.

3.1.2. *India*

Figure 6 shows the fitted model for the Indian sample.

6 Note that we do not test formally for mediation or indirect effects in our modelling exercises, given the very many and complex set of indirect effects in our models.

Figure 6. *Empirical model of conditional development of receptive vocabulary in India*



Notes: All coefficients are statistically significant (at a maximum p < .05), and standardised. N = 1980 Further model details, including bootstrapped confidence intervals, are in Appendix A2.

The overall level of model fit, as assessed with the three chosen measures, is moderately good. It shares many of the features of the model for Ethiopia, including a dense mesh of paths emanating from maternal education and economic well-being, through receptive vocabulary at age 5, early reading skills, and on to growth in receptive vocabulary and reading comprehension. It also shares the set of paths that proceed from economic well-being through hours spent completing household chores, and hours spent studying, through to growth in receptive vocabulary. The picture in this respect is similar to that in Ethiopia: economic well-being and maternal education are the starting points in a mesh of paths that lead to higher levels of receptive vocabulary early on, and to better early reading skills, and finally to greater growth in receptive vocabulary and better reading comprehension at age 15. A separate set of paths leads from greater economic well-being to less time spent on household chores and tasks, and more time spent studying, and through there, to greater growth in receptive vocabulary and better reading comprehension at age 15.

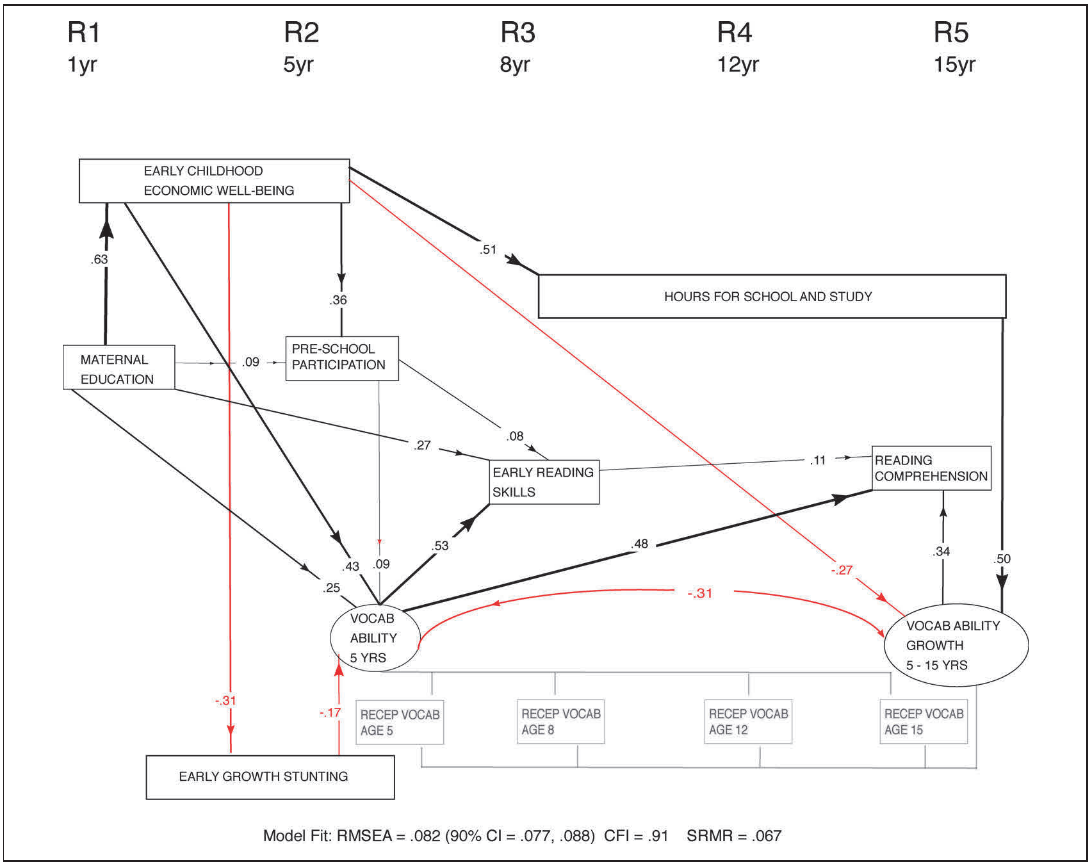
There are though some important differences between the models. Whereas participation in what is likely to be better quality of preschool education appeared to mediate some effects of economic well-being and maternal education, it does not do so in India. Whereas early growth stunting at ages 1 and 5 was weakly related in Ethiopia to outcomes, it seems more thoroughly enmeshed in mediators and outcomes in India: that is, children who are more stunted have lower receptive vocabulary at age 5, lower reading skill level at age 8, and show less growth in receptive vocabulary. Note that growth stunting itself is related to lower economic well-being, and to poorer maternal mental-health.

There are two additional complexities with the Indian model, both of which may reflect catch-up phenomena: that is, early economic well-being is negatively related to growth in receptive vocabulary, suggesting that children who are well off early on improve less than children less well-off. This is not surprising if one considers that well-off children have higher levels of receptive vocabulary at age 5, and might just represent catch-up (or perhaps a 'regression to the mean' phenomenon). This possible catch-up phenomenon is also suggested by the strong, negative correlation between the slope and intercept terms in the latent variable.

3.1.3. Vietnam

Figure 7 shows the fitted model for the Vietnamese sample.

Figure 7. *Empirical model of conditional development of receptive vocabulary in Vietnam*



Notes: All coefficients are statistically significant (at a maximum p < .05), and standardised. N = 1952. Further model details, including bootstrapped confidence intervals, are in Appendix A3.

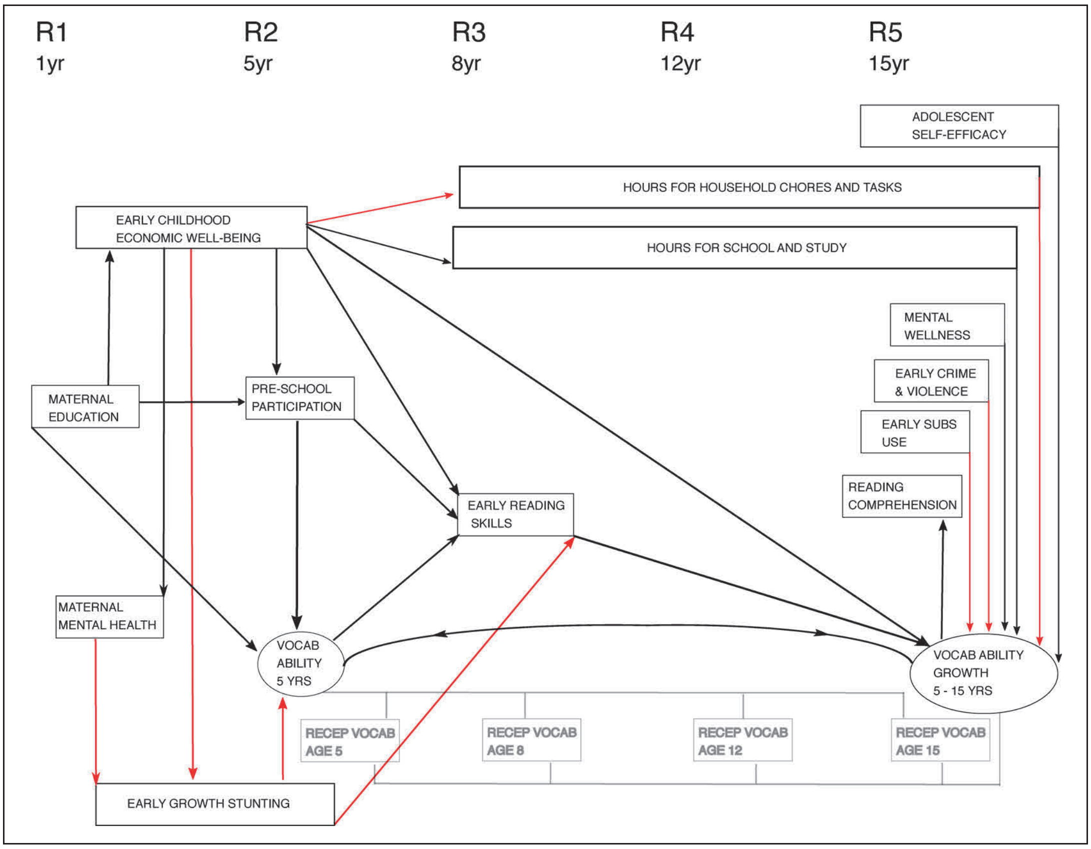
The overall level of model fit, as assessed with the three chosen measures, is at the upper end of what would conventionally be considered moderate fit (RMSEA is slightly higher, at .082; CFI is at the lower end of moderate fit, at .91; and SRMR is at a better level than considered moderate fit (at .067). It shares many of the features of the models for Ethiopia and India, including a dense mesh of paths emanating from maternal education and economic well-being, through receptive vocabulary at age 5, early reading skills, and on to reading comprehension. The pathways to growth in receptive vocabulary are somewhat different from Ethiopia, though, but share the negative relations between early economic well-being and growth in receptive vocabulary, as well as the strong negative relation between early receptive vocabulary and growth – this is likely to be the same catch-up phenomenon as in the Indian sample. The path to growth in receptive vocabulary from hours spent in school and at home, studying, is similar, but less densely connected to economic well-being. Participation in, and quality of preschool activity is weakly but significantly connected in several pathways, leading to early receptive vocabulary level, to early reading skills, to growth in receptive vocabulary, and to reading comprehension at age 15.

Despite these specific differences, the model for Vietnam is clearly very similar to those for India and Ethiopia.

3.2. Receptive vocabulary growth in Peru

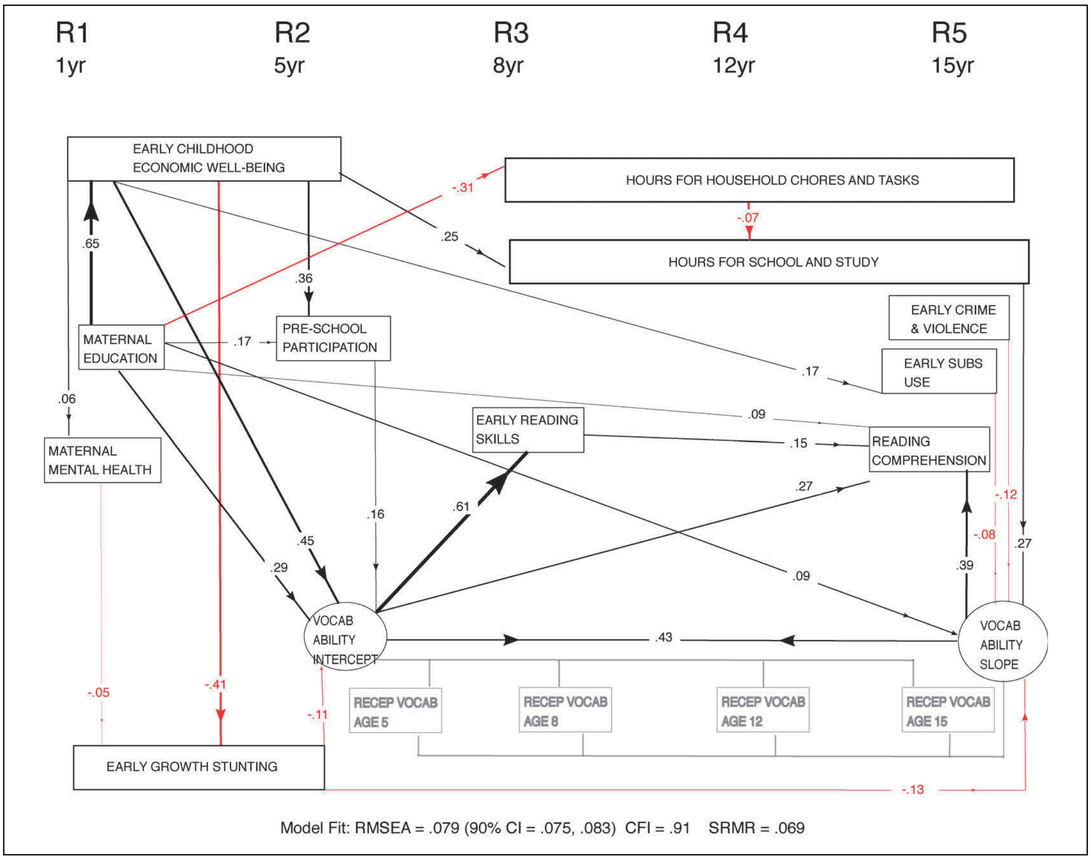
Peru is presented separately as it has certain variables that either were not available in the other countries or had too few records: mental wellness, early crime and violence, and early substance use. Figure 8 presents the temporal design for predicting growth in receptive vocabulary and reading comprehension skills at 15 years old for Peru.

Figure 8. *Theoretical model of conditional development of receptive vocabulary in Peru*



The fitted model for the Peruvian sample is shown in Figure 9.

Figure 9. *Empirical model of conditional development of receptive vocabulary in Peru*



Notes: All coefficients are statistically significant (at a maximum p < .05), and standardised. N = 1938. Further model details, including bootstrapped confidence intervals, are in Appendix A4.

The overall level of model fit, as assessed with the three chosen measures, is moderately good: sRMR = .067, which is lower than the value of .08 we justified earlier; RMSEA = 073, which is between the conventionally accepted levels of good (.05) and mediocre (.08) fit, respectively; and CFI = .90, which is at the lower level of conventionally acceptable fit. It shares many of the features of the models for Ethiopia and India, especially: (i) the mesh of paths emanating from maternal education and early economic well-being; (ii) the indirect effect of the former on receptive vocabulary at age 5, growth in receptive vocabulary, and reading comprehension, through preschool quality and participation, and early reading skills; and (iii) the indirect effect of early economic well-being on hours spent doing household chores and tasks, on hours spent studying or at school, and on growth in receptive vocabulary. The enmeshment of early growth stunting through early economic well-being, and the effects of this on early receptive vocabulary, and growth in receptive vocabulary, are also evident, as shown in Figure 10. There are two additional paths of interest in the Peru model: first, there is a pathway between economic well-being, maternal mental health, and early growth stunting, which we also observed in the data for India; and second, there is a negative path between early crime and violence (by age 15), and growth in receptive vocabulary, and this path does not have any significant relations with variables preceding it chronologically.

As in the models for the other three countries, the picture that emerges is of the central importance of early economic well-being and maternal education, and the possible mediation of their effects on receptive vocabulary levels and growth in receptive vocabulary, through preschool quality and participation, the development of early reading skills, lower levels of household chores and tasks, and more time spent in school and studying at home. Early growth stunting has significant, if small, negative effects in this complex chain, as do maternal mental wellness at a very early age of the Younger Cohort child, and adolescent activity of the Younger Cohort child in crime and violence.

4. Findings 2: Growth in mathematics abilities

The LGM that we constructed for change in mathematics performance over three rounds followed the same principles for the most part as that outlined for vocabulary and reading comprehension. For all countries, we undertook a preliminary modelling of change in mathematics performance, to test whether a linear model would suffice (using age per round as the coefficient set, rather than round number), and this proved to be the case. There was considerable variation in individual trajectories over time, though. As for vocabulary, we made change in mathematics performance the longitudinal element of our LGM (commencing when the child was 8 years old (Round 3) and built a model of intercept and rate of change, per individual child, extending from Round 1. As explained earlier, receptive vocabulary was measured in such a way over rounds to permit ‘real’ growth in receptive vocabulary, taking common items per round, but this was not feasible for the test of mathematics used per round. Instead, we transformed mathematics tests scores to standard normal scores at each round. This had the disadvantage of fixing mean performance per round, but it did allow us to examine individual variability of children at each round, and in individual growth across rounds. We should also note that although we used an auto-regressive structure in the receptive vocabulary models for taking into account the collinearity between the mathematics scores at Rounds 2 to 5, we were not able to do this for any of the mathematics models: since the latent growth function was based on three measurements of mathematics performance, there were too few degrees of freedom available to impose the additional constraints.[7]

Early round predictors of both starting level of performance, and change in mathematics performance over time, were as discussed for receptive vocabulary but with the addition of *Early quantitative skills*, which is the child’s performance on the quantities subscale of the CDA at 5 years of age. Vocabulary and reading were not included in the mathematics models. As for the previous modelling exercises, the large sample sizes in each country meant that many of the statistical significance tests for individual paths were highly powered. As before, we present all statistically significant paths for each country, and then present a summary diagram that captures the common pattern for all four countries.

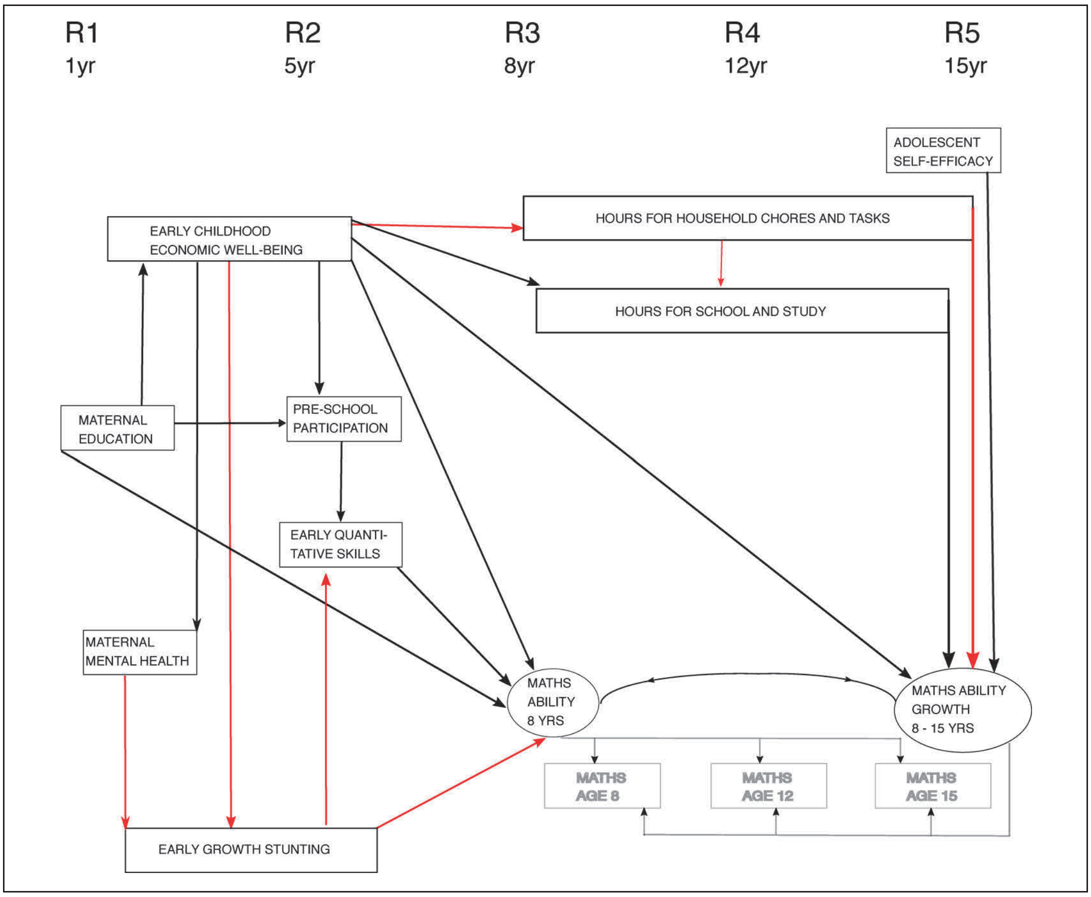
7 We should note that auto-regressive structures are not typically taken into account in Latent Growth Models. If there are substantial auto-correlations between the outcome variables across time periods, it is possible that the variance of the slope coefficients could be inflated, in the same way that multicollinearity might inflate standard errors of regression coefficients.

4.1. Growth in mathematics abilities: India, Ethiopia, and Vietnam

First, we present a temporal theoretical model for India, Ethiopia and Vietnam in Figure 10 (which also serves as the base for the model in Peru) Thereafter we present findings for each of the countries in turn. Peru is presented separately in Section 4.2. as it includes mental wellness, early crime and violence, and early substance use, none of which were included in the other models.

As for vocabulary growth models, in all figures that follow, black lines represent positive directional effects, red lines negative directional effects. Arrows indicate direction. Rectangular and ellipsoid shapes represent manifest and latent variables, respectively.

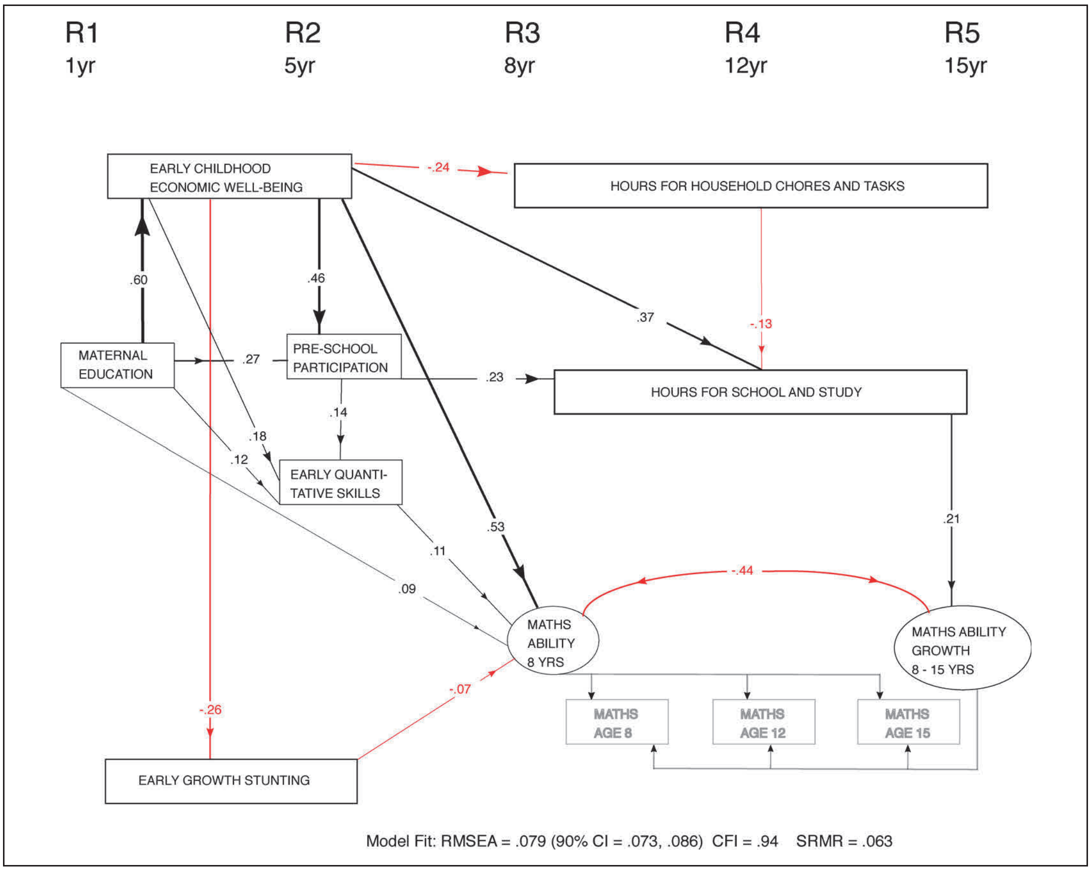
Figure 10. *Theoretical model of conditional development of mathematics in Ethiopia, India, and Vietnam*



4.1.1. Ethiopia

The fitted model for the Ethiopian sample is shown in Figure 11. The overall level of model fit, as assessed with the three chosen measures, is moderately good: sRMR, at .063, is better fitting than the conventionally acceptable level (.08); CFI = 0.94 is close to the .95 level, commonly accepted as well-fitting; RMSEA at .079 is just below a level considered a moderate degree of fit.

Figure 11. *Empirical model of conditional development of mathematics in Ethiopia*



Notes: All coefficients are statistically significant (at a maximum p = .05), and standardised. N = 1951. Further model details including bootstrapped confidence intervals, are in Appendix A5.

As in the models for receptive vocabulary in all four countries, the conditional model for mathematics in Ethiopia shows a strong and inter-meshed network of paths from early economic well-being and maternal education. Economic well-being is strongly related to preschool participation, directly and indirectly through to early quantitative skills at age 5, and on to mathematics ability at age 8 years. Maternal education has the same set of direct and indirect paths, through to early mathematics performance, although the effects appear to be substantially weaker.

Just as the fitted models for receptive vocabulary contained paths from economic well-being through hours spent on household chores and time spent studying at school and at home, on to growth in receptive vocabulary, we see the same pattern for the mathematics model in Ethiopia: there is a set of negative paths from economic well-being, through hours spent studying, which continues as a positive relation to growth in mathematics performance (which in this context means a better relative standing in the cohort at age 15 than earlier). That is, children who are less well off early on (at 1 and 5 years old) end up spending more hours on household chores and tasks, and fewer hours at school and studying at home. In turn, those who spend more hours at school and studying show greater improvement in mathematics.

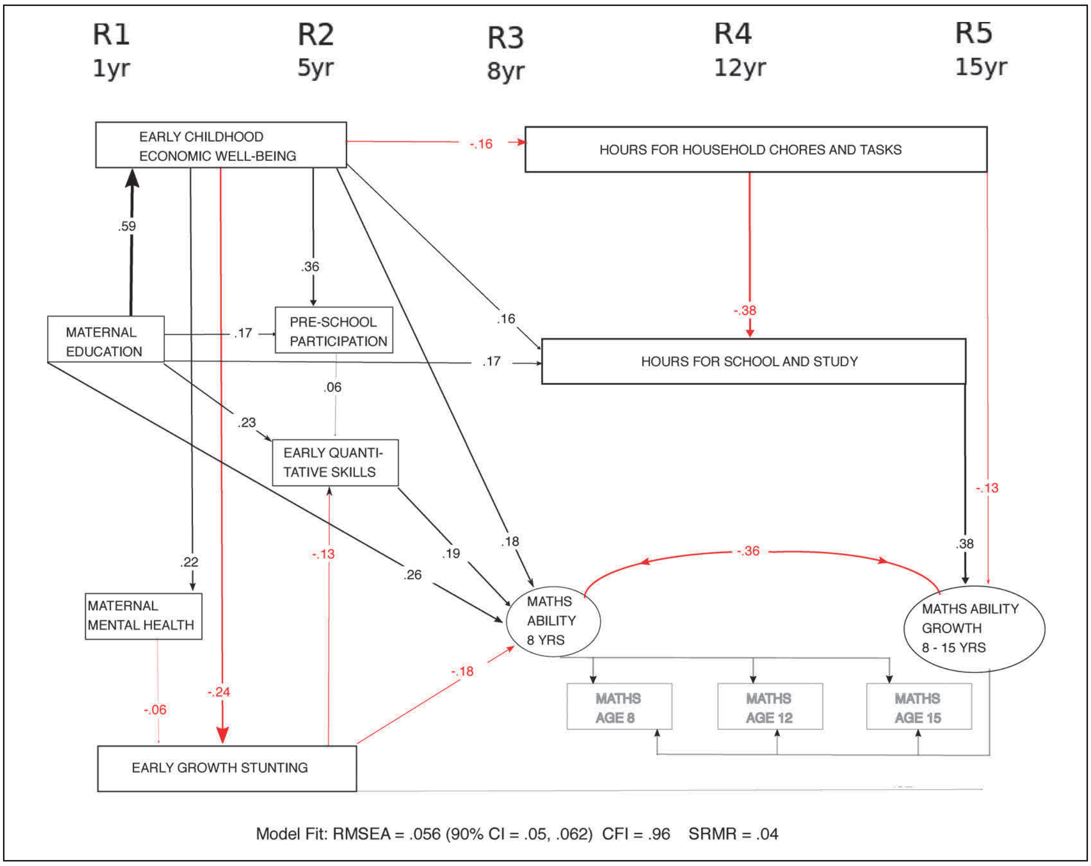
Early growth stunting (i.e. in the first five years) shows a similar set of weak, negative relations to economic well-being, and to early mathematics ability, as we saw in the receptive vocabulary models.

A final point worth noting is the negative covariance between maths intercept and maths slope, which suggests that children who had low scores initially showed greater relative growth than children with higher scores. This could be a catch-up effect again, or perhaps an instance of 'regression to the mean' (participants who scored well below the mean in the first time period could be expected, just by random variation, to score better on later tests).

4.1.2. India

The fitted model for India, shown in Figure 12, exhibits good fit on the three chosen measures: RMSEA, at .056, is just about within the level accepted as good fit, and other two measures are both lower than the conventional level for good fit (CFI = .96, and sRMR = .04, against conventions of .95 and .08, respectively).

Figure 12. *Empirical model of conditional development of mathematics in India*



Notes: All coefficients are statistically significant (at a maximum p < .05), and standardised. N = 1972. Further model details, including bootstrapped confidence intervals, are in Appendix A6.

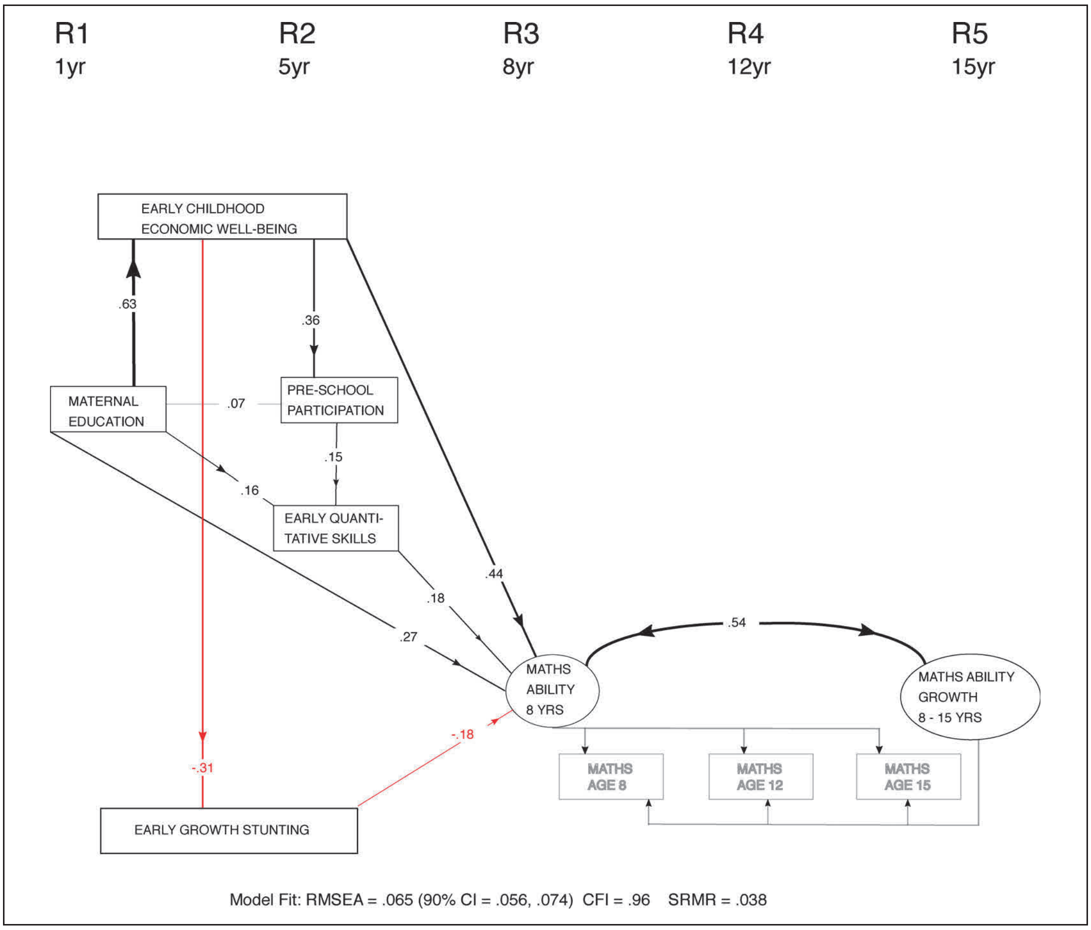
The similarities to the model for Ethiopia are clear, with central roles for early economic well-being and maternal education. There are some differences that should be noted. Maternal mental health (when the child is 1 year old) is negatively related to early growth stunting (women who are more at risk for mental health problems are more likely to have children with stunted growth in early childhood). Poor growth status is itself connected multiply (and negatively) to early quantitative skills, and to maths performance at age 8.

Time spent by children on household chores, and on attending school and studying at home are connected, as in the other countries, to growth in maths performance, all in the expected direction. It is worth pointing out that the relation between hours spent doing household chores and studying/attending school is particularly strong in India (β = -.38). Appendix A6 reports 95 per cent confidence intervals for the coefficients in the model, and for the coefficient in question the interval is -.33, -.44. This is non-overlapping with the coefficient for the Ethiopia model (-.08, -.17), although it might not be wise to compare the coefficients across entirely different models.

4.1.3. *Vietnam*

The model for Vietnam, as shown in Figure 13, is well-fitting in terms of the three standard measures. CFI and sRMR are both below the conventional levels for good models, but RMSE is in between the levels accepted as showing good and moderate fit.

Figure 13. *Empirical model of conditional development of mathematics in Vietnam*



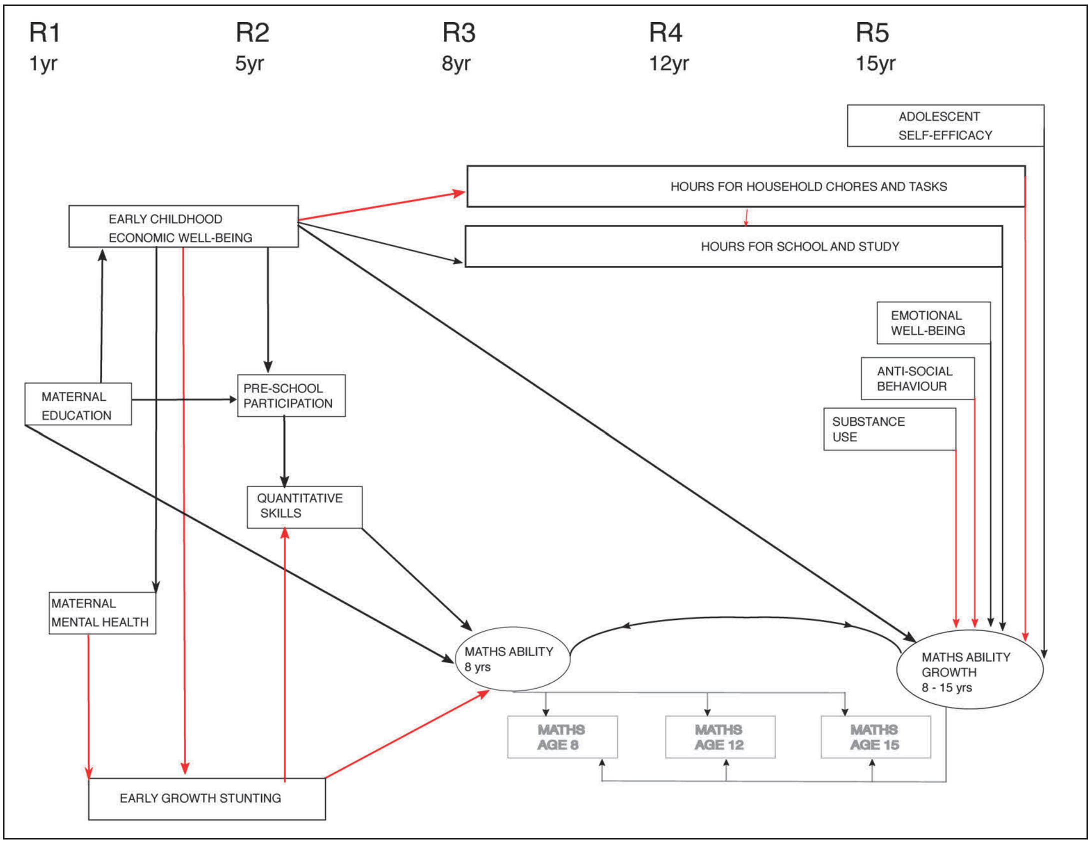
Notes: All coefficients are statistically significant (at a maximum p < .05), and standardised. N = 1951, df = 19. Further model details, including bootstrapped confidence intervals, are in Appendix A7.

The model is simpler than those for Ethiopia and India, containing no significant predictors of growth in mathematics. The dense mesh of paths emanating from early economic well-being, and maternal education is present once again, and there are indirect effects on maths ability at age 8 from both economic well-being and maternal education through preschool participation, and early quantitative skills at age 5. However, the paths connecting early economic well-being to growth in maths performance through reduced time spent on household chores and tasks, and increased time spent on studying at school and at home, are absent from the model (as they were not statistically significant). There is also no path connecting maternal mental health in the first year after the Younger Cohort child's birth to early growth stunting, as there was in the Indian sample. The covariance between the intercept and slope term in the latent growth variable is positive and strong in the Vietnamese model, which suggests that children who had higher scores relative to the rest of the sample at age 8 show greater growth in maths test scores over time.

4.2. Growth in mathematics abilities: Peru

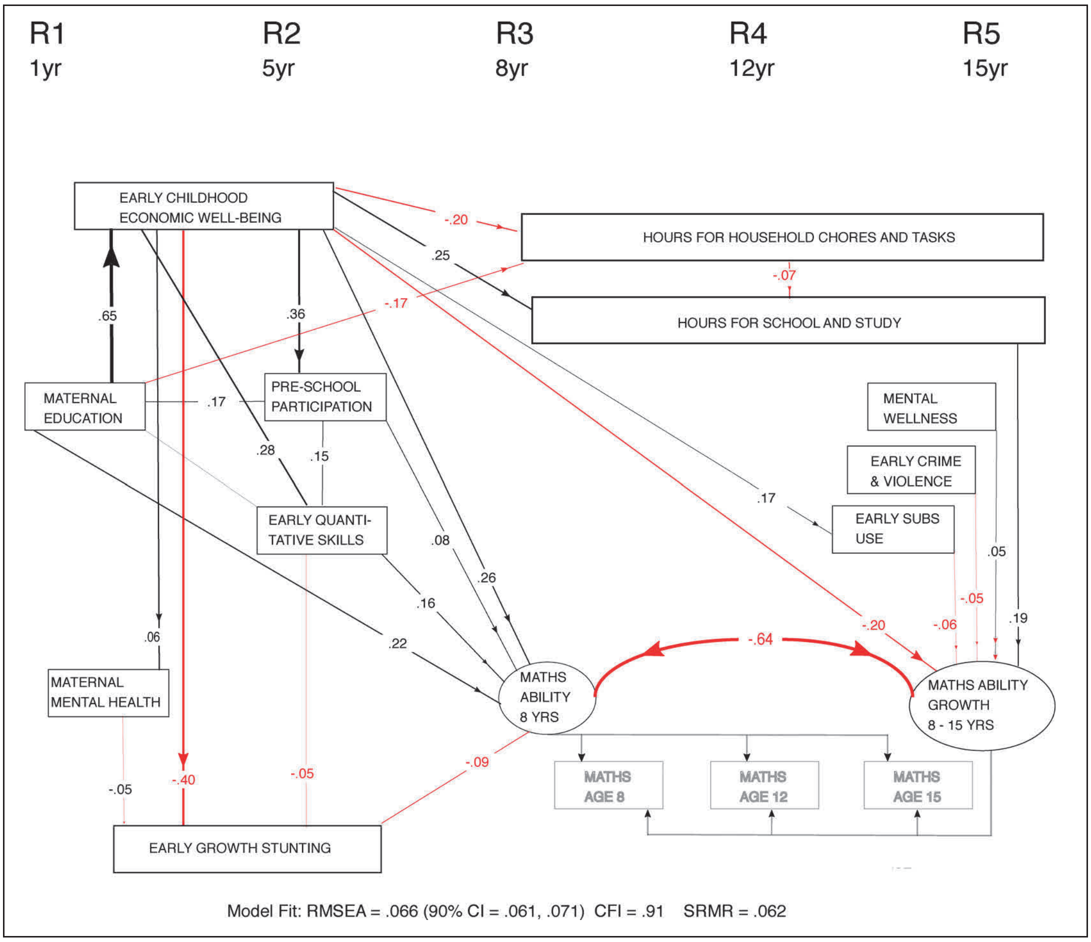
As noted above, the model for Peru differs from the others as psychosocial variables likely to affect learning outcomes are included (mental wellness, early crime and violence, and early substance use). Figure 14 presents the theoretical model for Peru.

Figure 14. *Theoretical model of conditional development of mathematics in Peru*



The model for Peru shown in Figure 15 has adequate fit on two measures (sRMR and RMSEA are within the acceptable range), but not as clearly on the third (CFI is below the convention for a good fitting model (0.95)).

Figure 15. *Empirical model of conditional development of mathematics in Peru*



Notes: All coefficients are statistically significant (at a maximum p < .05), and standardised. N = 1972, df = 64. Further model details, including bootstrapped confidence intervals, are in Appendix A8.

In contrast to the models for India, Ethiopia, and Vietnam, the conditional model of mathematics at age 8 and its growth over the subsequent seven years is considerably more complex. This is partly because we had a greater number of variables in the theoretical model for Peru, but it is also clear that predictors that were not significantly related to growth in mathematics in the other countries were significant in the case of Peru. Thus, we see the same dense mesh of relations between early economic well-being, maternal education, the mediating variables preschool participation, and early quantitative skills, and the outcome variables (the intercept of the latent variable: essentially, mathematics test score at age 8, relative to the entire sample), the slope of the intercept latent variable (growth in mathematics ability, relative to the entire sample), and reading comprehension, at age 15. We also see the less dense mesh of paths involving early growth stunting (at ages 1 and 5), strongly and negatively connected to economic well-being, and weakly and negatively connected to early quantitative skills, and the intercept of the mathematics latent growth variable. Maternal mental health when the child was 1 year old is also weakly connected, from economic well-being (positively) to early growth stunting (negatively), as was the case for India. Additional pathways from economic well-being, and maternal education, through increased time spent on household chores and tasks (negative), and increased time spent on study at school and at home (positive), are present, as they also are for some (but not all) countries. Two additional, weak albeit significant pathways in the Peru sample extend from variables recording adolescent participation in crime and violence, and mental wellness, to growth in mathematics, but without preceding connections in the model (they are not related to either economic well-being or maternal education, in particular). There is, however, a surprising positive connection between economic well-being and adolescent substance use, which continues as an expected negative connection to growth in mathematics (descriptive investigation bears the positive connection out, as children economically better off in the first two rounds indeed report higher rates of substance use in adolescence). Unlike the case in Vietnam, where the intercept and slope components of the mathematics latent growth variable were strongly positively connected, in Peru this relationship is negative, and strong, suggesting a catch-up phenomenon once again.

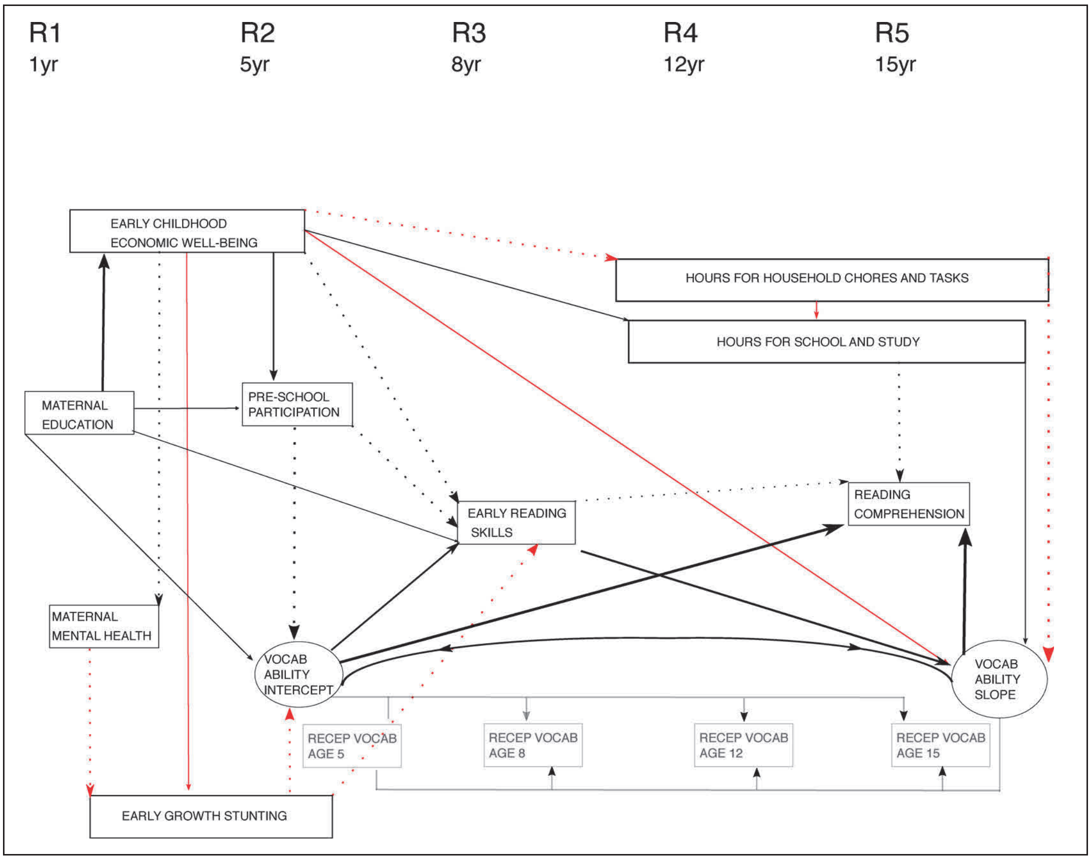
5. Common findings for predictors of the outcomes across the four countries

As is evident from the preceding presentation of the models for receptive vocabulary and mathematics across four different countries, the networks of paths for the countries are strikingly similar. While Peru has additional information on factors affecting adolescent outcomes, in other respects it is similar to the other counties. Common findings provide strong evidence that the paths apply across very different populations, providing a source of cross-national validation for the theoretical models. Commonality is evident for both language and mathematics outcomes. However, there are some differences between the countries, and these should not be ignored.

5.1. Common findings for growth in receptive vocabulary and for reading comprehension

Figure 16 presents a summary model of the ‘big picture’ of relatively strong paths common to all the countries for the conditional growth of receptive vocabulary between the ages of 5 and 15. We show paths that are almost always present with solid lines (in black or red, to indicate positive and negative relations respectively), and those that are present half the time with dashed lines (again, in black or red, to indicate direction). As is the case in earlier figures, rectangular and ellipsoid shapes represent manifest and latent variables, respectively. There are no coefficients in the diagram, as it is a summary across countries.

Figure 16. *Synthesis of common model findings for conditional development of receptive vocabulary across countries*



The composite picture for vocabulary growth and for reading comprehension shows that in at least two countries, early childhood economic well-being is positively associated with maternal mental health: poorer women are at more at risk for mental health problems. This transmits to a risk for growth stunting in their children (red dotted line). As is noted in the rationale for the models, several studies have reported elevated stunting when the child's caregiver has elevated scores on the SRQ20 (which measures risk for anxiety and depression). Poverty independently predicts growth stunting (regardless of maternal well-being). In two countries (red dotted lines), stunting leads to lower scores on receptive vocabulary at 5 years old. The effect of poverty (red solid line) is also evident much later as it reduces the extent to which children's receptive vocabulary has developed by age 15. In two of the countries, poorer children spend more time on domestic responsibilities and this reduces time for school and studies, resulting in lower growth in receptive vocabulary. On the positive side, better-educated mothers have effects on their children's early vocabulary and reading skills, they are more likely to send their children to a better preschool, and in two countries, this in turn strengthens early vocabulary and reading at age 8 – quite probably indicating greater readiness for school. Those skills in early childhood influence vocabulary growth, reading at 8 years old, and more advanced reading skills at 15 years old.

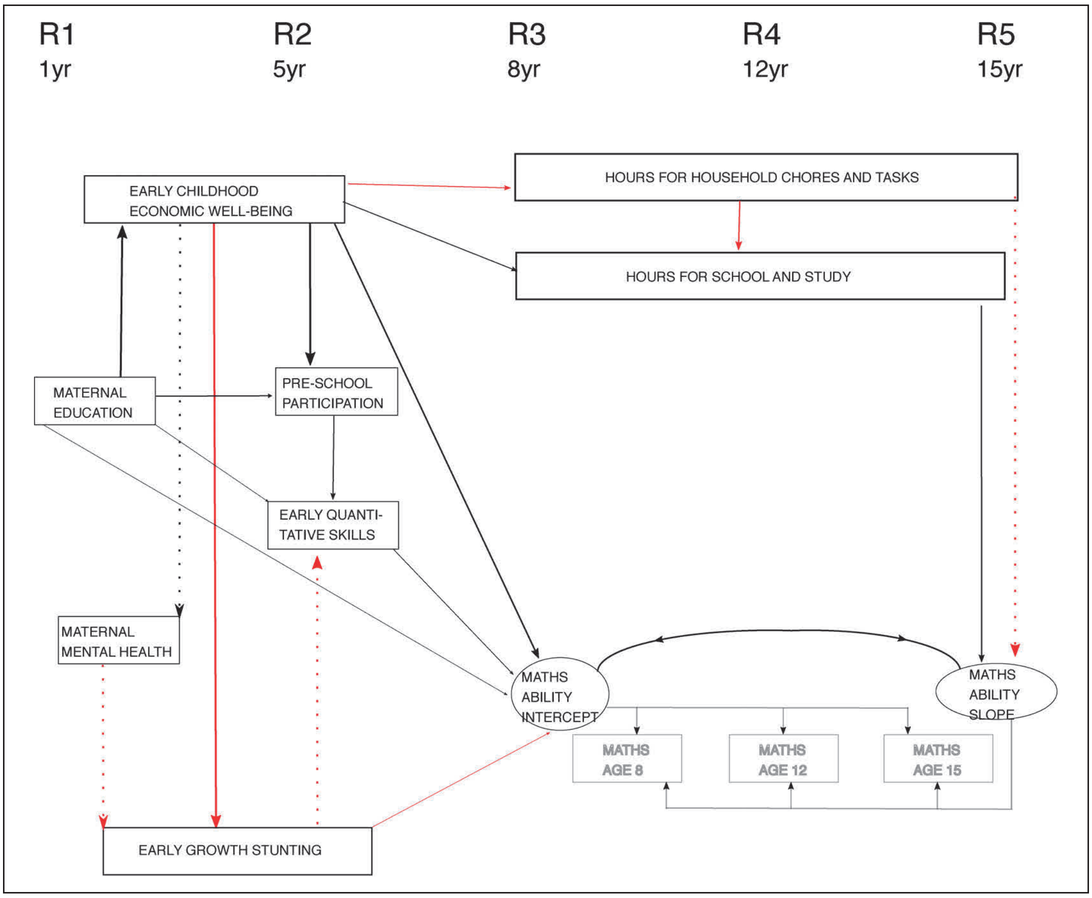
The strongest messages are that poverty in early childhood has both short and long-term effects on language development and literacy skills. Poverty amplifies mental health challenges in caregivers who are probably already vulnerable, and in turn increases stunting risk which compromises the development of literacy skills. Participation in what are likely to be better early learning programmes makes a positive difference to these early skills in two countries (Vietnam and Peru). Finally, and while children's support to the household in the form of chores and other tasks is recognised as being valued by children and their families alike, the models point to costs to both time for schooling and studies, and for the development of literacy skills.

Reading comprehension is a key skill that enables performance in a range of areas, including mathematics (Cappella and Weinstein 2001). The ability to solve mathematics problems is strongly related to reading comprehension abilities (Vilenius-Tuohimaa, Aunola, and Nurmi 2008), and becomes particularly important as adolescents prepare to take the examinations that gain them access to the final years of secondary schooling. Failure at this point precludes access to further education and training and to better paid work.

5.2. Common findings for growth in mathematics abilities

Figure 17 presents a summary model of the 'big picture' of relatively strong paths common to all the countries for the conditional growth of mathematics between the ages of 8 and 15. We show paths that are almost always present with solid lines (in black or red, to indicate positive and negative relations respectively), and those that are present half the time with dashed lines (again, in black or red, to indicate direction). As with earlier figures, rectangular and ellipsoid shapes represent manifest and latent variables, respectively. There are no coefficients in the diagram, as it is a summary across countries.

Figure 17. *Synthesis of common model findings for conditional development of mathematics across countries*



The composite picture for mathematics is simpler than that for the language model. The pattern for the early childhood period is very similar. The effects of early childhood, economic well-being on maternal mental health (red dotted line) and growth stunting remain, and maternal education levels accrue advantage to children's developing quantitative skills (measured by the CDA) at 5 years old. As many Young Lives studies have shown, better-off children have a head start in mathematics tests at 8 years old. In middle childhood and early adolescence, economic advantage is associated with more study time and less household time on household responsibilities.

Children with growth stunting have lower scores on the CDA, and that disadvantage influences their mathematics performance at 8 years old. Of particular interest is evidence across countries that participation in a preschool programme that is likely to be of better quality has positive effects on early quantitative skills, and that advantage influences mathematics abilities at 8 years old. As in the language model, children who contribute greater amounts of time to household tasks and chores have less time for their studies, and in two countries, more of these responsibilities over the period between 8 to 15 years old translates into less growth in their mathematics abilities.

6. Conclusion

Latent Growth Modeling undertaken on two key domains of development in four countries provides unusually powerful cross-country evidence that conditions and proximal relationships in the early childhood environment (Bronfenbrenner and Morris 2006), are particularly important. These include the resources and services (water and sanitation) that constitute the Young Lives wealth index (our measure of economic well-being) and which either advantage or place the young child at risk for developmental hazards such as illness and growth stunting. The models suggest that children whose mothers are rendered vulnerable by the stresses of poverty are at risk for poor growth in the early years, and that this impacts early language, cognition and numeracy skills. Early disadvantages in poorer children are compounded by the challenges of combining household responsibilities with schooling from middle childhood through adolescence, but those whose mothers have more education are more likely to attend early learning programmes.

6.1. Limitations

Modelling exercises were limited to variables that were available and suitable for analysis. A key gap is data on school quality. We do not know how the influence of home and school factors would balance out if the latter could be included in the model. This is a future project to be undertaken with children who participated in both the school and household surveys.

It is unfortunate that neither the agency (self-efficacy) nor self-esteem scales contributed to the models. The agency scale was not a good choice as a measure of self-efficacy as internal consistency was too low for it to be regarded as a satisfactory measure. This should caution against future use of the measure. Self-efficacy needs to be explored further using the more robust Generalised Self- Efficacy Scale of Schwarzer and Jerusalem (1995), and could be considered for revised models.

The particular analytic method we chose, Latent Growth Modeling, is a relatively new development in the statistical analysis of longitudinal data, and there are many aspects about it that are currently not well understood, or about which there is disagreement. We have pointed out that many of the standard fit indices might not apply in the same way to LGM models, and it may be that the indices we report here are not entirely accurate. There are other caveats that one must bear in mind with LGM. The first is that LGM modelling makes strong assumptions about directionality of effects (as does its parent modelling tradition, Structural Equation Modeling (SEM)). It may be that some of the assumptions about directionality that we have made are questionable, and it may turn out in the long run that we are wrong in this respect. This is one of the hazards of any kind of modelling, though it is an important caveat. Second, as we pointed out in the analytic method section, we were unable to take advantage of some aspects of LGM, due to various constraints. We would have liked to have used a fully latent model, where all the predictor variables are defined as latent variables, as this recognises the important problem of measurement error, and tries to do something about it in the model itself. We could not do this, for several reasons, and in fact it seems unlikely to be possible with this kind of dataset, where many variables are not measured originally as scales or do not lend themselves to re-constitution as latent variables. A second possible way to run LGMs is to model multiple latent growth curves within the same model, in order to capture growth in one or more of the predictor variables. This was theoretically possible for the measure of economic well-being, which was measured in all five rounds, and for several other of our predictor variables. However, there are constraints in LGM that make this difficult – for instance, one needs observations to be recorded in the same time windows – and additional difficulties of highly auto-correlated measures made this solution impracticable for present purposes. Hopefully there will be solutions in the not too distant future that will make modelling of this kind possible.

References

Behrman, J.R., J. Hoddinott, J.A. Maluccio, E. Soler-Hampejsek, E.L. Behrman, R. Martorell, M. Ramirez-Zea, and A.D. Stein (2006) *What Determines Adult Cognitive Skills? Impacts of Preschooling, Schooling, and Post-Schooling Experience in Guatemala*, PSC Working Paper Series PSC 06-0, University of Pennsylvania.

Bennett, I.M., W. Schott, S. Krutikova, and J.R. Behrman (2016) 'Maternal Mental Health, and Child Growth and Development, in Four Low-income and Middle-income Countries', *Journal of Epidemiology and Community Health* 70.2: 168–173.

Briones, C. (2017) *'How Many Rooms are There in Your House? Constructing the Young Lives Wealth Index*, Technical Note 43, Oxford: Young Lives.

Bortolotti, S.L., R. Tezza, D.F. Andrade, A.C. Bornia, and A.F. Júnior (2012) 'Relevance and Advantages of Using the Item Response Theory', *Quality & Quantity* 47.4: 2341-2360.

Bronfenbrenner, U., and P. Morris (2006) 'The Bioecological Model of Human Development', In R.M. Lerner, and W. Damon (eds.), *Theoretical Models of Human Development*, (5th Edition, 793-828, New York: Wiley.

Cappella, E., and R.S. Weinstein (2001) 'Turning Around Reading Achievement: Predictors of High School Students' Academic Resilience', *Journal of Educational Psychology* 93.4: 758.

Collins, L.M. (2006) 'Analysis of Longitudinal Data: The Integration of Theoretical Model, Temporal Design, and Statistical Model', *Annual Review of Psychology* 57: 505-528.

Collins, L.M., J.L. Schafer, and C. Kam (2001) 'A Comparison of Inclusive and Restrictive Strategies in Modern Missing Data Procedures', *Psychological Methods* 6.4: 330-351.

Crookston, B.T., W. Schott, S. Cueto, K.A. Dearden, P. Engle, A. Georgiadis, ... and J.R. Behrman (2013) 'Postinfancy Growth, Schooling, and Cognitive Achievement: Young Lives', *The American Journal of Clinical Nutrition* 98.6: 1555-1563.

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

Cueto, S., G. Guerrero, J. Leon, M. Zapata, and S. Freire (2014) 'The Relationship Between Socioeconomic Status at Age One, Opportunities to Learn and Achievement in Mathematics in Fourth Grade in Peru', *Oxford Review of Education* 40.1: 50-72.

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

Dawes, A., J. Streak, S. Levine, and D. Ewing (2012) 'Child Agricultural Work in South Africa: A Contested Space', in J. Boyden and M. Bourdillon (eds.) *Childhood Poverty: Multidisciplinary Approaches*, 283-297, Basingstoke: Palgrave MacMillan.

Dong, Y., and C.-Y.J. Peng (2013) 'Principled Missing Data Methods for Researchers', *SpringerPlus* 2: 222.

Dercon, S., and A. Sánchez (2013) 'Height in Mid-childhood and Psychosocial Competencies in Late Childhood: Evidence From Four Developing Countries', *Economics & Human Biology* 11.4: 426-432.

Dercon, S., and P. Krishnan (2009) 'Poverty and the Psychosocial Competencies of Children: Evidence from the Young Lives Sample in Four Developing Countries', *Children, Youth and Environments* 19.2: 138-163.

Duc, L.T., and J.R. Behrman (2017) 'Heterogeneity in Predictive Power of Early Childhood Nutritional Indicators for Mid-childhood Outcomes: Evidence from Vietnam', *Economics & Human Biology* 26*:* 86-95.

Duncan, G.J., C.J. Dowsett, A. Claessens, K. Magnuson, A.C. Huston, P. Klebanov, … and C. Japel (2007) 'School Readiness and Later Achievement', *Developmental Psychology* 44.1: 1428-1446.

Duncan, G.J., K.M. Ziol-Guest, and A. Kalil (2010) 'Early-childhood Poverty and Adult Attainment, Behavior and Health', *Child Development* 81.1: 306–325.

Duncan, T.E., and S.C. Duncan (2009) 'The ABC's of LGM: an Introductory Guide to Latent Variable Growth Curve Modeling', *Social and Personality Psychology Compass 3.*6: 979-991.

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

Dunn, L., E. Padilla, D. Lugo and L. Dunn (1986) 'Manual del Examinador para el Test de Vocabulario en Imágenes Peabody (Peabody Picture Vocabulary Test) – Adaptación Hispanoamericana (Hispanic-American Adaptation)', Minnesota: AGS.

Ekber, T. and K. Çelik (2009) 'The Relationship Between the Students' Academic Achievement and their Socioeconomic Level: Cross Regional Comparison', *Procedia-Social and Behavioral Sciences* 1.1: 1199-1204.

Favara, M. (2017) 'Do Dreams Come True? Aspirations and Educational Attainments of Ethiopian Boys and Girls', *Journal of African Economies* 26.5: 561-583.

Favara, M., and A. Sanchez (2017) 'Psychosocial Competencies and Risky Behaviours in Peru', *IZA Journal of Labor and Development* 6.3: 1-40.

Feeny, T., and J. Boyden (2003) *Children and Poverty. A Review of Contemporary Literature and Thought on Children and Poverty: Rethinking the Causes, Experiences and Effects,* Richmond VA. Christian Children's Fund.

Fernald, L.C.H., P. Kariger, P. Engle, and A. Raikes (2009) 'Examining Early Child Development in Low-Income Countries: A Toolkit for the Assessment of Children in the First Five Years of Life', Washington DC: The International Bank for Reconstruction and Development/The World Bank.

Georgiadis, A., L. Benny, L.T. Duc, S. Galab, P. Reddy, and T. Woldehanna (2017) 'Growth Recovery and Faltering Through Early Adolescence in Low- and Middle-income Countries: Determinants and Implications for Cognitive Development', *Social Science and Medicine* 179: 81–90.

Gibb, S.J., D.M. Fergusson, and L.J. Horwood (2012) 'Childhood Family Income and Life Outcomes in Adulthood: Findings from a 30-Year Longitudinal Study in New Zealand', *Social Science & Medicine* 74.12: 1979-1986.

Glewwe, P., Q. Chen, and B. Katare (2015) 'What Determines Learning among Kinh and Ethnic Minority Students in Vietnam? An Analysis of the Round 2 Young Lives Data', *Asia and the Pacific Policy Studies* 2.3: 494–516.

Goodman, R. (1997) 'The Strengths and Difficulties Questionnaire: a Research Note', *Journal of Child Psychology and Psychiatry 38.*5: 581–6.

Gove, A., and A. Wetterberg (2011) 'The Early Grade Reading Assessment. Applications and Interventions to Improve Basic Literacy', North Carolina: Research Triangle Institute.

Gottfredson, L.S. (1981) 'Circumscription and Compromise: A Developmental Theory of Occupational Aspirations', *Journal of Counseling Psychology* 28.6: 545.

Grantham-McGregor, S., Y. Cheung, S. Cueto, P. Glewwe, L. Richter, and B. Strupp (2007) 'Developmental Potential in the First 5 years for Children in Developing Countries', *The Lancet* 369: 60–70.

Guerrero, G., C. Sugimaru, A. Cussianovich, B. De Fraine, and S. Cueto (2016) *Education Aspirations among Young People in Peru and their Perceptions of Barriers to Higher Education*, Working Paper 148, Oxford: Young Lives.

Harpham, T., S. Huttley, M. De Silva, and T. Abramsky (2005) 'Maternal Mental Health and Child Nutritional Status in Four Developing Countries', *Journal of Epidemiology & Community Health* 59.12: 1060–1064.

Harpham, T., M. Reichenheim, R. Oser, et al. (2003) 'How to do (or not to do) … Measuring Mental Health in a Cost-effective Manner', *Health and Policy Planning* 18.*3:* 344–9.

Hecht, S.A., S.R. Burgess, J.K. Torgesen, R.K. Wagner, and C.A. Rashotte (2000) 'Explaining Social Class Differences in Growth of Reading Skills from Beginning Kindergarten Through Fourth-Grade: The Role of Phonological Awareness, Rate of Access, and Print Knowledge', *Reading and Writing* 12.1: 99-128.

Hooper, D., J. Coughlan, and M.R. Mullen (2008) 'Structural Equation Modelling: Guidelines for Determining Model Fit', *The Electronic Journal of Business Research Methods* 6.1: 53-60.

Hu, L.T., and P.M. Bentler (1999) 'Cutoff Criteria for Fit Indexes in Covariance Structure Analysis: Conventional Criteria Versus New Alternatives', *Structural Equation Modeling* 6.1: 1-55.

Kenny, D. (2017) 'Measuring Model Fit', http://davidakenny.net/cm/fit.htm (accessed 1 March 2018).

Leon, J., and A. Singh (2017) *Equating Test Scores for Receptive Vocabulary Across Rounds and Cohorts in Ethiopia, India and Vietnam*, Technical Note 40, Oxford: Young Lives.

Krutikova, S., C. Rolleston, and E. Aurino (2014) 'How Much Difference Does School Make and For Whom? A Two-Country Study of the Impact of School Quality on Educational Attainment', in M. Bourdillon and J. Boyden (eds.) *Growing Up in Poverty: Findings from Young Lives*, 201-224, Basingstoke: Palgrave McMillan.

Lund, C., A. Breen, A.J. Flisher, R. Kakuma, J. Corrigall, J.A. Joska, ... and V. Patel (2010) 'Poverty and Common Mental Disorders in Low and Middle-income Countries: a Systematic Review', *Social Science & Medicine* 71.3: 517-528.

Lundeen, A.E, A.D. Stein, L. Adair, on behalf of the COHORTS Investigators (2014) 'Height-For-Age Z Scores Increase Despite Increasing Height Deficits Among Children in 5 Developing Countries', *American Journal of Clinical Nutrition* 100*:* 821-825.

Marsh, H.W., R.H. Parada, and V. Ayotte (2004) 'A Multidimensional Perspective of Relations Between Self-Concept (Self Description Questionnaire II) and Adolescent Mental Health (Youth Self-Report)', *Psychological Assessment* 16.1: 27-41.

McLoyd, V. (1998) 'Socio-economic Disadvantage and Child Development', *American Psychologist* 53.*2:* 185-204.

Mercy, J.A., and L.C. Steelman (1982) 'Familial Influence on the Intellectual Attainment of Children', *American Sociological Review 47.*4: 532-542.

Nores, M., and W. Barnett (2010) 'Benefits of Early Childhood Interventions Across the World: (Under) Investing in the Very Young', *Economics of Education Review* 29.2: 271-282.

Orkin, K. (2012) 'Are Work and Schooling Complementary or Competitive for Children in Rural Ethiopia?' in J. Boyden and M. Bourdillon (eds.) *Childhood Poverty: Multidisciplinary Approaches*, 298–315, Basingstoke: Palgrave Macmillan.

Pankhurst, A., G. Crivello, and A. Tiumelissan (2016) *Children's Work in Family and Community Contexts: Examples from Young Lives Ethiopia*, Working Paper 147, Oxford: Young Lives.

Preacher, K.J., A.L. Wichman, R.C. MacCallum, and N.E. Briggs (2008) *Latent Growth Curve Modeling,* London: Sage.

Psacharopoulos, G. and H. Patrinos (2004) 'Returns to Investment in Education: A Further Update', *Education Economics* 12.2: 111–34.

Reardon, S.F. (2011) 'The Widening Academic Achievement Gap Between the Rich and the Poor: New Evidence and Possible Explanations', in G.J. Duncan and R. Murnan (eds.) *Whither Opportunity? Rising Inequality, Schools, and Children's Life Chances*, 91-116, New York: Russell Sage Foundation.

Rebello-Britto, P., P. Engle, and C. Super (eds.) (2013) *Handbook of Early Childhood Development Research and Its Impact on Global Policy,* New York: Oxford University Press.

Rolleston, C., and Z. James (2011) 'The Role of Schooling in Skill Development: Evidence from Young Lives in Ethiopia, India, Peru and Vietnam', background paper prepared for the Education for All Global Monitoring Report 2012, Paris: UNESCO.

Rosseel, Y. (2012) 'lavaan: An R Package for Structural Equation Modeling'*, Journal of Statistical Software* 48.2: 1-36.

Schwab-Stone, M.E., T.S. Ayers, W. Kasprow, C. Voyce, C. Barone, T. Shriver, and R.P. Weissberg (1995) 'No Safe Haven: A Study of Violence Exposure in an Urban Community', *Journal of the American Academy of Child & Adolescent Psychiatry* 34.10: 1343–1352.

Serneels, P., and S. Dercon (2014) *Aspirations, Poverty and Education*, Working Paper 125, Oxford: Young Lives.

Schwarzer, R., and M. Jerusalem(1995) 'Generalized Self-Efficacy Scale', In J. Weinman, S. Wright, and M. Johnston, (eds.) *Measures in Health Psychology: A User's Portfolio*, 35-37, Windsor: NFER-NELSON.

Serneels, P., and S. Dercon (2014) *Aspirations, Poverty and Education. Evidence from India*, Working Paper 125, Oxford: Young Lives.

Siegler, R.S., G.J. Duncan, P.E. Davis-Kean, K. Duckworth, A. Claessens, M. Engel, M., ... and M. Chen, M. (2012) 'Early Predictors of High School Mathematics Achievement', *Psychological Science* 23.7: 691-697.

Tafere, Y. (2017) *Educational Trajectories From Childhood to Early Adulthood: Aspirations, Gender and Poverty in Ethiopia*, Working Paper 171, Oxford: Young Lives.

Tuan, T., T. Harpham, and N.T. Huong (2004) 'Validity and Reliability of the Self-reporting Questionnaire 20 Items in Vietnam', *Hong Kong Journal of Psychiatry* 14.3: 15-18.

Vilenius-Tuohimaa, P.M., K. Aunola, and J.E. Nurmi (2008) 'The Association Between Mathematical Word Problems and Reading Comprehension', *Educational Psychology* 28.4: 409-426.

Watts, T.W., G.J. Duncan, M. Chen, A. Claessens, P.E. Davis-Kean, K. Duckworth, ... and M.I. Susperreguy (2015) 'The Role of Mediators in the Development of Longitudinal Mathematics Achievement Associations', *Child Development* 86.6: 1892-1907.

Wachs, T., and A. Rahman (2013) 'The Nature and Impact of Risk and Protective Influences on Children's Development in Low and Middle-income Countries', in P. Rebello-Britto, P. Engle, and C. Super (eds.) *Handbook of Early Childhood Development Research and its Impact on Global Policy*, 85-122), New York: Oxford University Press.

Watts, T.W., G.J. Duncan, M. Chen, A. Claessens, P.E. Davis-Kean, K. Duckworth, and M.I. Susperreguy (2015) 'The Role of Mediators in the Development of Longitudinal Mathematics Achievement Associations', *Child Development* 86.6: 1892-1907.

White, K.R. (1982) 'The Relation Between Socio-economic Status and Academic Achievement', *Psychological Bulletin* 91: 461–481.

Wigfield, A., and J.S. Eccles (2000) 'Expectancy–value Theory of Achievement Motivation', *Contemporary Educational Psychology* 25.1: 68-81.

Woldehanna, T., and A. Gebremedhin (2015) *Is Child Work Detrimental to the Educational Achievement of Children? Results from Young Lives in Ethiopia*, Working Paper 140, Oxford: Young Lives.

Woldehanna, T., and M.W. Araya (2017) *Early Investment in Preschool and Completion of Secondary Education in Ethiopia: Lessons Drawn from Young Lives*, Working Paper 168, Oxford: Young Lives.

Woodhead, M., J. Rossiter, A. Dawes, and A. Pankhurst (2017) ‘Scaling-up Early Learning as a Sustainable Development Priority: A Case Study of Ethiopia’, in L. Miller, C. Cameron, C. Dalli, and N. Barbour (eds.) *Sage Handbook of Early Childhood Policy*, 219-235, New York: Sage.

Woodhead, M., P. Ames, U. Vennam, W. Abebe, and N. Streuli (2009) *Equity and Quality? Challenges for Early Childhood and Primary Education in Ethiopia, India and Peru*, Working Paper No. 55, Studies in Early Childhood Transitions, The Hague: Bernard van Leer Foundation.

World Health Organisation (1994) ‘A User’s Guide to the Self Reporting Questionnaire SRQ20’, Geneva: World Health Organisation.

Yorke, L., and M.J. Ogando Portela (2018) *Psychosocial Scales in the Young Lives Round 4 Survey: Selection, Adaptation and Validation*, Technical Note 45, Oxford: Young Lives.

Appendices

Appendix A

These appendices contain the conditional LGM modelling results. We used R version 3.4.3 (R Core Team 2017), with package Lavaan (version 0.5; Rosseel 2012) for the latent growth modelling. In the appendices we report the following, under corresponding headings.

1. Overall model fit statistics

2. Fitted coefficients for model parameters. These are standardised coefficients (we do not report unstandardised coefficients, as variables were scaled at initial stages of modelling to avoid disparate variance size problems, and are not interpretable as scale values). We report standard errors and normal theory probability tests of coefficient size (against HO of no effect). However, we also report 95% bootstrapped confidence intervals for each estimated coefficient, in order to buttress against possible distributional problems, and to give some direct sense of coefficient variability. Bootstrap sampling was done with Lavaan, using 1,000 re-samples. The key for reading these is as follows:

Lavaan notation for parameters

Formula type	operator	mnemomic
latent variable definition	=~	is measured by
Regression	~	is regressed on
(residual) (co)variance	~~	is correlated with
Intercept	~1	intercept

Variable names/abbreviations

vocab_i = latent intercept (consituted by zero weighting of vocab_R2…vocab_R5)
vocab_s = latent slope (constituted by age weighting of vocab_R2…vocab_R5)
vocab_R2 = receptive vocabulary at Round 2 (age ≈ 5)
vocab_R3 = receptive vocabulary at Round 2 (age ≈ 5)
vocab_R4 = receptive vocabulary at Round 2 (age ≈ 5)
vocab_R5 = receptive vocabulary at Round 2 (age ≈ 5)
wi_r12 = aggregate wealth index over Rounds 1 and 2
stunt_r12 = early growth stunting over Rounds 1 and 2
pre_school_var = pre-school participation and quality
egra_rasch = early reading test score
house_chore_r345 = hours spent doing house chores and similar tasks, aggregated over Rounds 3 to 5
schoolstudyhrs_r345 = hours spent studying, aggregated over Rounds 3 to 5
read_comp_r5 = reading comprehension score at Round 5
momedu = maternal education
agemon = age of child in months
early_subs = substance use at age 15
Early crime and violence = participation in crime, gangs and violence at age 15

3. Particular notes about each model are made in a final section. These are self-explanatory, and are only made where we need to clarify something in particular.

Appendix A1. Modeling results for conditional growth of receptive vocabulary in Ethiopia

Overall model fit statistics

lavaan (0.5-23.1097) converged normally after 182 iterations	
Number of observations	1439
Number of missing patterns	18
Estimator	ML
Minimum Function Test Statistic	504.205
Degrees of freedom	50
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	7488.987
Degrees of freedom	78
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.939
Tucker-Lewis Index (TLI)	0.904
Root Mean Square Error of Approximation:	
RMSEA	0.079
90 per cent Confidence Interval	0.073 0.086
P-value RMSEA <=	0.05 0.000
Standardised Root Mean Square Residual:	
SRMR	0.053

Fitted coefficients/parameter estimates

lhs	op	Rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_i	=~	vocab_R2	0.6695	0.0322	20.7920	0.0000	0.6062	0.7191
vocab_i	=~	vocab_R3	0.4513	0.0236	19.1165	0.0000	0.4020	0.4932
vocab_i	=~	vocab_R4	0.4579	0.0249	18.3989	0.0000	0.4055	0.5028
vocab_i	=~	vocab_R5	0.4659	0.0251	18.5952	0.0000	0.4114	0.5124
vocab_i	~~	vocab_i	0.2670	0.0658	4.0566	0.0000	0.1287	0.3670
vocab_i	~1		-2.9696	0.4185	-7.0951	0.0000	-3.7818	-2.2412
vocab_s	=~	vocab_R2	0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_s	=~	vocab_R3	0.2443	0.0107	22.7655	0.0000	0.2218	0.2646
vocab_s	=~	vocab_R4	0.5783	0.0234	24.6797	0.0000	0.5306	0.6224
vocab_s	=~	vocab_R5	0.8407	0.0306	27.4695	0.0000	0.7819	0.8985
vocab_s	~~	vocab_s	0.7654	0.0238	32.2214	0.0000	0.7135	0.8145
vocab_s	~1		-0.1255	0.5707	-0.2200	0.8259	-1.3831	1.2679
vocab_i	~~	vocab_s	-0.1010	0.1094	-0.9231	0.3560	-0.2663	0.2003
vocab_R2	~~	vocab_R2	0.5517	0.0431	12.7962	0.0000	0.4830	0.6325
vocab_R3	~~	vocab_R3	0.5958	0.0165	36.1190	0.0000	0.5646	0.6270
vocab_R4	~~	vocab_R4	0.2678	0.0132	20.2830	0.0000	0.2281	0.3180
vocab_R5	~~	vocab_R5	0.1349	0.0178	7.5695	0.0000	0.0930	0.1795
vocab_R3	~	vocab_R2	0.0950	0.0188	5.0648	0.0000	0.0590	0.1345
vocab_R4	~	vocab_R3	0.0607	0.0213	2.8468	0.0044	0.0213	0.1001

lhs	op	Rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_R5	~	vocab_R4	-0.1344	0.0182	-7.3800	0.0000	-0.1686	-0.0973
wi_r12	~	momedu	0.6244	0.0167	37.4179	0.0000	0.5897	0.6586
stunt_r12	~	wi_r12	-0.2668	0.0248	-10.7418	0.0000	-0.3103	-0.2223
pre_school_var	~	wi_r12	0.4926	0.0224	21.9550	0.0000	0.4431	0.5436
pre_school_var	~	momedu	0.3018	0.0242	12.4751	0.0000	0.2401	0.3572
egra_rasch	~	wi_r12	0.3105	0.0308	10.0952	0.0000	0.2542	0.3762
egra_rasch	~	stunt_r12	-0.0674	0.0247	-2.7306	0.0063	-0.1158	-0.0186
egra_rasch	~	momedu	0.0985	0.0314	3.1364	0.0017	0.0255	0.1678
house_chore_r345	~	wi_r12	-0.1824	0.0255	-7.1558	0.0000	-0.2274	-0.1339
schoolstudyhrs_r345	~	wi_r12	0.3562	0.0275	12.9526	0.0000	0.3072	0.4078
schoolstudyhrs_r345	~	house_chore_r345	-0.1238	0.0210	-5.8958	0.0000	-0.1732	-0.0740
schoolstudyhrs_r345	~	pre_school_var	0.2832	0.0278	10.2000	0.0000	0.2332	0.3278
vocab_i	~	wi_r12	0.6178	0.0399	15.4832	0.0000	0.5387	0.7127
vocab_i	~	agemon	0.2523	0.0273	9.2515	0.0000	0.2039	0.3051
vocab_i	~	momedu	0.2750	0.0349	7.8700	0.0000	0.1937	0.3517
vocab_s	~	schoolstudyhrs_r345	0.3732	0.0280	13.3158	0.0000	0.3071	0.4354
vocab_s	~	egra_rasch	0.2352	0.0299	7.8661	0.0000	0.1617	0.3053
read_comp_r5	~	vocab_s	0.3201	0.0303	10.5728	0.0000	0.2594	0.3860
read_comp_r5	~	vocab_i	0.2059	0.0375	5.4968	0.0000	0.1356	0.2797
read_comp_r5	~	schoolstudyhrs_r345	0.1527	0.0304	5.0298	0.0000	0.0906	0.2157
read_comp_r5	~	pre_school_var	0.1126	0.0316	3.5657	0.0004	0.0517	0.1719
read_comp_r5	~	egra_rasch	0.1242	0.0250	4.9622	0.0000	0.0745	0.1716
wi_r12	~1		1.2142	0.0405	30.0023	0.0000	1.1470	1.2839
momedu	~1		0.7173	0.0300	23.9036	0.0000	0.6823	0.7522
pre_school_var	~1		0.7645	0.0454	16.8459	0.0000	0.7058	0.8236
egra_rasch	~1		19.6752	0.3846	51.1523	0.0000	18.9902	20.3801
read_comp_r5	~1		-2.8482	0.4454	-6.3950	0.0000	-3.7197	-1.9833
stunt_r12	~1		1.2951	0.0484	26.7479	0.0000	1.2098	1.3819
schoolstudyhrs_r345	~1		2.0887	0.0810	25.7981	0.0000	1.9144	2.2756
agemon	~1		16.2510	0.3041	53.4450	0.0000	15.8320	16.6828
house_chore_r345	~1		1.7514	0.0537	32.6339	0.0000	1.6599	1.8450
vocab_R2	~~	egra_rasch	0.1384	0.0307	4.5127	0.0000	0.0789	0.2016
vocab_R3	~~	egra_rasch	0.2703	0.0260	10.3939	0.0000	0.2168	0.3196
vocab_R4	~~	egra_rasch	0.1282	0.0290	4.4224	0.0000	0.0681	0.1941
wi_r12	~~	wi_r12	0.6101	0.0208	29.2708	0.0000	0.5663	0.6522
stunt_r12	~~	stunt_r12	0.9288	0.0133	70.0637	0.0000	0.9037	0.9506
pre_school_var	~~	pre_school_var	0.4806	0.0185	26.0241	0.0000	0.4394	0.5196
egra_rasch	~~	egra_rasch	0.8378	0.0179	46.8430	0.0000	0.8012	0.8675
house_chore_r345	~~	house_chore_r345	0.9667	0.0093	104.0065	0.0000	0.9483	0.9821
schoolstudyhrs_r345	~~	schoolstudyhrs_r345	0.6154	0.0201	30.6553	0.0000	0.5789	0.6479
read_comp_r5	~~	read_comp_r5	0.5819	0.0220	26.4466	0.0000	0.5339	0.6247
momedu	~~	momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
agemon	~~	agemon	1.0000	0.0000	NA	NA	1.0000	1.0000
vocab_R2	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: Error terms for vocabulary scores in Rounds 2 to 4 were allowed to correlate with the error terms for early reading scores, in order to reduce high model residuals. An auto-regressive structure (lag 1) was used to represent time-dependent correlation between the vocabulary scores at Rounds 2 to 5. 95% confidence intervals around standardised coefficients were computed through bootstrapping (1,000 resamples).

Appendix A2. Modeling results for conditional growth of receptive vocabulary in India

Overall model fit statistics

Lavaan (0.5-23.1097) converged normally after 141 iterations	
Number of observations	2011
Number of missing patterns	31
Estimator	ML
Minimum Function Test Statistic	648.269
Degrees of freedom	65
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	6013.081
Degrees of freedom	91
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.902
Tucker-Lewis Index (TLI)	0.862
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-11470.914
Loglikelihood unrestricted model (H1)	-11146.780
Number of free parameters	54
Akaike (AIC)	23049.828
Bayesian (BIC)	23352.573
Sample-size adjusted Bayesian (BIC)	23181.012
Root Mean Square Error of Approximation:	
RMSEA	0.067
90 per cent Confidence Interval	0.062 0.072
P-value RMSEA <= 0.05	0.000
Standardised Root Mean Square Residual:	
SRMR	0.060

Fitted coefficients/parameter estimates

lhs	op	Rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_i	=~	vocab_R2	0.8422	0.0213	39.6281	0.0000	0.8022	0.8844
vocab_i	=~	vocab_R3	0.8159	0.0264	30.9178	0.0000	0.7615	0.8646
vocab_i	=~	vocab_R4	0.8177	0.0268	30.4782	0.0000	0.7527	0.8832
vocab_i	=~	vocab_R5	0.7678	0.0248	30.9834	0.0000	0.7017	0.8116
vocab_i	~~	vocab_i	0.7883	0.0217	36.3668	0.0000	0.7301	0.8353
vocab_i	~1		0.0504	0.2589	0.1946	0.8457	-0.3831	0.3989
vocab_s	=~	vocab_R2	0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_s	=~	vocab_R3	0.3105	0.0113	27.5305	0.0000	0.2899	0.3330
vocab_s	=~	vocab_R4	0.7260	0.0270	26.8894	0.0000	0.6641	0.7842
vocab_s	=~	vocab_R5	0.9738	0.0307	31.7633	0.0000	0.9047	1.0267
vocab_s	~~	vocab_s	0.8044	0.0162	49.7634	0.0000	0.7585	0.8389
vocab_s	~1		-2.1375	0.3518	-6.0763	0.0000	-2.9436	-1.3951
vocab_i	~~	vocab_s	-0.8421	0.0152	-55.3572	0.0000	-0.8705	-0.8050
vocab_R2	~~	vocab_R2	0.2907	0.0358	8.1187	0.0000	0.2178	0.3565
vocab_R3	~~	vocab_R3	0.7366	0.0101	73.1562	0.0000	0.7146	0.7543
vocab_R4	~~	vocab_R4	0.5639	0.0166	34.0432	0.0000	0.5182	0.6101

lhs	op	Rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_R5	~~	vocab_R5	0.5549	0.0209	26.4940	0.0000	0.5042	0.6034
vocab_R3	~	vocab_R2	-0.2626	0.0204	-12.8446	0.0000	-0.3044	-0.2201
vocab_R4	~	vocab_R3	0.1186	0.0210	5.6442	0.0000	0.0669	0.1668
vocab_R5	~	vocab_R4	-0.0411	0.0146	-2.8160	0.0049	-0.0708	-0.0064
wi_r12	~	momedu	0.5908	0.0147	40.1999	0.0000	0.5586	0.6289
mother_mhealth	~	wi_r12	-0.2182	0.0212	-10.2753	0.0000	-0.2639	-0.1796
stunt_r12	~	wi_r12	-0.2441	0.0220	-11.0708	0.0000	-0.2849	-0.2004
stunt_r12	~	mother_mhealth	0.0604	0.0226	2.6677	0.0076	0.0221	0.1083
pre_school_var	~	wi_r12	0.3585	0.0240	14.9217	0.0000	0.3194	0.4257
pre_school_var	~	momedu	0.1666	0.0249	6.6817	0.0000	0.1030	0.1979
egra_rasch	~	stunt_r12	-0.1237	0.0218	-5.6735	0.0000	-0.1620	-0.0856
egra_rasch	~	momedu	0.1533	0.0220	6.9732	0.0000	0.1039	0.1986
house_chore_r345	~	momedu	-0.1672	0.0226	-7.4007	0.0000	-0.1947	-0.1299
schoolstudyhrs_r345	~	wi_r12	0.2515	0.0204	12.3460	0.0000	0.2123	0.2929
schoolstudyhrs_r345	~	house_chore_r345	-0.3610	0.0195	-18.5540	0.0000	-0.4193	-0.3160
vocab_i	~	wi_r12	0.2689	0.0265	10.1480	0.0000	0.2149	0.3205
vocab_i	~	agemon	0.0664	0.0148	4.4856	0.0000	0.0446	0.0900
vocab_i	~	momedu	0.1924	0.0194	9.9114	0.0000	0.1604	0.2347
vocab_i	~	stunt_r12	-0.1173	0.0157	-7.4593	0.0000	-0.1446	-0.0878
vocab_s	~	wi_r12	-0.2527	0.0284	-8.8829	0.0000	-0.3158	-0.1928
vocab_s	~	schoolstudyhrs_r345	0.3312	0.0192	17.2186	0.0000	0.2737	0.3865
vocab_s	~	egra_rasch	0.2826	0.0187	15.1383	0.0000	0.2497	0.3192
read_comp_r5	~	vocab_s	0.9725	0.0410	23.7432	0.0000	0.8860	1.0435
read_comp_r5	~	vocab_i	0.9069	0.0430	21.0715	0.0000	0.8141	0.9932
wi_r12	~1		1.8817	0.0432	43.5109	0.0000	1.8093	1.9592
momedu	~1		0.6976	0.0250	27.8774	0.0000	0.6662	0.7211
pre_school_var	~1		2.4842	0.0791	31.3999	0.0000	2.3473	2.5898
egra_rasch	~1		20.0557	0.3264	61.4494	0.0000	19.4131	20.7731
read_comp_r5	~1		-3.1202	0.2436	-12.8092	0.0000	-3.5262	-2.5640
stunt_r12	~1		1.2315	0.0687	17.9245	0.0000	1.0434	1.3392
schoolstudyhrs_r345	~1		5.0474	0.1034	48.8222	0.0000	4.6670	5.4851
agemon	~1		17.2812	0.2787	62.0055	0.0000	16.9334	17.7025
house_chore_r345	~1		1.6750	0.0360	46.5555	0.0000	1.6069	1.7610
mother_mhealth	~1		1.8940	0.0544	34.8167	0.0000	1.7942	1.9960
vocab_R2	~~	egra_rasch	0.2260	0.0408	5.5425	0.0000	0.1563	0.2932
vocab_R3	~~	egra_rasch	0.4131	0.0229	18.0131	0.0000	0.3679	0.4468
wi_r12	~~	wi_r12	0.6509	0.0174	37.4830	0.0000	0.6045	0.6879
mother_mhealth	~~	mother_mhealth	0.9524	0.0093	102.7488	0.0000	0.9304	0.9677
stunt_r12	~~	stunt_r12	0.9304	0.0113	82.5504	0.0000	0.9088	0.9462
pre_school_var	~~	pre_school_var	0.7731	0.0167	46.2152	0.0000	0.7287	0.8073
egra_rasch	~~	egra_rasch	0.9554	0.0090	106.3252	0.0000	0.9330	0.9687
house_chore_r345	~~	house_chore_r345	0.9720	0.0076	128.6249	0.0000	0.9621	0.9831
schoolstudyhrs_r345	~~	schoolstudyhrs_r345	0.7885	0.0161	49.0565	0.0000	0.7350	0.8181
read_comp_r5	~~	read_comp_r5	0.4834	0.0243	19.8824	0.0000	0.4358	0.5280
momedu	~~	Momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
agemon	~~	Agemon	1.0000	0.0000	NA	NA	1.0000	1.0000
vocab_R2	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: Error terms for vocabulary scores in Rounds 2 to 3 were allowed to correlate with the error terms for early reading scores, in order to reduce high model residuals. An auto-regressive structure (lag 1) was used to represent time-dependent correlation between the vocabulary scores at Rounds 2 to 5. 95% confidence intervals around standardised coefficients were computed through bootstrapping (1,000 resamples).

Appendix A3. Modeling results for conditional growth of receptive vocabulary in Vietnam

Overall model fit statistics

lavaan (0.5-23.1097) converged normally after 123 iterations	
	Used Total
Number of observations	1990 1999
Number of missing patterns	38
Estimator	ML
Minimum Function Test Statistic	647.642
Degrees of freedom	45
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	6689.142
Degrees of freedom	66
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.909
Tucker-Lewis Index (TLI)	0.867
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-11497.562
Loglikelihood unrestricted model (H1)	-11173.741
Number of free parameters	45
Akaike (AIC)	23085.125
Bayesian (BIC)	23336.940
Sample-size adjusted Bayesian (BIC)	23193.973
Root Mean Square Error of Approximation:	
RMSEA	0.082
90 per cent Confidence Interval	0.077 0.088
P-value RMSEA <= 0.05	0.000
Standardised Root Mean Square Residual:	
SRMR	0.066

Fitted coefficients/parameter estimates

lhs	op	rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_i	=~	vocab_R2	0.7470	0.0160	46.8117	0.0000	0.7177	0.7745
vocab_i	=~	vocab_R3	0.6644	0.0195	34.1426	0.0000	0.6293	0.6974
vocab_i	=~	vocab_R4	0.5472	0.0171	32.0673	0.0000	0.5055	0.5925
vocab_i	=~	vocab_R5	0.5457	0.0176	31.0245	0.0000	0.5053	0.5872
vocab_i	~~	vocab_i	0.4296	0.0257	16.7287	0.0000	0.3776	0.4816
vocab_i	~1		-3.0678	0.3418	-8.9760	0.0000	-3.7772	-2.3551
vocab_s	=~	vocab_R2	0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_s	=~	vocab_R3	0.2656	0.0124	21.3740	0.0000	0.2369	0.2954
vocab_s	=~	vocab_R4	0.5103	0.0228	22.3503	0.0000	0.4583	0.5619
vocab_s	=~	vocab_R5	0.7271	0.0307	23.6844	0.0000	0.6586	0.7946
vocab_s	~~	vocab_s	0.8070	0.0234	34.4925	0.0000	0.7358	0.8737
vocab_s	~1		6.5746	0.3303	19.9054	0.0000	5.6056	7.8193
vocab_i	~~	vocab_s	-0.3137	0.0509	-6.1684	0.0000	-0.4130	-0.1990
vocab_R2	~~	vocab_R2	0.4421	0.0238	18.5449	0.0000	0.4001	0.4849

lhs	op	rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_R3	~~	vocab_R3	0.5812	0.0146	39.9448	0.0000	0.5520	0.6146
vocab_R4	~~	vocab_R4	0.4707	0.0164	28.7490	0.0000	0.4246	0.5147
vocab_R5	~~	vocab_R5	0.4942	0.0234	21.1098	0.0000	0.4277	0.5597
vocab_R3	~	vocab_R2	-0.0273	0.0208	-1.3130	0.1892	-0.0692	0.0171
vocab_R4	~	vocab_R3	0.0894	0.0179	5.0036	0.0000	0.0552	0.1257
vocab_R5	~	vocab_R4	-0.1687	0.0156	-10.8327	0.0000	-0.1987	-0.1378
wi_r12	~	momedu	0.6350	0.0137	46.2274	0.0000	0.6026	0.6600
stunt_r12	~	wi_r12	-0.3117	0.0206	-15.1504	0.0000	-0.3492	-0.2686
pre_school_var	~	wi_r12	0.3654	0.0258	14.1420	0.0000	0.3065	0.4167
pre_school_var	~	momedu	0.0806	0.0273	2.9565	0.0031	0.0286	0.1357
egra_rasch	~	vocab_i	0.5293	0.0235	22.4827	0.0000	0.4827	0.5779
egra_rasch	~	pre_school_var	0.0770	0.0241	3.1949	0.0014	0.0241	0.1241
schoolstudyhrs_r345	~	wi_r12	0.4973	0.0174	28.5939	0.0000	0.4621	0.5330
vocab_i	~	wi_r12	0.4286	0.0270	15.8515	0.0000	0.3789	0.4829
vocab_i	~	Agemon	0.1907	0.0202	9.4190	0.0000	0.1487	0.2355
vocab_i	~	Momedu	0.2490	0.0249	10.0020	0.0000	0.2025	0.2957
vocab_i	~	stunt_r12	-0.1731	0.0202	-8.5654	0.0000	-0.2115	-0.1346
vocab_i	~	pre_school_var	0.0915	0.0220	4.1596	0.0000	0.0432	0.1366
vocab_s	~	wi_r12	-0.2675	0.0370	-7.2273	0.0000	-0.3454	-0.1885
vocab_s	~	schoolstudyhrs_r345	0.5061	0.0308	16.4380	0.0000	0.4022	0.5987
read_comp_r5	~	vocab_s	0.3428	0.0317	10.8058	0.0000	0.2758	0.4022
read_comp_r5	~	vocab_i	0.4822	0.0286	16.8513	0.0000	0.4240	0.5414
read_comp_r5	~	egra_rasch	0.1125	0.0258	4.3567	0.0000	0.0580	0.1581
wi_r12	~1		1.6388	0.0484	33.8483	0.0000	1.5469	1.7468
Momedu	~1		1.2810	0.0308	41.6455	0.0000	1.2299	1.3296
pre_school_var	~1		3.6843	0.0983	37.4904	0.0000	3.4349	4.0032
egra_rasch	~1		18.6164	0.3512	53.0112	0.0000	17.8823	19.3736
read_comp_r5	~1		-3.3768	0.5708	-5.9163	0.0000	-4.4056	-2.2526
stunt_r12	~1		1.3395	0.0529	25.3083	0.0000	1.2276	1.4344
schoolstudyhrs_r345	~1		3.4264	0.0962	35.6281	0.0000	3.1077	3.7664
Agemon	~1		17.6428	0.2834	62.2464	0.0000	17.1113	18.1585
wi_r12	~~	wi_r12	0.5968	0.0174	34.2127	0.0000	0.5644	0.6369
stunt_r12	~~	stunt_r12	0.9028	0.0128	70.3706	0.0000	0.8780	0.9278
pre_school_var	~~	pre_school_var	0.8226	0.0157	52.2455	0.0000	0.7936	0.8549
egra_rasch	~~	egra_rasch	0.6837	0.0220	31.0911	0.0000	0.6371	0.7292
schoolstudyhrs_r345	~~	schoolstudyhrs_r345	0.7527	0.0173	43.5089	0.0000	0.7160	0.7865
read_comp_r5	~~	read_comp_r5	0.6495	0.0227	28.6635	0.0000	0.6028	0.6956
momedu	~~	Momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
Agemon	~~	Agemon	1.0000	0.0000	NA	NA	1.0000	1.0000
vocab_R2	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: An auto-regressive structure (lag 1) was used to represent time-dependent correlation between the vocabulary scores at Rounds 2 to 5. 95% confidence intervals around standardised coefficients were computed through bootstrapping (1000 resamples).

Appendix A4. Modeling results for conditional growth of receptive vocabulary in Peru

Overall model fit statistics

lavaan (0.5-23.1097) converged normally after 167 iterations	
Number of observations	2052
Number of missing patterns	114
Estimator	ML
Minimum Function Test Statistic	1107.578
Degrees of freedom	92
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	10267.154
Degrees of freedom	120
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.900
Tucker-Lewis Index (TLI)	0.869
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-21838.030
Loglikelihood unrestricted model (H1)	-21284.241
Number of free parameters	60
Akaike (AIC)	43796.060
Bayesian (BIC)	44133.654
Sample-size adjusted Bayesian (BIC)	43943.029
Root Mean Square Error of Approximation:	
RMSEA	0.073
90 per cent Confidence Interval	0.070 0.077
P-value RMSEA <= 0.05	0.000
Standardised Root Mean Square Residual:	
SRMR	0.067

Fitted coefficients/parameter estimates

lhs	op	rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_i	=~	vocab_R2	0.7883	0.0120	65.8779	0.0000	0.7612	0.8130
vocab_i	=~	vocab_R3	0.7598	0.0158	48.1544	0.0000	0.7250	0.7991
vocab_i	=~	vocab_R4	0.8207	0.0189	43.4449	0.0000	0.7775	0.8647
vocab_i	=~	vocab_R5	0.7997	0.0183	43.6975	0.0000	0.7558	0.8469
vocab_i	~~	vocab_i	0.3309	0.0194	17.0503	0.0000	0.2929	0.3669
vocab_i	~1		0.0825	0.0991	0.8332	0.4047	-0.1142	0.2871
vocab_s	=~	vocab_R2	0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_s	=~	vocab_R3	0.1617	0.0101	15.9486	0.0000	0.1367	0.1864
vocab_s	=~	vocab_R4	0.4076	0.0242	16.8230	0.0000	0.3503	0.4607
vocab_s	=~	vocab_R5	0.5674	0.0318	17.8245	0.0000	0.4870	0.6414
vocab_s	~~	vocab_s	0.8867	0.0228	38.8503	0.0000	0.8169	0.9326
vocab_s	~1		7.7632	0.5142	15.0983	0.0000	6.8041	9.0567
vocab_i	~~	vocab_s	0.4251	0.0854	4.9785	0.0000	0.2473	0.6351
vocab_R2	~~	vocab_R2	0.3785	0.0189	20.0611	0.0000	0.3391	0.4206
vocab_R3	~~	vocab_R3	0.2551	0.0112	22.7166	0.0000	0.2333	0.2787
vocab_R4	~~	vocab_R4	0.1874	0.0103	18.2373	0.0000	0.1497	0.2281
vocab_R5	~~	vocab_R5	0.2428	0.0140	17.3729	0.0000	0.1869	0.3019
vocab_R3	~	vocab_R2	0.0432	0.0161	2.6842	0.0073	0.0049	0.0771

lhs	op	rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
vocab_R4	~	vocab_R3	-0.1614	0.0214	-7.5253	0.0000	-0.2115	-0.1116
vocab_R5	~	vocab_R4	-0.2995	0.0180	-16.6647	0.0000	-0.3396	-0.2596
wi_r12	~	momedu	0.6546	0.0134	48.8513	0.0000	0.6278	0.6791
mother_mhealth	~	wi_r12	-0.0642	0.0227	-2.8300	0.0047	-0.1047	-0.0200
stunt_r12	~	wi_r12	-0.4026	0.0195	-20.6380	0.0000	-0.4403	-0.3662
stunt_r12	~	mother_mhealth	0.0525	0.0212	2.4826	0.0130	0.0116	0.0954
pre_school_var	~	wi_r12	0.3725	0.0263	14.1505	0.0000	0.3234	0.4214
pre_school_var	~	momedu	0.1620	0.0276	5.8626	0.0000	0.1101	0.2087
house_chore_r345	~	momedu	-0.2993	0.0221	-13.5657	0.0000	-0.3423	-0.2560
schoolstudyhrs_r345	~	wi_r12	0.2716	0.0237	11.4437	0.0000	0.2234	0.3187
schoolstudyhrs_r345	~	house_chore_r345	-0.0976	0.0235	-4.1495	0.0000	-0.1582	-0.0397
egra_rasch	~	vocab_i	0.6153	0.0168	36.5543	0.0000	0.5796	0.6479
early_subs	~	wi_r12	0.1679	0.0236	7.0978	0.0000	0.1211	0.2136
vocab_i	~	wi_r12	0.4593	0.0249	18.4701	0.0000	0.4067	0.5056
vocab_i	~	momedu	0.2890	0.0248	11.6688	0.0000	0.2389	0.3393
vocab_i	~	pre_school_var	0.1472	0.0213	6.9223	0.0000	0.1061	0.1890
vocab_i	~	stunt_r12	-0.1143	0.0210	-5.4305	0.0000	-0.1619	-0.0660
vocab_s	~	stunt_r12	-0.1255	0.0376	-3.3400	0.0008	-0.2068	-0.0504
vocab_s	~	schoolstudyhrs_r345	0.2673	0.0361	7.4088	0.0000	0.1775	0.3578
vocab_s	~	early_crime	-0.1348	0.0360	-3.7453	0.0002	-0.2218	-0.0543
read_comp_r5	~	vocab_s	0.3786	0.0293	12.9271	0.0000	0.3170	0.4421
read_comp_r5	~	egra_rasch	0.1519	0.0260	5.8499	0.0000	0.0956	0.2072
read_comp_r5	~	momedu	0.0897	0.0321	2.7977	0.0051	0.0281	0.1611
read_comp_r5	~	vocab_i	0.2824	0.0460	6.1451	0.0000	0.1846	0.3777
wi_r12	~1		0.8493	0.0456	18.6298	0.0000	0.7741	0.9266
momedu	~1		1.7317	0.0361	48.0133	0.0000	1.6801	1.7921
pre_school_var	~1		3.5308	0.0947	37.2682	0.0000	3.2728	3.8087
egra_rasch	~1		18.3930	0.3535	52.0356	0.0000	17.7291	19.2138
read_comp_r5	~1		-3.0300	0.5511	-5.4981	0.0000	-4.2346	-1.8933
stunt_r12	~1		1.4208	0.0531	26.7348	0.0000	1.3218	1.5259
mother_mhealth	~1		1.4707	0.0535	27.5052	0.0000	1.3744	1.5672
house_chore_r345	~1		2.6906	0.0519	51.8461	0.0000	2.5737	2.8061
schoolstudyhrs_r345	~1		7.0485	0.1531	46.0492	0.0000	6.5210	7.6427
early_subs	~1		2.2539	0.0720	31.2948	0.0000	2.1364	2.3886
early_crime	~1		1.4608	0.0347	42.1026	0.0000	1.3956	1.5372
mental_well_r5	~1		4.3481	0.0802	54.2300	0.0000	4.2025	4.4931
wi_r12	~~	wi_r12	0.5715	0.0175	32.5715	0.0000	0.5388	0.6058
mother_mhealth	~~	mother_mhealth	0.9959	0.0029	342.2207	0.0000	0.9890	0.9996
stunt_r12	~~	stunt_r12	0.8325	0.0159	52.4478	0.0000	0.7998	0.8603
pre_school_var	~~	pre_school_var	0.7560	0.0178	42.4930	0.0000	0.7237	0.7875
house_chore_r345	~~	house_chore_r345	0.9104	0.0132	68.9502	0.0000	0.8829	0.9345
schoolstudyhrs_r345	~~	schoolstudyhrs_r345	0.9063	0.0132	68.4968	0.0000	0.8767	0.9302
egra_rasch	~~	egra_rasch	0.6214	0.0207	30.0009	0.0000	0.5802	0.6640
early_subs	~~	early_subs	0.9718	0.0079	122.4149	0.0000	0.9544	0.9853
read_comp_r5	~~	read_comp_r5	0.5438	0.0198	27.5205	0.0000	0.5042	0.5850
momedu	~~	momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
early_crime	~~	early_crime	1.0000	0.0000	NA	NA	1.0000	1.0000
mental_well_r5	~~	mental_well_r5	1.0000	0.0000	NA	NA	1.0000	1.0000
vocab_R2	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
vocab_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: An auto-regressive structure (lag 1) was used to represent time-dependent correlation between the vocabulary scores at Rounds 2 to 5. 95% confidence intervals around standardised coefficients were computed through bootstrapping (1,000 resamples).

Appendix A5. Modeling results for conditional growth of mathematics in Ethiopia

Overall model fit statistics

lavaan (0.5-23.1097) converged normally after 122 iterations	
Number of observations	1999
Number of missing patterns	48
Estimator	ML
Minimum Function Test Statistic	474.039
Degrees of freedom	35
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	6771.933
Degrees of freedom	55
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.935
Tucker-Lewis Index (TLI)	0.897
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-10601.696
Loglikelihood unrestricted model (H1)	-10364.676
Number of free parameters	42
Akaike (AIC)	21287.392
Bayesian (BIC)	21522.609
Sample-size adjusted Bayesian (BIC)	21389.173
Root Mean Square Error of Approximation:	
RMSEA	0.079
90 per cent Confidence Interval	0.073 0.086
P-value RMSEA <= 0.05	0.000
Standardised Root Mean Square Residual:	
SRMR	0.063

Fitted coefficients/parameter estimates

Lhs	op	Rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
maths_i	=~	maths_R3	0.8792	0.0200	43.8583	0.0000	0.8374	0.9147
maths_i	=~	maths_R4	0.8117	0.0175	46.3742	0.0000	0.7766	0.8481
maths_i	=~	maths_R5	0.7840	0.0224	34.9536	0.0000	0.7402	0.8277
maths_i	~~	maths_i	0.5680	0.0271	20.9602	0.0000	0.5077	0.6185
maths_i	~1		-2.7575	0.3118	-8.8429	0.0000	-3.4359	-2.1997
maths_s	=~	maths_R3	0.0000	0.0000	NA	NA	0.0000	0.0000
maths_s	=~	maths_R4	0.3776	0.0224	16.8392	0.0000	0.3252	0.4174
maths_s	=~	maths_R5	0.6383	0.0409	15.6240	0.0000	0.5495	0.7103
maths_s	~~	maths_s	0.9553	0.0144	66.2605	0.0000	0.9152	0.9790
maths_s	~1		-0.6472	0.1080	-5.9905	0.0000	-0.8797	-0.4301
maths_i	~~	maths_s	-0.4367	0.0525	-8.3184	0.0000	-0.5263	-0.3151

Lhs	op	Rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
maths_R3	~~	maths_R3	0.2271	0.0352	6.4429	0.0000	0.1634	0.2988
maths_R4	~~	maths_R4	0.3506	0.0137	25.5958	0.0000	0.3226	0.3788
maths_R5	~~	maths_R5	0.2261	0.0237	9.5552	0.0000	0.1797	0.2762
wi_r12	~	momedu	0.5986	0.0149	40.0488	0.0000	0.5657	0.6287
stunt_r12	~	wi_r12	-0.2593	0.0217	-11.9467	0.0000	-0.2962	-0.2200
pre_school_var	~	wi_r12	0.4590	0.0203	22.5772	0.0000	0.4144	0.5051
pre_school_var	~	momedu	0.2708	0.0221	12.2434	0.0000	0.2181	0.3246
schoolstudyhrs_r345	~	house_chore_r345	-0.1246	0.0190	-6.5473	0.0000	-0.1704	-0.0838
schoolstudyhrs_r345	~	pre_school_var	0.2301	0.0263	8.7592	0.0000	0.1846	0.2736
schoolstudyhrs_r345	~	wi_r12	0.3670	0.0243	15.1046	0.0000	0.3227	0.4099
house_chore_r345	~	wi_r12	-0.2392	0.0223	-10.7275	0.0000	-0.2775	-0.2019
cda_rscore	~	wi_r12	0.1840	0.0352	5.2305	0.0000	0.1081	0.2568
cda_rscore	~	pre_school_var	0.1379	0.0334	4.1322	0.0000	0.0729	0.2016
cda_rscore	~	momedu	0.1163	0.0321	3.6257	0.0003	0.0489	0.1830
maths_i	~	wi_r12	0.5246	0.0248	21.1279	0.0000	0.4670	0.5801
maths_i	~	stunt_r12	-0.0728	0.0193	-3.7756	0.0002	-0.1077	-0.0354
maths_i	~	cda_rscore	0.1144	0.0208	5.4996	0.0000	0.0726	0.1597
maths_i	~	momedu	0.0873	0.0232	3.7581	0.0002	0.0365	0.1416
maths_i	~	agemon	0.0767	0.0185	4.1346	0.0000	0.0396	0.1138
maths_s	~	schoolstudyhrs_r345	0.2114	0.0341	6.1997	0.0000	0.1450	0.2913
stunt_r12	~1		1.2510	0.0394	31.7391	0.0000	1.1865	1.3208
agemon	~1		16.3852	0.2660	61.6087	0.0000	16.0269	16.7460
momedu	~1		0.7097	0.0254	27.9029	0.0000	0.6781	0.7403
pre_school_var	~1		0.7835	0.0383	20.4765	0.0000	0.7257	0.8382
cda_rscore	~1		5.3874	0.1265	42.5838	0.0000	5.1258	5.6451
schoolstudyhrs_r345	~1		2.1751	0.0740	29.3992	0.0000	2.0257	2.3477
house_chore_r345	~1		1.9177	0.0442	43.3512	0.0000	1.8392	1.9906
wi_r12	~1		1.0530	0.0327	32.2504	0.0000	1.0062	1.1071
maths_R3	~~	schoolstudyhrs_r345	0.4074	0.0470	8.6670	0.0000	0.3254	0.5106
wi_r12	~~	wi_r12	0.6416	0.0179	35.8504	0.0000	0.6048	0.6800
stunt_r12	~~	stunt_r12	0.9327	0.0113	82.8483	0.0000	0.9123	0.9516
pre_school_var	~~	pre_school_var	0.5671	0.0169	33.5382	0.0000	0.5273	0.6058
schoolstudyhrs_r345	~~	schoolstudyhrs_r345	0.6615	0.0189	35.0676	0.0000	0.6233	0.6947
house_chore_r345	~~	house_chore_r345	0.9428	0.0107	88.3613	0.0000	0.9230	0.9592
cda_rscore	~~	cda_rscore	0.8589	0.0159	54.0325	0.0000	0.8226	0.8875
momedu	~~	momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
agemon	~~	agemon	1.0000	0.0000	NA	NA	1.0000	1.0000
maths_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
maths_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
maths_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: 95% confidence intervals around standardised coefficients were computed through bootstrapping (1,000 resamples). Error variances of maths at Round 3 and school study hours were allowed to correlate in order to reduce high model residuals.

Appendix A6. Modeling results for conditional growth of mathematics in India

Overall model fit statistics

lavaan (0.5-23.1097) converged normally after 134 iterations	
Number of observations	2011
Number of missing patterns	39
Estimator	ML
Minimum Function Test Statistic	291.766
Degrees of freedom	40
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	5498.115
Degrees of freedom	66
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.954
Tucker-Lewis Index (TLI)	0.924
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-17792.049
Loglikelihood unrestricted model (H1)	-17646.166
Number of free parameters	50
Akaike (AIC)	35684.099
Bayesian (BIC)	35964.418
Sample-size adjusted Bayesian (BIC)	35805.565
Root Mean Square Error of Approximation:	
RMSEA	0.056
90 per cent Confidence Interval	0.050 0.062
P-value RMSEA <= 0.05	0.050
Standardised Root Mean Square Residual:	
SRMR	0.042

Fitted coefficients/parameter estimates

lhs	op	rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
maths_i	=~	maths_R3	0.7988	0.0223	35.8203	0.0000	0.7524	0.8410
maths_i	=~	maths_R4	0.7877	0.0209	37.7181	0.0000	0.7471	0.8272
maths_i	=~	maths_R5	0.7780	0.0273	28.5216	0.0000	0.7239	0.8298
maths_i	~~	maths_i	0.6955	0.0273	25.4903	0.0000	0.6372	0.7415
maths_i	~1		-2.6736	0.3647	-7.3303	0.0000	-3.3832	-2.0040
maths_s	=~	maths_R3	0.0000	0.0000	NA	NA	0.0000	0.0000
maths_s	=~	maths_R4	0.3264	0.0328	9.9366	0.0000	0.2603	0.3939
maths_s	=~	maths_R5	0.5641	0.0593	9.5123	0.0000	0.4461	0.6793
maths_s	~~	maths_s	0.7857	0.0765	10.2653	0.0000	0.6014	0.8990
maths_s	~1		-1.9649	0.7505	-2.6183	0.0088	-3.3999	-0.3395
maths_i	~~	maths_s	-0.3622	0.0803	-4.5078	0.0000	-0.4874	-0.1632
maths_R3	~~	maths_R3	0.3619	0.0356	10.1584	0.0000	0.2927	0.4339
maths_R4	~~	maths_R4	0.3561	0.0138	25.7185	0.0000	0.3271	0.3840
maths_R5	~~	maths_R5	0.2508	0.0249	10.0760	0.0000	0.2036	0.2993

lhs	op	rhs	est.std	se	z	p-value	lower CI (.025)	upper CI (.975)
wi_r12	~	momedu	0.5932	0.0152	39.1325	0.0000	0.5603	0.6252
mother_mhealth	~	wi_r12	-0.2188	0.0220	-9.9297	0.0000	-0.2588	-0.1757
stunt_r12	~	wi_r12	-0.2391	0.0224	-10.6531	0.0000	-0.2817	-0.1960
stunt_r12	~	mother_mhealth	0.0553	0.0231	2.3979	0.0165	0.0062	0.1050
pre_school_var	~	wi_r12	0.3583	0.0243	14.7381	0.0000	0.3062	0.4104
pre_school_var	~	momedu	0.1708	0.0253	6.7445	0.0000	0.1185	0.2258
schoolstudyhrs_r345	~	wi_r12	0.1607	0.0246	6.5221	0.0000	0.1131	0.2095
schoolstudyhrs_r345	~	house_chore_r345	-0.3805	0.0188	-20.2650	0.0000	-0.4351	-0.3258
schoolstudyhrs_r345	~	momedu	0.1724	0.0249	6.9230	0.0000	0.1353	0.2143
house_chore_r345	~	wi_r12	-0.1617	0.0228	-7.1044	0.0000	-0.2039	-0.1211
cda_rscore	~	pre_school_var	0.0592	0.0243	2.4314	0.0150	0.0127	0.1020
cda_rscore	~	stunt_r12	-0.1314	0.0229	-5.7446	0.0000	-0.1742	-0.0861
cda_rscore	~	momedu	0.2344	0.0239	9.7977	0.0000	0.1845	0.2759
mathematics_i	~	wi_r12	0.1760	0.0274	6.4302	0.0000	0.1193	0.2285
mathematics_i	~	stunt_r12	-0.1756	0.0260	-6.7556	0.0000	-0.2260	-0.1221
mathematics_i	~	cda_rscore	0.1860	0.0226	8.2439	0.0000	0.1398	0.2299
mathematics_i	~	momedu	0.2623	0.0276	9.5007	0.0000	0.2075	0.3193
mathematics_i	~	agemon	0.0628	0.0208	3.0229	0.0025	0.0239	0.1010
mathematics_s	~	stunt_r12	0.1661	0.0411	4.0408	0.0001	0.0842	0.2603
mathematics_s	~	schoolstudyhrs_r345	0.3763	0.1225	3.0716	0.0021	0.1158	0.6018
mathematics_s	~	house_chore_r345	-0.1279	0.0706	-1.8104	0.0702	-0.2773	-0.0026
stunt_r12	~1		1.2300	0.0703	17.4959	0.0000	1.0891	1.3753
agemon	~1		17.3148	0.2853	60.6927	0.0000	16.9215	17.7370
momedu	~1		0.6908	0.0259	26.6795	0.0000	0.6605	0.7205
pre_school_var	~1		2.4978	0.0808	30.9085	0.0000	2.3722	2.6340
cda_rscore	~1		5.7744	0.1348	42.8381	0.0000	5.4901	6.0850
schoolstudyhrs_r345	~1		5.4227	0.1087	49.9033	0.0000	5.0588	5.8391
house_chore_r345	~1		1.9501	0.0601	32.4692	0.0000	1.8231	2.0883
mother_mhealth	~1		1.8950	0.0566	33.4549	0.0000	1.7931	1.9950
wi_r12	~1		1.8962	0.0450	42.1781	0.0000	1.8257	1.9727
mathematics_R3	~~	schoolstudyhrs_r345	0.3209	0.0399	8.0400	0.0000	0.2439	0.4051
mathematics_R4	~~	schoolstudyhrs_r345	0.3373	0.0715	4.7208	0.0000	0.1861	0.4796
mathematics_R5	~~	schoolstudyhrs_r345	0.2198	0.1389	1.5829	0.1135	-0.0506	0.4901
wi_r12	~~	wi_r12	0.6481	0.0180	36.0425	0.0000	0.6092	0.6861
mother_mhealth	~~	mother_mhealth	0.9521	0.0096	98.7380	0.0000	0.9330	0.9691
stunt_r12	~~	stunt_r12	0.9340	0.0112	83.4218	0.0000	0.9120	0.9524
pre_school_var	~~	pre_school_var	0.7699	0.0171	44.9582	0.0000	0.7283	0.8046
schoolstudyhrs_r345	~~	schoolstudyhrs_r345	0.7344	0.0173	42.3537	0.0000	0.6922	0.7731
house_chore_r345	~~	house_chore_r345	0.9739	0.0074	132.3359	0.0000	0.9584	0.9853
cda_rscore	~~	cda_rscore	0.9027	0.0132	68.2499	0.0000	0.8751	0.9260
momedu	~~	momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
agemon	~~	agemon	1.0000	0.0000	NA	NA	1.0000	1.0000
mathematics_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
mathematics_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
mathematics_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: Error terms for mathematics scores in Rounds 3 to 5 were allowed to correlate with the error terms for hours spend studying at school and at home, in order to reduce high model residuals. 95% confidence intervals around standardised coefficients were computed through bootstrapping (1,000 resamples).

Appendix A7. Modeling results for conditional growth of mathematics in Vietnam

Overall model fit statistics

lavaan (0.5-23.1097) converged normally after 94 iterations	
Number of observations	Used Total 1989 1999
Number of missing patterns	38
Estimator	ML
Minimum Function Test Statistic	177.037
Degrees of freedom	19
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	4324.990
Degrees of freedom	36
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.963
Tucker-Lewis Index (TLI)	0.930
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-6803.533
Loglikelihood unrestricted model (H1)	-6715.014
Number of free parameters	35
Akaike (AIC)	13677.065
Bayesian (BIC)	13872.904
Sample-size adjusted Bayesian (BIC)	13761.707
Root Mean Square Error of Approximation:	
RMSEA	0.065
90 per cent Confidence Interval	0.056 0.074
P-value RMSEA <= 0.05	0.003
Standardised Root Mean Square Residual:	
SRMR	0.038

Fitted coefficients/parameter estimates

lhs	op	rhs	est.std	se	z	p-value		
maths_i	=~	maths_R3	0.6288	0.0304	20.6974	0.0000	0.5608	0.6840
maths_i	=~	maths_R4	0.6252	0.0286	21.8612	0.0000	0.5611	0.6783
maths_i	=~	maths_R5	0.6163	0.0310	19.8931	0.0000	0.5477	0.6728
maths_i	~~	maths_i	0.3805	0.0599	6.3520	0.0000	0.2092	0.4776
maths_i	~1		-3.8765	0.5225	-7.4188	0.0000	-5.0404	-2.9017
maths_s	=~	maths_R3	0.0000	0.0000	NA	NA	0.0000	0.0000
maths_s	=~	maths_R4	0.2234	0.0490	4.5625	0.0000	0.0914	0.3055
maths_s	=~	maths_R5	0.3853	0.0864	4.4620	0.0000	0.1579	0.5284
maths_s	~~	maths_s	1.0000	0.0000	NA	NA	1.0000	1.0000
maths_s	~1		-0.0179	0.0607	-0.2958	0.7674	-0.1509	0.1321
maths_i	~~	maths_s	0.5279	0.4433	1.1908	0.2337	-0.0117	5.0389

lhs	op	rhs	est.std	se	z	p-value		
maths_R3	~~	maths_R3	0.5761	0.0383	15.0369	0.0000	0.5025	0.6591
maths_R4	~~	maths_R4	0.4682	0.0154	30.4639	0.0000	0.4380	0.4982
maths_R5	~~	maths_R5	0.3170	0.0287	11.0399	0.0000	0.2574	0.3742
wi_r12	~	momedu	0.6276	0.0140	44.8529	0.0000	0.5956	0.6566
stunt_r12	~	wi_r12	-0.3095	0.0204	-15.1394	0.0000	-0.3480	-0.2665
pre_school_var	~	wi_r12	0.3647	0.0257	14.1848	0.0000	0.3045	0.4131
pre_school_var	~	momedu	0.0689	0.0273	2.5210	0.0117	0.0154	0.1216
cda_rscore	~	pre_school_var	0.1511	0.0239	6.3259	0.0000	0.0975	0.1891
cda_rscore	~	momedu	0.1608	0.0238	6.7468	0.0000	0.1144	0.2057
cda_rscore	~	agemon	0.1627	0.0225	7.2217	0.0000	0.1207	0.2055
maths_i	~	wi_r12	0.4413	0.0361	12.2211	0.0000	0.3711	0.5251
maths_i	~	stunt_r12	-0.1876	0.0273	-6.8818	0.0000	-0.2443	-0.1343
maths_i	~	cda_rscore	0.1780	0.0271	6.5646	0.0000	0.1247	0.2336
maths_i	~	momedu	0.2650	0.0335	7.9194	0.0000	0.2056	0.3367
maths_i	~	agemon	0.0824	0.0289	2.8470	0.0044	0.0243	0.1436
stunt_r12	~1		1.3402	0.0531	25.2413	0.0000	1.2356	1.4369
agemon	~1		17.7163	0.2823	62.7568	0.0000	17.1924	18.2190
momedu	~1		1.2838	0.0307	41.7648	0.0000	1.2289	1.3366
pre_school_var	~1		3.7163	0.0980	37.9340	0.0000	3.4823	4.0427
cda_rscore	~1		2.1911	0.4078	5.3725	0.0000	1.4216	3.0634
wi_r12	~1		1.6687	0.0486	34.3296	0.0000	1.5784	1.7689
wi_r12	~~	agemon	0.1140	0.0223	5.1073	0.0000	0.0702	0.1549
pre_school_var	~~	agemon	0.1153	0.0219	5.2671	0.0000	0.0677	0.1653
maths_R3	~~	agemon	0.2679	0.0276	9.7142	0.0000	0.2071	0.3260
wi_r12	~~	wi_r12	0.6061	0.0176	34.5089	0.0000	0.5688	0.6453
stunt_r12	~~	stunt_r12	0.9042	0.0127	71.4630	0.0000	0.8789	0.9290
pre_school_var	~~	pre_school_var	0.8307	0.0154	53.9599	0.0000	0.8013	0.8615
cda_rscore	~~	cda_rscore	0.9036	0.0128	70.7717	0.0000	0.8780	0.9302
momedu	~~	momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
agemon	~~	agemon	1.0000	0.0000	NA	NA	1.0000	1.0000
maths_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
maths_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
maths_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: 95% confidence intervals around standardised coefficients were computed through bootstrapping (1,000 resamples). Error variances of maths at Round 3 and school study hours were allowed to correlate in order to reduce high model residuals.

Appendix A8. Modeling results for conditional growth of mathematics in Peru

Overall model fit statistics

lavaan (0.5-23.1097) converged normally after 102 iterations	
Number of observations	2052
Number of missing patterns	80
Estimator	ML
Minimum Function Test Statistic	584.592
Degrees of freedom	64
P-value (Chi-square)	0.000
Model test baseline model:	
Minimum Function Test Statistic	6387.174
Degrees of freedom	91
P-value	0.000
User model versus baseline model:	
Comparative Fit Index (CFI)	0.917
Tucker-Lewis Index (TLI)	0.882
Loglikelihood and Information Criteria:	
Loglikelihood user model (H0)	-18013.568
Loglikelihood unrestricted model (H1)	-17721.272
Number of free parameters	55
Akaike (AIC)	36137.136
Bayesian (BIC)	36446.597
Sample-size adjusted Bayesian (BIC)	36271.858
Root Mean Square Error of Approximation:	
RMSEA	0.063
90 per cent Confidence Interval	0.058 0.068
P-value RMSEA <= 0.05	0.000
Standardised Root Mean Square Residual:	
SRMR	0.056

Fitted coefficients/parameter estimates

lhs	op	rhs	est.std	se	z	pvalue	lower CI (.025)	upper CI (.975)
maths_i	=~	maths_R3	1.0197	0.0336	30.3429	0.0000	0.9470	1.0870
maths_i	=~	maths_R4	1.0175	0.0322	31.5676	0.0000	0.9432	1.0786
maths_i	=~	maths_R5	1.0372	0.0418	24.8056	0.0000	0.9502	1.1209
maths_i	~~	maths_i	0.6711	0.0306	21.9258	0.0000	0.6039	0.7246
maths_i	~1		-2.0254	0.1410	-14.3624	0.0000	-2.3272	-1.7246
maths_s	=~	maths_R3	0.3442	0.0212	16.2081	0.0000	0.3006	0.3850
maths_s	=~	maths_R4	0.6869	0.0377	18.2177	0.0000	0.6053	0.7600
maths_s	=~	maths_R5	1.0503	0.0639	16.4433	0.0000	0.9168	1.1771
maths_s	~~	maths_s	0.9371	0.0142	65.9816	0.0000	0.9016	0.9586
maths_s	~1		-0.6783	0.2072	-3.2737	0.0011	-1.0972	-0.2897
maths_i	~~	maths_s	-0.6478	0.0342	-18.9358	0.0000	-0.7053	-0.5623
wi_r12	~	momedu	0.6545	0.0134	48.6830	0.0000	0.6289	0.6799
mother_mhealth	~	wi_r12	-0.0676	0.0227	-2.9779	0.0029	-0.1114	-0.0265
stunt_r12	~	wi_r12	-0.3994	0.0195	-20.4541	0.0000	-0.4339	-0.3627
stunt_r12	~	mother_mhealth	0.0549	0.0212	2.5910	0.0096	0.0161	0.0982
pre_school_var	~	wi_r12	0.3700	0.0263	14.0538	0.0000	0.3166	0.4198

lhs	op	rhs	est.std	se	z	pvalue	lower CI (.025)	upper CI (.975)
pre_school_var	~	momedu	0.1632	0.0277	5.8861	0.0000	0.1142	0.2151
schoolstudyhrs_r345	~	wi_r12	0.2712	0.0238	11.4009	0.0000	0.2227	0.3174
schoolstudyhrs_r345	~	house_chore_r345	-0.0965	0.0234	-4.1195	0.0000	-0.1594	-0.0371
house_chore_r345	~	wi_r12	-0.1906	0.0301	-6.3442	0.0000	-0.2471	-0.1320
house_chore_r345	~	momedu	-0.1704	0.0305	-5.5920	0.0000	-0.2333	-0.1099
cda_rscore	~	pre_school_var	0.1190	0.0245	4.8609	0.0000	0.0728	0.1618
cda_rscore	~	wi_r12	0.3170	0.0234	13.5383	0.0000	0.2649	0.3646
early_subs	~	wi_r12	0.1647	0.0218	7.5399	0.0000	0.1220	0.2122
maths_i	~	wi_r12	0.2711	0.0308	8.8117	0.0000	0.2138	0.3262
maths_i	~	stunt_r12	-0.0765	0.0188	-4.0698	0.0000	-0.1146	-0.0373
maths_i	~	cda_rscore	0.1320	0.0186	7.0977	0.0000	0.0926	0.1758
maths_i	~	momedu	0.2112	0.0239	8.8254	0.0000	0.1694	0.2600
maths_i	~	pre_school_var	0.0719	0.0198	3.6376	0.0003	0.0322	0.1110
maths_s	~	wi_r12	-0.2065	0.0341	-6.0499	0.0000	-0.2767	-0.1437
maths_s	~	schoolstudyhrs_r345	0.1576	0.0227	6.9442	0.0000	0.1124	0.2072
maths_s	~	early_subs	-0.0613	0.0229	-2.6778	0.0074	-0.1098	-0.0145
maths_s	~	early_crime	-0.0548	0.0229	-2.3891	0.0169	-0.0985	-0.0101
maths_s	~	mental_well_r5	0.0485	0.0218	2.2205	0.0264	0.0068	0.0915
stunt_r12	~1		1.4130	0.0535	26.4299	0.0000	1.3113	1.5117
momedu	~1		1.7285	0.0361	47.8685	0.0000	1.6702	1.7853
pre_school_var	~1		3.5369	0.0947	37.3353	0.0000	3.2935	3.8272
cda_rscore	~1		5.0930	0.1504	33.8655	0.0000	4.8065	5.3939
wi_r12	~1		0.8621	0.0458	18.8110	0.0000	0.7837	0.9331
early_subs	~1		2.2620	0.0685	33.0185	0.0000	2.1389	2.3891
early_crime	~1		1.4605	0.0347	42.1497	0.0000	1.3924	1.5389
schoolstudyhrs_r345	~1		7.0215	0.1522	46.1366	0.0000	6.4608	7.6445
house_chore_r345	~1		2.8478	0.0552	51.6170	0.0000	2.7320	2.9742
mother_mhealth	~1		1.4782	0.0536	27.5548	0.0000	1.3805	1.5721
mental_well_r5	~1		4.3453	0.0802	54.1558	0.0000	4.2008	4.4914
early_subs	~~	early_crime	0.3987	0.0207	19.3057	0.0000	0.3477	0.4478
maths_R3	~~	maths_R3	0.2643	0.0270	9.7741	0.0000	0.2112	0.3191
maths_R4	~~	maths_R4	0.3343	0.0129	25.8983	0.0000	0.3103	0.3605
maths_R5	~~	maths_R5	0.1327	0.0281	4.7235	0.0000	0.0761	0.1851
wi_r12	~~	wi_r12	0.5717	0.0176	32.4886	0.0000	0.5378	0.6044
mother_mhealth	~~	mother_mhealth	0.9954	0.0031	324.5355	0.0000	0.9876	0.9993
stunt_r12	~~	stunt_r12	0.8345	0.0158	52.8539	0.0000	0.8059	0.8635
pre_school_var	~~	pre_school_var	0.7574	0.0178	42.6696	0.0000	0.7244	0.7853
schoolstudyhrs_r345	~~	schoolstudyhrs_r345	0.9013	0.0139	65.0722	0.0000	0.8686	0.9286
house_chore_r345	~~	house_chore_r345	0.8921	0.0141	63.2372	0.0000	0.8630	0.9146
cda_rscore	~~	cda_rscore	0.8493	0.0153	55.5211	0.0000	0.8183	0.8759
early_subs	~~	early_subs	0.9729	0.0072	135.2455	0.0000	0.9550	0.9851
momedu	~~	momedu	1.0000	0.0000	NA	NA	1.0000	1.0000
early_crime	~~	early_crime	1.0000	0.0000	NA	NA	1.0000	1.0000
mental_well_r5	~~	mental_well_r5	1.0000	0.0000	NA	NA	1.0000	1.0000
maths_R3	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
maths_R4	~1		0.0000	0.0000	NA	NA	0.0000	0.0000
maths_R5	~1		0.0000	0.0000	NA	NA	0.0000	0.0000

Model notes: 95% confidence intervals around standardised coefficients were computed through bootstrapping (1000 resamples). Error variances of substance use and adolescent participation in crime and violence were allowed to correlate in order to reduce high model residuals.

Appendix B. Early crime and violence and early substance use items (Peru only)

Early crime and violence
1: Have you ever been beaten up or physically hurt in other ways by the following people? (summed)
2: During the last 30 days, on how many days did you carry a weapon such as a knife, machete or gun to be able to protect yourself?
3: How many of your best friends have been/are members of a gang?
4: Have you ever been member of a gang?
5: Have you been arrested by the police or taken into custody for an illegal or delinquent offense?
6: How many partners have you EVER had intercourse with?
Early substance use
1: How old were you when you tried a cigarette for the first time?
2: How often do you usually drink alcohol?
3: How many of your best friends drink alcohol at least once a month?

Notes: Items are available in the Peru Round 5 Younger Cohort self-administered questionnaire. They are based (with some modifications) on commonly used youth risk behaviour scales, including: the Center for Disease Control (CDC) YRBSS (1999-2017) (www.cdc.gov/healthyyouth/data/yrbs/questionnaires.htm); and the Social and Health Assessment (SAHA) scales) (Schwab-Stone et al. 1995).

Predictors of Mathematics and Literacy Skills at 15 Years Old in Ethiopia, India, Peru and Vietnam: A Longitudinal Analysis

This working paper reports on longitudinal analyses of predictors of growth in receptive vocabulary and mathematics abilities by age 15 conducted using data from the Younger Cohort in the four Young Lives countries. Theoretical models of plausible influences on these outcomes were constructed. Latent Growth Modelling undertaken on these two key aspects of development provides unusually powerful cross-country evidence that household economic well-being and proximal relationships in the early childhood environment are particularly important, either placing young children at risk for developmental hazards such as growth stunting or conferring advantage for future development.

The models indicate that children whose mothers are rendered psychologically vulnerable by the stresses of poverty are at risk of poor physical growth in the early years, and that this impacts language, cognition and numeracy skills across childhood. Early disadvantages in poorer children are compounded by challenges of combining household responsibilities with schooling from middle childhood through adolescence. Those children with better-educated mothers are more likely to attend preschool programmes, which contribute to improved learning outcomes. While not directly measured, it is likely that those who made most gains attended programmes that were better provisioned. While there is variation across the countries, the high degree of similarity suggests good external validity of the general findings.

Young Lives

An International Study of Childhood Poverty

About Young Lives

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

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

Young Lives Partners

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

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

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

Young Lives