RISE Working Paper 21/082
November 2021

Socio-Emotional and Academic Learning Before and After COVID-19 School Closures: Evidence from Ethiopia

Stephen Bayley, Darge Wole Meshesha, Paul Ramchandani, Pauline Rose, Tassew Woldehanna, and Louise Yorke

Abstract

This paper presents the findings of research undertaken in Ethiopia to examine the effects of COVID-19 school closures on children's holistic learning, including both socio-emotional and academic learning. It draws on data collected in 2019 (prior to the pandemic) and 2021 (after schools re-opened) to compare primary pupils' learning before and after the school closures. In particular, the study adapts self-reporting scales that have been used in related contexts to measure Grade 3 and 6 children's social skills, self-efficacy, emotional regulation and mental health and wellbeing, along with literacy and numeracy. Lesson observations were also undertaken to explore teachers' behaviours to foster socio-emotional learning (SEL) in the classroom.

The findings advance current knowledge in several respects. First, they quantify the decline in Ethiopian pupils' social skills over the period of the school closures. Second, they identify a significant and strong relationship between learners' social skills and their numeracy, even after taking other factors into account. Third, they reveal a significant association between children's social skills and their mental health and wellbeing, highlighting the importance of interpersonal interactions to safeguard children's holistic welfare. The paper concludes by proposing a model for understanding the relationship between learners' SEL and academic outcomes, and with recommendations for education planning and practice, in Ethiopia and elsewhere.

Socio-Emotional and Academic Learning Before and After COVID-19 School Closures: Evidence from Ethiopia

Stephen Bayley
University of Cambridge

Darge Wole Meshesha
Addis Ababa University

Paul Ramchandani
University of Cambridge

Pauline Rose
REAL Centre, University of Cambridge

Tassew Woldehanna
Addis Ababa University

Louise Yorke
REAL Centre, University of Cambridge

Acknowledgements:

This research is funded by the LEGO Foundation and Research for Improving Systems of Education (RISE) programme. We would like to thank members of the RISE Ethiopia team at Addis Ababa University for feedback and advice during the study, including Dr Belay Hagos, Professor Amare Asegdom, Professor Tirussew Tefera and Professor Girma Lemma. Further thanks go to Ricardo Sabates, Dawit Tibebu Tiruneh, Janice Kim, Mesele Arraya and Chanie Ejigu. We are also grateful to Eve Hadshar, Celia Hsiao, Paul Frisoli, Garrett Jaeger and Margaret Laurie for their helpful input into this paper, as well as members of the RISE Directorate for their feedback. Finally, we would like to thank the schools, teachers and pupils who participated in the research.

This is one of a series of working papers from "RISE"—the large-scale education systems research programme supported by funding from the United Kingdom's Foreign, Commonwealth and Development Office (FCDO), the Australian Government's Department of Foreign Affairs and Trade (DFAT), and the Bill and Melinda Gates Foundation. The Programme is managed and implemented through a partnership between Oxford Policy Management and the Blavatnik School of Government at the University of Oxford.

Please cite this paper as:
Bayley, S., Wole, D., Ramchandani, P., Rose, P., Woldehanna, T. Yorke, L. 2021. Paper Socio-Emotional and Academic Learning Before and After COVID-19 School Closures: Evidence from Ethiopia. RISE Working Paper Series. 21/082. https://doi.org/10.35489/BSG-RISE-WP_2021/082

Socio-Emotional and Academic Learning Before and After COVID-19 School Closures: Evidence from Ethiopia

Stephen Bayley, Darge Wole Meshesha, Paul Ramchandani, Pauline Rose, Tassew Woldehanna and Louise Yorke

Contents

1. Introduction

The prolonged closure of schools since 2020 in response to COVID-19 has prompted major international concerns about the impact of the pandemic on children's learning, particularly for the most marginalised. As a result, global research efforts have been initiated to quantify the extent of learning loss, not least to assist national education systems to support *all* pupils as schools resume. Many of these studies focus largely on children's academic learning particularly with respect to literacy and numeracy. This overlooks the effects on children's socio-emotional learning (SEL) which is vital for understanding their holistic development and longer-term life outcomes.

This paper outlines the findings of research undertaken in Ethiopia to address that gap. Specifically, it describes a 2020-2021 study conducted through the Research on Improving Systems of Education (RISE) programme with support from the LEGO Foundation. The study builds on previous rounds of RISE data collection including assessments of students' academic learning and Grade 4 pupils' SEL between May and June 2019 (Yorke, Wole & Rose, 2021a). The current research follows the same children between January and March 2021. These children are now largely aged 9 to 13, and enrolled in Grades 3 and 6. This study extends the measures of SEL to encompass social skills, emotional regulation and self-efficacy, and also includes mental health and wellbeing in addition to academic outcomes like numeracy and literacy. The study thereby highlights the value of a wider conceptualisation of learning, not least in light of the ongoing COVID-19 pandemic (Bayley et al., 2021; Yorke et al., 2021b). It further draws on close and continuing collaboration between the Institute of Education Research, Addis Ababa University, the Policy Studies Institute, Addis Ababa and the REAL Centre, University of Cambridge.

The paper is structured as follows: section 2 outlines the context and maps the literature relating to children's SEL, mental health and wellbeing in low- and middle-income countries, including during COVID-19 as background to this study; section 3 describes the methodology including how the scales used were validated; section 4 outlines the key results; sections 5 and 6 discuss the findings and their limitations respectively; and section 7 concludes the paper with recommendations for future research and education planning going forward.

2. Context and Literature Review

This section provides a brief overview[i] of the Ethiopian context and key literature regarding the conceptualisation of SEL, mental health and wellbeing, and the related evidence from studies in low- and lower-middle-income countries. It builds on our earlier papers that discussed the importance of SEL (Yorke et al., 2021b) and the main approaches used to measure SEL in such contexts (Bayley et al., 2021). Together, they provide the basis for the research and analyses described in this paper, including the value of engaging with local specialists and stakeholders to contextualise the inquiry.

2.1 Ethiopian Context

The Federal Democratic Republic of Ethiopia comprises a large, diverse and landlocked country situated in the Horn of Africa. After Nigeria, Ethiopia is the most populous African nation with an estimated 112 million people in 2019, around 80 percent of whom live in rural areas, and who make up more than 90 different linguistic and ethnic groups (Federal Ministry of Education, 2015; World Bank, 2021). The country is organised through ten regional states and two city administrations, one of which contains the capital, Addis Ababa. The government places a strong emphasis on rapid economic growth as a key pillar of its development strategy, however, for now the human development indicators remain low, with Ethiopia ranking 173rd out of 189 rated countries (United Nations Development Programme, 2020). Certain regions and areas have also been affected by ongoing conflict and violence which further undermine efforts at national development.

For many years, the diversity and dispersal of communities across rural Ethiopia posed particular problems for delivering education (Federal Ministry of Education, 2015). However, the country has seen considerable investment since 2008 to increase access, quality and most recently equity, not least through initiatives like the General Education Quality Improvement Project (GEQIP). Currently in its third iteration, GEQIP sought first to strengthen education institutions, teaching and learning in primary schools, then added support for information and technology before including a stronger focus to reach disadvantaged learners, especially girls, pupils with special needs and children from pastoralist communities (Hoddinott et al., 2019). The government has also added an O-class to accommodate an extra year of pre-primary schooling before kindergarten (Borisova et al., 2017).

Regarding skills and SEL, the Education and Training Policy (ETP) outlines a wide range of intended learning outcomes, including the knowledge, abilities and attitudes to produce responsible citizens and meet the "country's need for skilled manpower" (Federal Democratic Republic Government of Ethiopia, 1994, p. 4). However, to date, most efforts to define and measure learning quality have been framed via the assessment of more conventional indicators like literacy and numeracy. Guidelines also specify pre-primary classrooms as the key space in which to foster children's SEL (Federal Ministry of Education, 2020b; Teferra et al., 2018). Together these raise questions regarding learners' SEL during the primary grades, including how pupils' skill development might have been impacted by COVID-19.

2.2 Key Literature

Since 2020, an estimated 1.6 billion students globally including 26 million learners in Ethiopia have missed out on valuable opportunities to pursue their education because of school closures due to the COVID-19 pandemic (UNESCO, n.d.; Yorke et al., 2020, 2021b). In addition to losses in literacy and numeracy, these closures have denied learners access to physical, emotional and psychological support through feeding programmes, safe and protective spaces, and to key SEL to build their skills for coping with the new levels of pressure, stress and anxiety (UNESCO, 2020). Recent evidence from the Young Lives initiative, including from Ethiopia, further indicates downturns in students' mental health and wellbeing, not least given the isolation of social distancing, the economic shock and instability of lockdowns, and the trauma of illness itself (Favara et al. 2021).

At its core, SEL comprises the process of acquiring a wide set of personal traits, skills and attributes. These competencies hold intrinsic value in and of themselves, but also play a key role in shaping other abilities and aptitudes. Particular understandings of SEL, especially the skills and qualities promoted in certain settings, may also vary according to cultural, political and economic factors, including who gets to set the learning priorities and agenda (Jukes et al., 2018).

Numerous frameworks exist for conceptualising, describing and operationalising SEL. Some of them use different terms such as 'non-cognitive skills', 'psychosocial skills' or '21st century learning' to emphasise particular aspects of learners' development (Duckworth & Yeager, 2015).[ii] Across frameworks, however, common themes regarding SEL concern individuals building int*er*personal skills to work with other people, int*ra*personal abilities to manage their own emotions and behaviours, and various broader competencies to achieve their goals, drawing on cognitive processes like problem solving and responsible decision making (Collaborative for Academic, Social, and Emotional Learning, 2020; Jones & Doolittle, 2017; Zins & Elias, 2007).

Literature on the benefits of SEL describes its importance for individuals in both academic and non-academic domains, as well as for more collective progress towards the Sustainable Development Goals (SDGs) (Organisation for Economic Co-operation and Development (OECD), 2021; Raikes et al., 2017). Indeed, international discourse around SEL has been used to explore and discuss the more holistic advantages of education beyond the traditional focus on literacy and numeracy, and thereby expand the meaning of 'learning' (Yorke et al., 2021b). Different aspects of schooling, for example, peer-to-peer interactions and emotional support, may enable and enhance children's self-efficacy, their

belief in their own ability to achieve a particular goal, complete certain tasks or succeed in specific situations (Bandura, 1977).

There are also important links between SEL and learners' mental health and personal wellbeing. Conceptions of mental health vary with culture and context but in general terms it is widely understood as comprising a state of wellbeing in which individuals are able to realise their abilities, cope with the stresses of everyday life, work productively and contribute to their communities (World Health Organization (WHO), 2018). It can be affected by a combination of social, biological and psychological factors, which may be positive, in the case of targeted community interventions and psychosocial support in schools, or negative, for example where prolonged exposure to toxic stress harms brain development and increases the risk of future illness (National Scientific Council on the Developing Child, 2014; Yorke et al., 2021b). The relationship between mental health and SEL also appears to be bidirectional, with SEL providing coping skills for improved wellbeing and emotional resilience, and poor mental health affecting individuals' disposition and capacity for learning, including their SEL (Diamond, 2014; Immordino-Yang & Damasio, 2007; UNESCO, 2020).

To date, the vast majority of literature concerning children's SEL, mental health and wellbeing in the context of formal education has drawn on research focused on learners living in high-income countries, particularly the USA. Even before COVID-19, the evidence from low- and lower-middle-income countries was piecemeal and sporadic (Inter-Agency Network for Education in Emergencies, 2016). This reflects in part certain complexities around the measurement of SEL, mental health and wellbeing in diverse contexts, for example the need to ensure cultural suitability and equivalence when adapting established tools for use in a new setting (Bayley et al., 2021).

However, there is now a growing body of related studies regarding children living in low- and lower-middle-income countries, many of which tend to centre on several key and inter-related themes. First, there is increasing international research concerning young learners' SEL as part of their broader holistic development. Some studies have explored interactions between early development including SEL and stunting, or the experiences of parents and teachers in fostering young children's SEL (Goldschmidt & Pedro, 2019; Tran et al., 2019; Watts et al., 2016). Global tools like the International Development and Early Learning Assessment (IDELA) and Measuring Early Learning Quality and Outcomes (MELQO) are also being used to capture data regarding young children's SEL and school readiness over time and at scale (Raikes et al., 2019; Wolf et al., 2017). For example, longitudinal research by Wolf and McCoy (2019) using IDELA examined bidirectional associations between Ghanaian children's SEL and early academic skills. They found that literacy and numeracy significantly predicted pupils' socio-emotional competences, but the reverse was not also true. They suggested that higher attainment in skills like reading, mathematics and problem solving could increase children's self-confidence, -esteem and -efficacy and encourage them to engage in more prosocial behaviours like sharing and cooperating.

Second, there is increasing research regarding SEL, mental health and wellbeing among learners in low- and lower-middle income countries affected by conflict, trauma or violence. These include psychosocial data captured from children and adolescents using measures like the Student Learning in Emergencies Checklist (SLEC-26) created for use in Palestine and the Holistic Assessment for Learning used in Syria (Forsberg et al., 2019; Tubbs Dolan, 2019). Such evidence can then be used to understand how the conflict affects learners' development and welfare, and to inform emergency responses and design appropriate support packages.

Third, measures for children's SEL, mental health and wellbeing have been used to generate data to evaluate the efficacy and impact of various educational programmes and interventions. Examples include research in India to examine the effect of an educational television series on 3-7-year olds' socio-emotional strategies, and to evaluate a randomised control trial to improve SEL through a parenting programme in Rwanda (Abimpaye et al., 2019; Borzekowski et al., 2019). Borisova and

colleagues (2017) also used IDELA to assess the impact of two early language and literacy interventions in Ethiopia and found no significant difference in SEL gains between the family-based and O-class models.

Finally, the Young Lives initiative has captured rich and valuable data since 2002 from several low- and lower-middle-income countries, including Ethiopia.[iii] Studies have used both quantitative and qualitative methods to capture information on learners' academic outcomes, wellbeing and life experiences to investigate diverse issues among children and youths (Camfield, 2012; Camfield & Tafere, 2009; Ford et al., 2021; Himaz, 2009). These have included the relationship between poverty and SEL and the emergence of self-efficacy and agency as learners progress through adolescence (Dercon & Krisnan, 2009; Espinoza & Ogando, 2019). Most recently Porter and colleagues (2021) have identified increased rates of anxiety and depression in light of COVID-19, as well as protective factors to mitigate their effects.

However, despite the growing body of data from low- and lower-middle-income countries, there remain important knowledge gaps around children's SEL, mental health and wellbeing. In particular, there are questions regarding primary-aged learners' holistic learning, especially in light of the recent school closures.

3. Methodology

This study investigates the effects of COVID-19 on children's holistic learning. In order to do so, it supplements data collected on academic learning and SEL by the RISE programme in 2019 with further data on children's learning in 2021. Regarding the former, literacy and numeracy assessments were conducted in 2019 with pupils typically aged 7 to 11 in Grades 1 and 4 across six regions in Ethiopia. Data on SEL were also captured from approximately 4,000 Grade 4 learners in 168 government primary schools between May and June 2019. In each case, the pupils were invited to rate themselves using established scales designed to measure social skills, student effort, family support for learning and teacher-student relationships.[iv] Specifically, they selected responses from three options to indicate whether they agreed, disagreed or partially agreed with statements such as '*If I hurt someone, I say sorry*' and '*I try to learn from my mistakes*'. Table 1 sets out the measures used and for further details of the 2019 SEL research, see Yorke et al. (2021a).

Table 1 – Scales used with Grade 4 Pupils in 2019 Data Collection

Aspect of SEL	Scale Name	Items	Reference
Social Skills	Children's Self Report Social Skills Scale (CS4)	7	Danielson & Phelps (2003)
Student Effort	Academic Self-Concept Questionnaire (ASCQ)	5	Liu, Wang & Parkin (2005)
Family Support for Learning	School Engagement Instrument (SEI)	5	Appleton, Christenson, Kim & Reschly (2006)
Teacher-Student Relationship		6	

3.1 Research Questions

The present study extends the scope and focus of the earlier research through tracking the same children in 2021. In most cases, the Grade 1 and 4 pupils in 2019 had progressed into Grades 3 and 6 respectively and were aged 9 to 13 at the time of the surveys. Data on children's SEL were captured from both cohorts to explore changes in skills over time within the older group, and the relationship between SEL and academic learning for both age cohorts. Specifically, we sought to answer the following research questions (RQs):

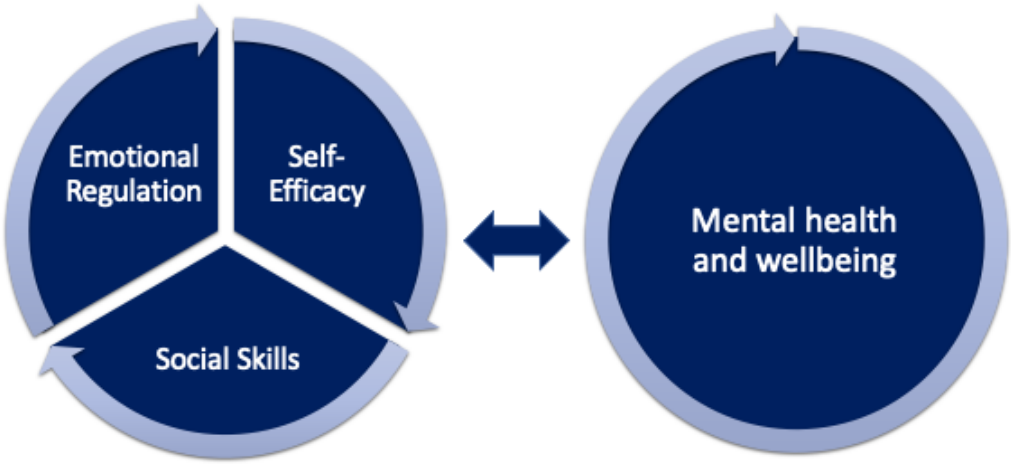
1. How have Ethiopian children's SEL and academic learning changed following school closures, once other factors are taken into account?
2a. What is the relationship between Ethiopian children's SEL and their numeracy[v] after school closures?
2b. To what extent does pupils' SEL predict their numeracy performance following school closures, once other factors are taken into account?
3. What is the relationship between Ethiopian children's SEL and their mental health and wellbeing, after COVID-19 school closures?
4. To what extent are teachers in Ethiopian classrooms supporting learners' SEL?

3.2 Selecting, Adapting and Administering Measures of SEL, Mental Health and Wellbeing

In order to respond to the research questions, we focus on three key aspects of SEL, namely learners' social skills, self-efficacy and emotional regulation, being identified as important for wider life outcomes by Yorke et al. (2021b). These areas cover a broad set of SEL competencies, encompassing both inter- and intrapersonal skills, and were deemed by Ethiopian colleagues and experts to be important for learners in the country. We further captured data on pupils' mental health and wellbeing, given its hypothesised relationship with SEL as illustrated in Figure 1 and growing evidence on the adverse impact of the COVID-19 pandemic on children's welfare (Yorke et al., 2021b).

In each case, we drew on previous data collected in 2019, and complemented this with data from the same pupils aged 9 to 13 collected in January to March 2021 after schools started re-opening. The data collected include self-reporting scales, which are considered to be the most effective means of capturing information from a large number of respondents within a limited timeframe and have been previously used to research children's SEL in diverse contexts (Bayley et al., 2021). They further allowed us to compare the older pupils' responses on the same social skills scale over time as they moved from Grade 4 to Grade 6, and to access feelings and perceptions that cannot be easily observed (Assessment Work Group, 2019).

Figure 1 – Hypothesised Relationship between SEL and Mental Health and Wellbeing (Recreated from Figure 2, Yorke et al. (2021b))



We drew on several sources and databases to identify suitable scales to measure children's SEL, mental health and wellbeing in Ethiopia.[vi] Given the importance of culture in understanding and assessing SEL, we sought to select tools that had been developed or used in similar contexts or low-income settings (Bayley et al., 2021). They further needed to be age-appropriate. Such consideration required reflection on different conceptions of childhood, the ages at which parents and teachers expect learners to contribute to family income or assume household responsibilities like caring for younger siblings, and the developmental milestones to be met for children to engage cognitively in the process of providing

responses and answering surveys (Burman, 2008; Raikes et al., 2017). The most prevalent tools used in low- and lower-middle-income contexts, such as IDELA and MELQO had been created for measuring related components of SEL in low-income countries, but they focus on school readiness among 3-6-year-olds, and so were unsuitable for respondents in the current study (Pisani, Borisova & Dowd, 2018; Raikes et al., 2019; Wolf et al., 2017; Wolf & McCoy, 2019).

Practical factors also guided the selection of the SEL measures. For example, as we were keen that others could adopt a similar approach in related settings, we sought tools that were freely accessible and available without the need for licence fees, and that could be adapted and translated for use in Ethiopia (Bayley et al., 2021). To allow for comparability of the data collected in 2019, our starting point was the scales used at that time (see above). An initial review identified nine scales, from which we created a pool of 91 items. Experts and colleagues from Addis Ababa University and the Ethiopian Policy Studies Institute then evaluated the items independently, considering their cultural suitability and relevance to the study. In several cases, they also proposed alternative wording or new items, for example, '*If I am hungry there is enough to eat at home*'. Finally, we examined areas of consensus and disagreement across the items to refine the pool and select the scales. These are shown in Table 2 while the full instruments are set out in Appendix 1.

Table 2 – SEL Scales with Grades 3 and 6 Pupils in 2021 Data Collection

Aspect of SEL	Scale Name	Items	Reference
Social Skills	Children's Self Report Social Skills Scale (CS4)*	7	Danielson & Phelps (2003)
	Matson Evaluation of Social Skills with Youngsters (MESSY)	5	Matson, Rotatori & Helsel (1983)
Self-Efficacy	Self-Regulation of Learning Self-Report Scale (SRL-SRS)	11	Toering, Elferink-Gemser, Jonker, van Heuvelen & Visscher (2012)
Emotional Regulation	Student Learning in Emergency Checklist (26) (SLEC-26)	10	Forsberg, Schultz, Lodi & Tubbs Dolan (2019)
Mental Health and Wellbeing	WHO Wellbeing Index	8	WHO Collaborating Center for Mental Health (1998)

* The items from this scale were also used in the 2019 data collection.

Once agreed, the scale items were translated carefully by experienced translators who were native speakers of the relevant language. Bilingual members of the RISE Ethiopia team provided supervision and guidance to clarify any ambiguities and ensure that the statements would be intelligible for the young respondents (Bayley et al., 2021). The RISE team subsequently trained 47 enumerators to conduct the surveys, first orally, then on paper and finally using tablets. This training, together with a pilot involving 393 pupils in two schools, provided an opportunity for the field staff to practise using the tools and to ensure that they functioned as intended. The pilot data also allowed us to check the psychometric properties of the scales to confirm that they offered appropriate reliability and validity.[vii]

During the main data collection, the final scales were read aloud to the Grade 3 learners one-to-one to avoid any need for literacy. The Grade 6 students were guided in groups to complete the surveys to ensure that they understood the questions, with each of them responding on their own using pencil and paper, keeping their responses confidential. In each case, pupils ranked their views of their own skills and experiences by expressing their agreement or disagreement with statements like '*I feel confident talking to others*' and '*I can handle whatever comes my way*'. Specifically, they rated themselves on a 5-point Likert scale comprising the following options: Strongly Disagree; Disagree; Undecided; Agree; and Strongly Agree. The least favourable responses were presented first to mitigate the effects of social desirability bias and any inclination among the children to simply agree with the statements (Bayley et al., 2021). The use of a 5-point scale also marked a change from the data collection in 2019, when learners were invited to rank their social skills and support for learning on three response options: Agree,

Somewhat Agree, and Do Not Agree. The shift to a 5-point scale was to allow greater variability in responses, with the aim of achieving better gradation and nuance in the pupils' responses (Krosnick & Presser, 2010; Leung, 2011).

Capturing data on children's SEL, mental health and wellbeing raised important ethical issues which were considered carefully throughout the study. We were mindful throughout the research of the time taken for pupils to complete the scales and adjusted the number of items accordingly to minimise disruption to their learning. By asking questions regarding the learners' feelings, personal experiences and household behaviours, the enumerators were also granted access into a private space (Bayley et al., 2021; Homan, 2001). The research team therefore treated the children and their data with sensitivity. They also obtained their voluntary informed assent at the start of each survey from the children, protecting the confidentiality of their responses, including from teachers, and critically avoiding any harm or discomfort that could arise through their participation.

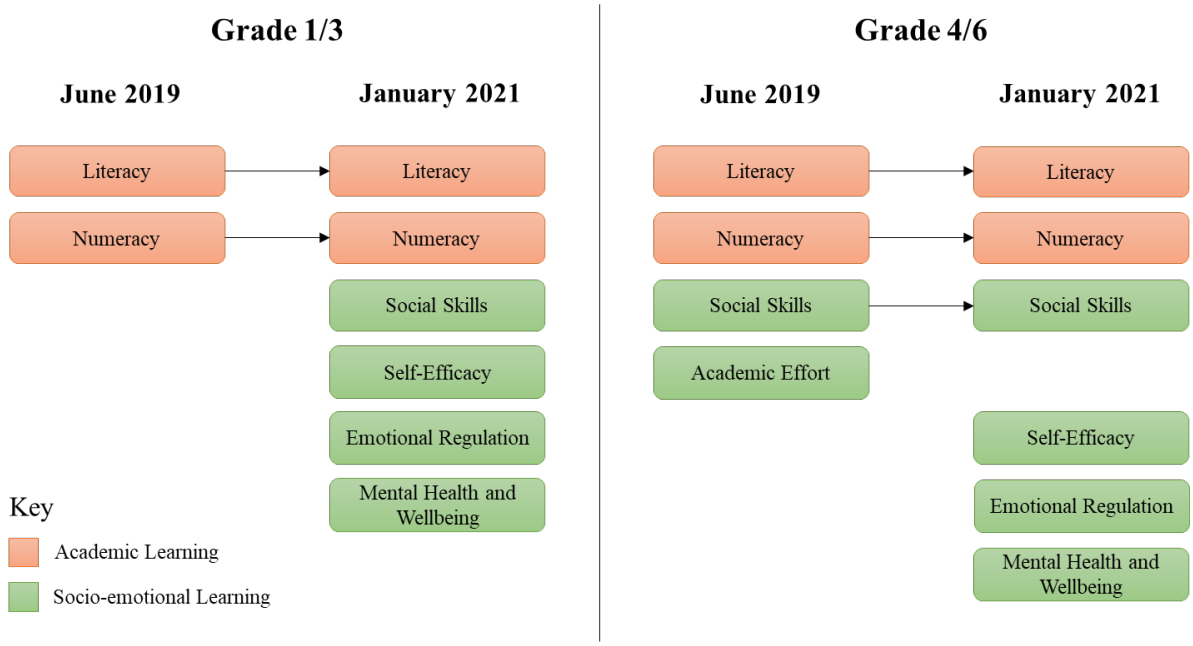
3.3 Control Measures and Lesson Observations

In addition to their SEL, mental health and wellbeing, all Grade 3 and 6 learners were tested on their literacy and numeracy using established RISE measures adapted from the 2018/2019 instruments. The Grade 4/6 test was an adapted version of the Young Lives survey, developed based on guidance from the Ethiopian Ministry of Education and the National Educational Assessment and Examinations Agency (NEAEA) (Kim et al., 2021). The Grade 3 test was adapted from MELQO with the NEAEA through the Early Learning Partnership, but also included tasks derived from the Young Lives survey, Ethiopian curriculum and Early Grade Mathematics Assessment. In each case, the tools were tailored to the expected abilities of the particular cohort, such that the Grade 3 tests comprised tasks like letter and number recognition, while the Grade 6 measures involved more advanced reading activities, comprehension questions and exercises to conduct mathematical calculations and solve problems.

Pupils' performance on the assessments were analysed through item response theory (IRT), which identifies link items of similar difficulty across datasets to allow direct comparison of students' skills and learning over multiple time points (see for example Tiruneh et al. (2021) on the use of IRT using test scores with RISE Ethiopia data). These scores were then transformed and put onto a common scale with a mean of 500 and a standard deviation of 100.

Figure 2 summarises the key academic and SEL variables captured at the two time points. The enumerators also recorded certain key information about the children in 2021, namely their gender, language spoken, age and class grade.[viii] Between 2019 and the 2021 data collection, most of the pupils had progressed into Grades 3 and 6, at which stage they were expected to be aged 9-10 and 12-13 years old respectively.

Figure 2 – Data Collection Across Cohorts and Time Points



Finally, in order to identify the types of teaching practices adopted in Ethiopian classrooms associated with supporting SEL, we conducted lesson observations using the World Bank's *Teach* tool (World Bank, 2019). This exploratory approach was the first use of the *Teach* lesson observation protocol in Ethiopia, following its recent use by members of the team in classrooms in Rwanda (Carter et al., 2020a). Fieldworkers working with the Ethiopian Policy Studies Institute had been extensively trained through the RISE programme to conduct lesson observations using protocols including the MELQO Measure of Early Learning Environments (Rossiter et al., 2018). As the observers spoke and read English, they were able to use the original tool without the need for translation into multiple Ethiopian languages. In each case, the team recorded the number of pupils in the class, the availability of resources like textbooks, and the number of minutes teachers spent on different activities, such as whole-class instruction and group-based learning. Of greatest relevance to the present study, the observers also used existing behaviour descriptions to rate each teacher on a 3-point scale from low to high on key practices to foster children's SEL. These focused on behaviours to nurture pupils' social and collaborative skills, their autonomy and perseverance, with teachers scored on items including the extent to which they facilitated students to engage in peer interactions, provided them with choice and encouraged them to set goals.

3.4 Sampling

The study gathered data from pupils and classrooms in 138 government schools selected from across Ethiopia. They were drawn from the RISE Ethiopia sample of 168 schools, each of which had participated in earlier rounds of data collection. The shortfall comprised four schools in Oromia, six in Benishangul Gumuz and all 20 schools in Tigray, which had to be omitted from the 2021 survey due to ongoing conflict, violence and insecurity in the local areas.

The RISE Ethiopia sample was developed to assess the impact of the Ethiopian government's GEQIP (see Hoddinott et al. (2019) for details of the sampling approach). Specifically, the sample included schools from five regions, Addis Ababa, Amhara, Oromia, the Southern Nations, Nationalities and Peoples' region (SNNP) and Tigray, which have also been included in the Young Lives sample, and as well as Somali and Benishangul Gumuz as emerging regions in the east and west of the country respectively. Across the more populous regions, namely Addis Ababa, Benishangul Gumuz, Somali and Tigray, the number of schools in the sample was allocated according to region size, also taking into

account the proportion of people living in urban and rural areas. The sample includes 64 percent rural schools and 36 percent urban schools across the six regional locations. This is intended to give an illustrative picture of Ethiopia overall, given that the country is predominantly rural.

Within schools, learners were selected based on their participation in the 2019 RISE data collection, when they were enrolled in Grades 1 and 4. At such time, pupils were sampled randomly from the relevant grades where their school had no more than two parallel Grade 1 and Grade 4 classes. In schools with more than two Grade 1 or 4 classes, two classes were first randomly selected before children were randomly sampled from across those two classes. Overall, the 2021 data collection captured SEL information from a panel of 6,018 pupils (49 percent female). Table 3 shows the breakdown of learners by gender and region.

Table 3 – 2021 Data Collection Breakdown by Pupil Gender and Region

Region	Number of Schools	Grade 1/3		Grade 4/6		Total
		Female	Male	Female	Male	
Addis Ababa	20	220	228	239	256	943
Amhara	25	268	249	275	285	1,077
Benishangul Gumuz	14	160	169	161	184	674
Oromia	37	425	415	386	466	1,692
SNNP	22	201	216	228	224	869
Somali	20	197	192	187	187	763
Total	**138**	**1,471**	**1,469**	**1,476**	**1,602**	**6,018**

For the lesson observations, the research team observed Grade 6 classes, typically mathematics lessons, by 138 teachers in each of the participating schools. These teachers were sampled according to the number of children surveyed in 2021 taught in their classes. Specifically, where there were multiple Grade 6 mathematics teachers in one school, we selected teachers for observation to maximise the number of surveyed pupils in their classes. In total, we matched the 138 observed teachers to 2,292 of the surveyed pupils who were in their classes. On several occasions, lesson scheduling did not permit the observation of mathematics lessons and so the research team observed the sampled Grade 6 mathematics teacher instructing another academic subject.

In total, 30 teachers, or 21.7 percent, were female. This is much lower than the proportion of female teachers in the overall sample, being 44.1 percent, suggesting lower representation of women among Grade 6 mathematics teachers. Indeed, this is consistent with recent statistics regarding the declining proportion of female teachers in higher levels of Ethiopian education, reducing from 92.7 percent in kindergarten to 41.1 percent in primary, and just 19.6 percent in secondary classrooms (Federal Ministry of Education, 2020a).

3.5 Descriptive Statistics of the Sample

Table 4 sets out key descriptive statistics for the sample of pupils used in our analysis. It shows mean ages of 10 years for the younger cohort and 13 years for the older cohort, broadly in line with expectations for grade.[ix] Around half of households identify having acceptable levels of food consumption. Across both groups, approximately one-quarter of caregivers are literate (based on information collected by asking the household respondent to read a sentence on a card).

Children's SEL, literacy and numeracy data were captured from learners who were assessed as part of the RISE June 2019 survey and who were still enrolled and attending school in January 2021 following the 2020 school closures. Between 2019 and 2021, 83.4 percent of the initial Grade 1 pupils and 90.3 percent of Grade 4 learners had advanced into Grades 3 and 6 respectively ('on track' learners in Table 4). When schools re-opened, pupils were automatically promoted to the next grade. The remaining 16.6

percent of pupils in the younger cohort and 9.7 percent of the older learners had fallen one or more years behind their peers, which is likely to be due to them repeating a grade. These children were included in the data collection and analysis, with account taken of the grade that children had actually reached. For convenience of reference, throughout the paper all pupils in the younger and older cohorts are referred to as 'Grade 3' and 'Grade 6' learners respectively, while recognising that some of them were still enrolled in lower grades. Similarly, some students in each cohort were at least two years above the official age: 17.1 percent of the younger children and 15.0 percent of the older students were two or more years older than the designated ages for their grade. This is also taken into account in the analysis that follows. With respect to teachers in the grade 4/6 sample, the majority have a diploma, and around 4.7 years of experience, on average.

Table 4 – 2021 Descriptive Statistics by Grade

Variable	**Grade 1/3**	**Grade 4/6**
Children and household characteristics		
Female* (%)	50.0	48.0
Average age* (years)	10.0 (1.5)	13.0 (1.5)
Over age* (%)	17.1	15.0
'On track'* (%)	83.4	90.4
Rural* (%)	64.8	64.2
Food consumption (% acceptable)	53.7	50.4
Caregiver literacy (% literate)	26.8	26.2
Average household size	5.7 (1.9)	5.9 (2.0)
Teacher characteristics		
Teacher holds certificate (%)	-	10.3
Teacher holds diploma (%)	-	81.7
Teacher holds degree (%)	-	6.4
Average teacher experience (years)	-	4.7 (4.4)
Observations	**2,940**	**3,078**

Notes: Standard deviations shown in brackets (where appropriate).

* indicates data captured from learners in 2021 (the remaining variables being derived from the 2019 data collection).

Pupils aged 12 and older were considered over age for Grade 3, and 15 and older for Grade 6. Learners were classified as 'on track' if they were promoted with their peers as expected between 2019 and 2021. The classification does not indicate whether a learner was in the right grade for his/her age at either timepoint, and so for the present purposes, a pupil could be both 'on track' and over age.

3.6 Student Attrition

Given our analysis focuses on those who remained in school between June 2019 and January 2021, we were mindful that this could potentially result in selection bias. In particular, if those from more disadvantaged backgrounds were no longer in the sample for example because they had migrated, dropped out or were absent due to the effects of COVID-19 school closures, this could affect the results. In order to understand the impact this could have on our analysis, we assessed the extent of attrition in the sample across different population groups.

We found that 19 percent of learners surveyed in the 2019 data collection were missing during the 2021 round.[x] The rates were very similar across the two grade cohorts: 19.3 percent of the younger children and 19.4 percent of the older children (Table 5).

Table 5 – Pupil Participation, Enrolment and Attrition by Cohort

Cohort	**Participated in June 2019 Collection**	**Surveyed in January 2021**	**Not surveyed in January 2021**	**Attrition Rate**
Grade 1/3	2,729	2,202	527	19.3
Grade 4/6	3,082	2,483	599	19.4
Cohort	**Enrolled in June 2019**	**Enrolled in January 221**	**Not enrolled in January 2021**	**Drop Out Rate**
Grade 1/3	2,560	2,236	324	12.7
Grade 4/6	2,865	2,505	360	12.6

The attrition rate broadly accords with the findings from RISE Ethiopia prior to COVID. For example, Weldesilassie, Hoddinott and Woldehanna (2020) report attrition of 24.1 and 16.2 percent for Ethiopian pupils in Grades 1 and 4 respectively *over one academic year*. Given the longer time lapse between rounds of data collection and the unprecedented school closures in response to the COVID-19 pandemic, attrition of 19 percent suggests that a high proportion of students were tracked successfully.

In order to understand the reasons for attrition, we examined the extent to which this was due to children dropping out of school following school closures. In total, 12.7 percent of Grade 1 and 12.6 percent of Grade 4 pupils dropped out before they reached Grades 3 and 6 respectively (see Table 5). The difference between the attrition and drop-out rates of around 7 percent represents learners who were enrolled in school but could not be physically located at the time of the 2021 data collection. Reasons given included both temporary absence for children who were ill or in hospital on the day of the surveys, to longer-term arrangements where learners had changed school or migrated to different areas with their families. Eight pupils also declined to take part in the study.

Overall, attrition from the sample was quite low given the wider circumstances. However, it is important to identify if there are patterns of attrition within and across groups of learners to understand whether there are systematic differences between children missing from the sample. To address this, Table 6 sets out attrition rates for pupils disaggregated according to key individual and household characteristics. The results show slightly higher attrition among girls than boys in both cohorts, but the differences between them are not statistically significant in either case. By contrast, attrition appears to be related to pupils' location, region and family wealth. Specifically, children from poorer households and living in rural locations were significantly less likely to be available to participate in the 2021 round of data collection than wealthier pupils or those living in urban areas. Attrition rates were also higher among learners attending schools in Benishangul Gumuz, Oromia and SNNP, compared to those in Addis Ababa, Amhara and Somali regions, and the differences by region are also statistically significant. As indicated above, the high rates of attrition in Benishangul Gumuz and Oromia reflect, at least in part, the unavailability of some schools in those regions to take part in the 2021 data collection because of local insecurity.

Table 6 – Comparison of Attrition Rates (%) by Child and Household Characteristics between Data Collection for 2019 and 2021 Academic Years

Child and Household Characteristics	Grade 1/3	Grade 4/6
Gender		
Girls	19.33	20.45
Boys	19.19	18.36
Pearson chi-squared	0.01	2.15
Location		
Rural	21.28	22.61
Urban	15.90	13.69
Pearson chi-squared	11.79**	35.91***
Region		
Addis Ababa	16.53	8.90
Amhara	16.70	15.41
Benishangul Gumuz	33.54	36.10
Oromia	19.44	21.40
SNNP	20.79	21.21
Somali	8.87	14.75
Pearson chi-squared	64.43***	116.11***
Wealth		
Tercile 1 (poorest)	25.38	25.51
Tercile 2	17.79	20.98
Tercile 3 (wealthiest)	14.73	11.79
Pearson chi-squared	35.06***	64.11***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

To interrogate these differences further, we conducted logistic regressions to explore the extent to which these factors predict children's attrition between 2019 and 2021, taking the other variables into account. We also included learners' age, which was found to be significantly associated with their attrition in previous rounds of RISE data collection (Weldesilassie et al., 2020). Table 7 sets out the results of the regressions for both cohorts. They show that there are indeed statistically significant differences in attrition between the regions but they are not necessarily consistent across the two cohorts, and the differences by rural and urban location cease to be significant when controlling for the other variables. However, variables for pupils' age and family wealth are significant for both groups, with older children more likely to drop out between 2019 and 2021, and lower attrition among learners from the wealthiest tercile.

Table 7 – Logistic Regression Results Showing the Likelihood of Attrition between 2019 and 2021

2021 Attrition		Grade 1/3	Grade 4/6
Gender (base group: female)		-0.01 (0.10)	-0.21 (0.12)
Age		0.17*** (0.03)	0.17*** (0.03)
Family Wealth (base group: first tercile/poorest)	Second Tercile	-0.16 (0.12)	-0.16 (0.15)
	Third Tercile	-0.33* (0.17)	-0.38* (0.19)
Location (base group: rural)		-0.16 (0.16)	0.21 (0.17)
Region (base group: Addis Ababa)	Amhara	-0.34 (0.21)	0.24 (0.25)
	Benishangul Gumuz	0.55* (0.22)	-0.79* (0.38)
	Oromia	-0.27 (0.20)	-0.03 (0.25)
	SNNP	-0.17 (0.21)	0.59* (0.26)
	Somali	-1.00*** (0.28)	0.60* (0.28)
Constant		-2.41*** (0.30)	-3.83*** (0.43)
Observations		2,726	2,565
Pseudo R-squared		.04	.04

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

In light of the results, we controlled for these variables when analysing pupils' SEL, to ensure that factors associated with their attrition did not confound the results pertaining to relationships between other aspects of their learning and development. We also calculated attrition weights following the approach used for the earlier round of RISE (see Weldesilassie et al. (2020)). The inclusion of a variable for children's region accounted for variations in the sampling approach between the 2019 and 2021 rounds of data collection, specifically the missing schools in Benishangul Gumuz and Oromia. The final weights were then used in the analyses described below to mitigate the influence of different patterns of attrition in the emerging results.

3.7 Validating the SEL Scales

Given the SEL instruments were adapted from established tools for the purpose of this study in Ethiopia, it was important to first validate them. To do this, we followed a similar process to that described by Yorke et al. (2021a) using related data collected for RISE Ethiopia.

An initial review of the 2021 data found that learners' responses were negatively skewed, with more students agreeing with the scale statements than disagreeing with them.[xi] The histograms in Appendix 2 show the distribution of the SEL scales, and pupils' numeracy scores, for each of the cohorts. The shape of these distributions and their skewness are consistent with the SEL data from the 2019 collection, notwithstanding the increase in response categories from three to five Likert points to allow for a greater distribution in scores (Yorke et al., 2021a). Other factors that could have affected the skewness of the data include social desirability bias, the inclination and predisposition for the children to provide socially accepted responses or agree with other people, and attrition between rounds of data collection (Schulz, 2005).

We then examined the internal consistency and reliability for the four scales, denoted by Cronbach's alpha (α) and set out in Table 8. In each case, it shows coefficients ranging from .88 to .93, which indicate a high level of reliability for the scales for both cohorts, well above the .70 threshold widely considered as acceptable (Field, 2009).

Table 8 – Reliability Analysis for Six SEL Scales

Scale	No. Items	Cronbach's Alpha (α)	
		Grade 3	Grade 6
Social Skills	12	.91	.93
Self-efficacy	11	.91	.89
Emotional Regulation	10	.89	.89
Mental Health and Wellbeing	8	.88	.88

Factor analyses were undertaken separately for each of the two cohorts, recognising that variables including learners' age, level of development and education could affect their responses, and therefore the validity of the scales. To ensure the rigour and robustness of the process, we used exploratory and confirmatory factor analyses with different subsets of the data, in each case randomly created so that no responses were used in both types of analysis (Wolf et al., 2017). The full details and results of the validation process are set out in Appendix 3.

In summary, the analyses for the Grade 6 learners' data showed that their responses loaded onto basic single-factor models for the different SEL scales. In each case, the eigenvalues for the factor exceeded the recommended threshold of 1 and standardised loadings ranged from .62 to .80 indicating that each scale measured one main trait (Kaiser, 1960). These findings were corroborated by the confirmatory factor analyses, which were used to test the hypothesised one-factor model for each scale and which showed similar coefficient loadings between .62 and .82. The goodness of fit indices, in particular the root mean square error of approximation, which is more robust for large sample sizes, showed that levels of fit were acceptable and in some cases good (Hu & Bentler, 1999).

The results for the Grade 3 pupils' data were less straightforward. The exploratory factor analyses for learners' social skills, emotional regulation and mental health and wellbeing revealed that the items did not load onto one factor. The confirmatory factor analyses further showed that the single-factor structures using all items were not the most appropriate models for these scales.

Given these issues, we made various adjustments to improve the models for the Grade 3 learners' responses. First, we added covariances between scale items where there was a theoretical basis for doing so, for example between the statements relating to pupils' emotional regulation '*It is easy for me to stick to my aims*' and '*It is easy for me to achieve my goals*'. Next, we removed items loading onto the second factor and tested the reduced models. We then ran further confirmatory analyses using both correlated factor and bifactor models. The former involves two separate factors accounting for a correlation between them, while the latter includes a general factor which loads onto all observable variables as well as two unrelated factors that load onto the sub-groups (Dunn & McCray, 2020). Table 9 compares the fit indices for different models for the mental health and wellbeing scale.

Table 9 – Goodness of Fit Indices for Different Models for Mental Health and Wellbeing

Mental Health and Wellbeing	Coefficient Loadings	Chi-squared	CFI	TLI	RMSEA	SRMR
Threshold for good fit	-	p>.05	>.95	>.90	<.06	<.08
Basic one-factor model	.70-.79	p < .05	.678	.549	.075	.101
Basic model with covariances	.59-.79	p < .05	.881	.825	.046	.048
Reduced one-factor model	.72-.79	p < .05	.972	.945	.033	.021
Correlated factors model	.73-.82	p < .05	.804	.711	.060	.063
Bifactor model	.07-.77	p < .05	.987	.969	.020	.015

Notes: CFI = comparative fit index; TLI = Tucker-Lewis index; RMSEA = root mean square of error of approximation; SRMR = standardised root mean square residual.

In light of theoretical considerations and these goodness-of-fit indices, we decided to proceed with the reduced models for the three scales that loaded onto more than one factor. This approach offered the

greatest simplicity and parsimony, and has been used to improve model fit in various other studies examining children's SEL in low-income countries (Forsberg et al., 2019; Raikes et al., 2019; Wolf et al., 2017). We nevertheless recognised that there could be some issues that account for the variation in the Grade 3 learners' responses to the questions which could affect the patterns. For example, the grouping of related items, where statements at the start of the scale loaded onto one factor and items at the end onto another, could signal that the pupils were becoming fatigued by the process. Alternatively, the younger children might have had more difficulty understanding the questions, or they might have been easily distracted or unsettled by the presence of the data collectors or the unfamiliar one-to-one test setting. These potentially provide lessons for future studies including younger children in the self-reporting of SEL, for example the value of using reduced scales to accommodate shorter attention spans.

To analyse learners' SEL, we first needed to put them onto common scales. We followed the approach used by Espinoza and Ogando (2019) with Young Lives data on adolescents' agency and self-efficacy. This entailed creating standardised *z*-scores with means of 0 and standard deviations of 1 for each item, and then averaging the *z*-scores within each scale. This approach enabled us to locate each respondent relative to the wider sample and could be used with SEL data both captured with 3-point Likert scales in 2019 and 5-point response options in 2021 (Dawis, 2000). In the case of the longitudinal analyses of social skills for RQ1, however, we only included the standardised scores for the seven common items used in both 2019 and 2021 to maximise their comparability.

3.8 Analytical Approach

The study used a combination of bivariate analyses and multivariate regressions to examine children's SEL and answer the RQs. Given the importance of factors like learners' gender, location and family wealth, we either disaggregated their data by group or controlled for such variables in the regressions. We used responses drawn from the 2021 data collection wherever possible, being particularly important for factors that had or were likely to have changed between 2019 and 2021, such as pupils' age and class grade. There were, however, certain variables that were only recorded in 2019, such as indicators of household wealth and caregiver literacy, which were captured through household surveys. In these cases, we used covariates derived from the earlier round of data collection.

To conduct the longitudinal analyses to address RQ1, which examines changes in SEL and academic learning during the school closures, we adopted a two-pronged approach with two separate datasets. First, we used *z*-scores from a panel dataset to compare the difference in pupils' responses between 2019 and 2021 and *t*-tests to examine the significance of those differences. We also ran additional *t*-tests to look for significant differences in learners' social skills and numeracy scores by gender and location across the two time points. Second, we used a cross-sectional dataset to analyse the relationship between 2019 and 2021 scores, conditional on other factors, akin to a value-added model (Tiruneh et al., 2021). In this case, we ran multivariate linear regressions, building up from a basic model focused on the association between learners' skills in 2019 and 2021. We then included variables for their wider skills in 2021, numeracy or social skills as appropriate, and other child, family, teacher and school location factors. Full details of the specific covariates used in these and subsequent regressions are set out in section 4.1.

The cross-sectional dataset, which includes richer information on SEL together with mental health and wellbeing following school closures, was used in analyses to answer the other RQs. For RQ2a and RQ3, we conducted multivariate linear regressions with the dataset to examine the relationship between SEL and numeracy for RQ2a and between SEL and mental health and wellbeing for RQ3. RQ2b more specifically focused on those who were identified as low performers in 2021 to enable logistic regressions and explore any association between their SEL and low performance over time. Finally, for RQ4 we included the ratings from lesson observations to run further linear regressions to examine the relationship between teacher behaviours and pupils' SEL responses.

In each of the regressions, we included attrition weights to account for patterns of drop-out between the 2019 and 2021 rounds of data collection. Such weights were calculated further to the analyses described in section 3.6 and in accordance with the approach outlined by Weldesilassie et al. (2020). We also clustered standard errors at the school level to accommodate the likely effects of similarities among pupils within the same schools and increased variation between different schools.

Finally, in addition to the reported models, we undertook robustness checks by running regressions including school fixed effects to account for unobserved heterogeneities in the same schools *over time*. Similarly, to account for the effects of standardisation and the switch from using 3-point to 5-point Likert scales, we tested the results using proportional social skills scores between 0 and 1, and recoded responses which placed pupils' 2021 answers onto a 3-point scale. The latter, for example, entailed combining children's scores where they answered 'Agree' or 'Strongly Agree' on items like '*I like to share things with others*'. In each case, the results were identical or very similar in terms of the significance of the variables of interest, with only small differences in their respective coefficients. As such, we do not report these additional models, but the results are available from the authors upon request.

4. Results and Findings

The preceding sections highlighted the emerging but still limited research regarding children's SEL in low- and lower-middle-income countries, and the methods used in Ethiopia to capture and analyse data from primary school pupils to address this gap and to better understand the impact of the COVID-19 school closures. This section now sets out the results of the study and the key findings by research question. In each case, we specify the particular focus, describe the data and investigate the variables of interest once other factors are taken into account, before providing our interpretation of the findings and their significance.

4.1 Changes in SEL and Academic Learning

Like most other countries worldwide, Ethiopia has faced difficult educational decisions following the re-opening of its schools after the COVID-19 closures. In part, these choices reflect the realities of children's diverse experiences at home and the extent to which they have been able to continue their learning. This study therefore offers valuable insight by drawing on longitudinal data collected from Ethiopian pupils both before the COVID-19 pandemic and since schools have resumed. In particular, this section aims to address RQ1: '*How have Ethiopian children's SEL and academic learning changed following school closures, once other factors are taken into account?*'

Specifically, we focused on changes in pupils' social skills and their numeracy. As indicated in Figure 2 above, data concerning learners' social and numeracy skills were collected as they moved from Grade 4 to 6, enabling a comparison of their responses within and across time points. Figure 3 shows the trends in social skills for the full panel of learners surveyed in both 2019 and 2021, disaggregated for gender and location. Tables 10 and 11 further show mean *z*-scores for the different groups of children and the results of *t*-tests to examine the significance of such differences.

Figure 3 – Changes in Grade 4/6 Social Skills between 2019 and 2021

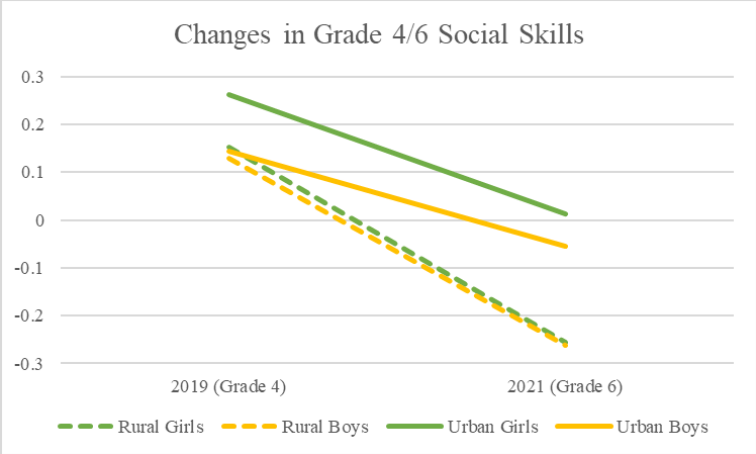


Table 10 – Mean *Z*-Scores and *T*-Test Results for Differences in Social Skills between 2019 and 2021

Group	2019 Social Skills *z*-score	2021 Social Skills *z*-score	Change	*t*-value
Overall	0.16	-0.17	-0.33	17.18***
Urban Girls	0.26	0.01	-0.25	6.31***
Rural Girls	0.15	-0.26	-0.41	10.48***
Urban Boys	0.14	-0.06	-0.20	4.93***
Rural Boys	0.13	-0.26	-0.39	10.87***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. $n = 2{,}448$.

Table 11 – Mean *Z*-Scores and *T*-Test Results for Differences in Social Skills in 2019 and 2021

Measure	Girls	Boys	*t*-value	Rural	Urban	*t*-value
2019 Social Skills	0.20	0.14	2.50*	0.14	0.20	-2.41*
2021 Social Skills	-0.15	-0.18	1.09	-0.26	-0.02	-7.51***
Change in Social Skills	-0.35	-0.32	-0.71	-0.40	-0.22	-4.54***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. $n = 2{,}448$.

Most learners started at a similar point in 2019 with only urban girls reporting noticeably higher levels of social skills before the school closures. Over time, however, the chart and Table 10 indicate a statistically significant decline in *all* pupils' social skills captured prior to and after the COVID-19 school closures, before taking other variables into account. In practice, this meant that smaller proportions of children in 2021 agreed with the statements '*I make friends easily*' and '*Other people like me*' than in 2019. The results also highlight significant differences between various groups of learners, according to their gender and location. Table 11 reports a significant difference between rural and urban learners at both time points, but only significantly higher social skills among girls in 2019. Although all groups show a decrease in social skills, the gradient is steepest for respondents in rural settings, suggesting widening disadvantage and the greatest losses among rural children. By contrast, the decline is smallest for urban boys, indicating that across the groups, their social skills were impacted by the school closures the least. Similarly, urban girls show smaller losses than their rural counterparts and overall higher social skills than both boys and rural learners when measured following the school closures.

Figure 4 and Tables 12 and 13 show a different pattern regarding pupils' numeracy scores across the same time period. In both 2019 and 2021, there are significant differences according to gender and location, with boys and learners in urban settings achieving significantly higher scores than girls and rural pupils respectively. During the period of school closures, all children show only slight learning

gains with these starting from a low base and varying in magnitude and significance according to gender and location. Only urban boys and rural girls show statistically significant increases in numeracy over the period of the school closures. However, the improvement in numeracy for rural girls still leaves them behind their counterparts. More generally, a wide gap remains between those in rural and urban areas.

Comparing these data with scores in earlier rounds of RISE assessments sheds further light on the different academic trajectories. Using data on Grade 4 pupils' numeracy scores at the start and end of the 2018/2019 academic year, we calculated an average increase of 34 points (or 0.34 standard deviations) over the period. Assuming that learners should have at least been able to make the same progress (of 34 points) between the end of Grade 4 and start of Grade 6, we extrapolated where those children should have reached on a test of similar difficulty in 2021 if the schools had not closed. However, in reality, between 2019 and 2021 their numeracy scores only increased by 12 points (0.12 standard deviations), suggesting that there was a loss in learning over the period of the school closures.

Similarly, we made estimates of expected progress for those in rural and urban areas for the period between the end of Grade 4 and start of Grade 6, based on the progress they each made in Grade 4 during the 2018/2019 academic year. Figure 5 shows these comparisons and the shortfall between expected and actual numeracy scores in 2021 disaggregated by location. For those in urban areas, their learning progressed at less than half of the speed that would have been expected if they continued at the same pace, an average increase of 15 points compared to the expected 37 points. Progress was even lower for rural students, whose learning progressed by only one third, an average of 10 points compared with 32 points, suggesting potentially growing inequalities between urban and rural pupils following the COVID-19 school closures.

Figure 4 – Changes in Grade 4/6 IRT Numeracy Scores between 2019 and 2021

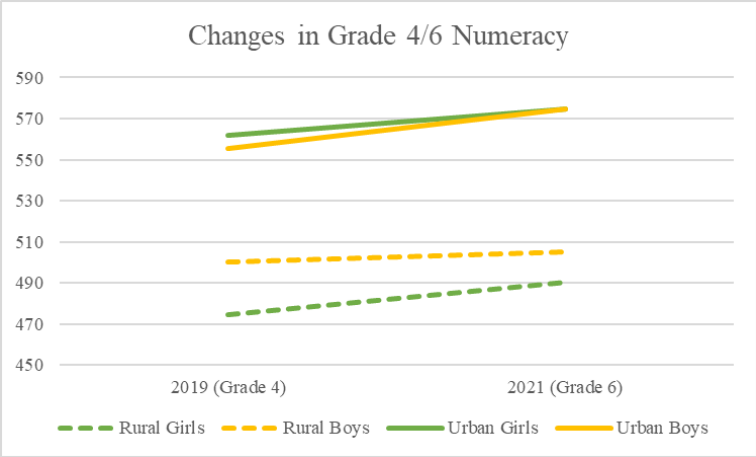


Table 12 – Mean IRT Scores and *T*-Test Results for Grade 4/6 Differences in Numeracy in 2019 and 2021

Measure	Girls	Boys	*t*-value	Rural	Urban	*t*-value
2019 Numeracy	506.97	520.42	-3.16**	488.07	558.60	-16.94***
2021 Numeracy	521.48	530.61	-2.09*	498.17	574.54	-17.99***
Change in Numeracy	14.51	10.19	1.33	10.09	15.94	-1.74

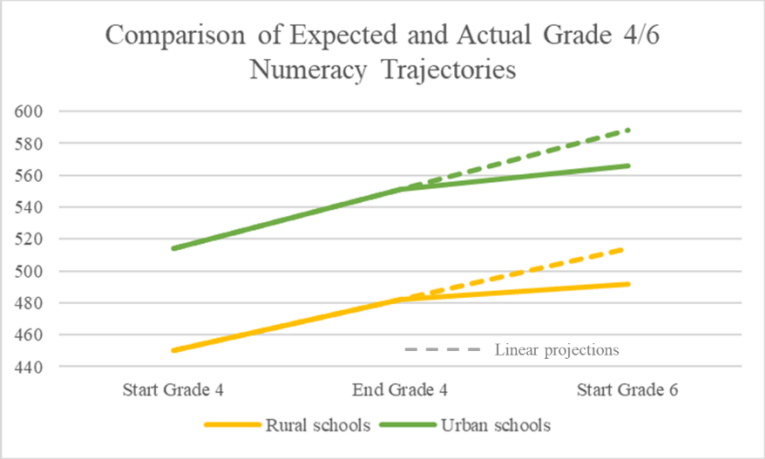
Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. $n = 2{,}346$.

Table 13 – Mean IRT Scores and *T*-Test Results for Differences in Numeracy between 2019 and 2021

Group	2019 Numeracy IRT score	2021 Numeracy IRT score	Change	*t*-value
Overall	514.02	526.26	12.24	-4.02***
Urban Girls	561.95	574.55	12.60	-1.70
Rural Girls	474.60	490.23	15.63	-3.39***
Urban Boys	555.51	574.53	19.02	-2.74**
Rural Boys	500.22	505.32	5.10	-1.02

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. $n = 2{,}346$.

Figure 5 – Comparison of Expected and Actual Grade 4/6 Numeracy Trajectories



Source: Recreated from Kim et al. (2021), Figure 1

We investigated these changes further, and the factors associated with them, by running multivariate regressions and building the models step-by-step by adding new variables. In each case, the standardised scores for children's social and numeracy skills in 2021 comprised the relevant dependent variables. We first examined the extent to which pupils' earlier skills in 2019 predict their later scores in 2021 (shown in model 1). We then explored the relationship between different skills in 2021 (model 2) and added control variables to account for variation in pupils' responses and scores that could relate to wider personal, family, school or geographical factors. The variables included correspond with ones used in earlier analyses of RISE numeracy data, drawing on those applicable to this study based on a review of the relevant literature (Tiruneh et al., 2021). At the individual student level, these comprised variables concerning their gender and age, as well as their class grade, recognising that some learners had not progressed 'on track' with their peers, and to account for the effects of grade repetition (also model 2). We further included variables for the pupils' caregiver's literacy (whether he or she could read a basic sentence) (model 3a), their family wealth and household size (model 3b), their teacher's years of experience and highest qualification (model 4), and the school region and location (whether in a rural or urban setting) (model 5).[xii] Tables 14 and 15 set out the regression results for social skills and numeracy respectively to examine changes following the school closures, once other factors are taken into account, while Table 16 compares the significant variables from the final models.

Table 14 – Regressions for Grade 4/6 2021 Social Skills

2021 Social Skills		**(1) Basic Model**	**(2) + Child Variables**	**(3a) + Caregiver Literacy**	**(3b) + Family Variables**	**(4) + Teacher Variables**	**(5) + School Location**
2019 Social Skills		0.08 (0.02)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.05 (0.04)	0.04 (0.03)
2021 Numeracy		-	0.24*** (0.03)	0.24*** (0.03)	0.21*** (0.03)	0.21*** (0.03)	0.18*** (0.02)
Gender (base group: female)		-	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.03 (0.04)	-0.02 (0.03)
Age		-	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.01 (0.01)	0.00 (0.01)
Class Grade		-	-0.02 (0.04)	-0.02 (0.05)	-0.01 (0.04)	-0.02 (0.05)	-0.03 (0.04)
Caregiver Literacy (base group: no/low literacy)		-	-	0.02 (0.04)	-0.04 (0.04)	-0.04 (0.04)	-0.03 (0.04)
Family Wealth (base group: first tercile (poorest))	Second Tercile	-	-	-	0.09 (0.07)	0.09 (0.07)	0.10 (0.07)
	Third Tercile	-	-	-	0.19** (0.07)	0.18* (0.07)	0.25* (0.10)
Household Size		-	-	-	-0.02 (0.01)	-0.02 (0.01)	-0.01 (0.01)
Teacher Experience		-	-	-	-	0.00 (0.01)	0.00 (0.01)
Teacher Qualification (base group: no training)	Certificate	-	-	-	-	-0.19 (0.20)	-0.14 (0.18)
	Diploma	-	-	-	-	-0.09 (0.13)	-0.07 (0.10)
	Degree	-	-	-	-	-0.08 (0.15)	-0.05 (0.14)
Location (base group: rural)		-	-	-	-	-	-0.15 (0.10)
Region (base group: Addis Ababa)	Amhara	-	-	-	-	-	0.03 (0.07)
	Benishangul Gumuz	-	-	-	-	-	-0.38* (0.16)
	Oromia	-	-	-	-	-	-0.10 (0.08)
	SNNP	-	-	-	-	-	-0.24 (0.12)
	Somali	-	-	-	-	-	-0.19 (0.23)
Constant		0.01 (0.02)	0.16 (0.31)	0.15 (0.31)	0.06 (0.30)	0.16 (0.33)	0.31 (0.33)
Observations		2,120	2,120	2,120	2,120	2,120	2,120
R-squared		.00	.11	.11	.12	.12	.15

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 15 – Regressions for Grade 4/6 2021 Numeracy

2021 Numeracy		**(1) Basic Model**	**(2) + Child Variables**	**(3a) + Caregiver Literacy**	**(3b) + Family Variables**	**(4) + Teacher Variables**	**(5) + School Location**
2019 Numeracy		0.72*** (0.02)	0.68*** (0.02)	0.67*** (0.02)	0.65*** (0.02)	0.66*** (0.02)	0.58*** (0.02)
2021 Social Skills		-	0.21*** (0.03)	0.21*** (0.03)	0.19*** (0.02)	0.19*** (0.02)	0.16*** (0.02)
Gender (base group: female)		-	-0.01 (0.03)	-0.01 (0.03)	0.00 (0.03)	-0.01 (0.03)	0.00 (0.03)
Age		-	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.01 (0.01)	0.02 (0.01)
Class Grade		-	0.13*** (0.03)	0.13*** (0.03)	0.14*** (0.03)	0.12*** (0.03)	0.15*** (0.03)
Caregiver Literacy (base group: no/low literacy)		-	-	0.12* (0.05)	0.04 (0.04)	0.04 (0.04)	0.02 (0.04)
Family Wealth (base group: first tercile (poorest))	Second Tercile	-	-	-	0.06 (0.04)	0.06 (0.04)	0.03 (0.04)
	Third Tercile	-	-	-	0.22*** (0.05)	0.21*** (0.05)	0.09 (0.05)
Household Size		-	-	-	-0.02** (0.01)	-0.02** (0.01)	0.00 (0.01)
Teacher Experience		-	-	-	-	0.01 (0.01)	0.01* (0.01)
Teacher Qualification (base group: no training)	Certificate	-	-	-	-	-0.19 (0.16)	0.09 (0.15)
	Diploma	-	-	-	-	-0.19 (0.15)	-0.06 (0.13)
	Degree	-	-	-	-	-0.24 (0.16)	-0.07 (0.14)
Location (base group: rural)		-	-	-	-	-	0.06 (0.06)
Region (base group: Addis Ababa)	Amhara	-	-	-	-	-	-0.19 (0.11)
	Benishangul Gumuz	-	-	-	-	-	-0.63*** (0.13)
	Oromia	-	-	-	-	-	-0.27** (0.09)
	SNNP	-	-	-	-	-	-0.38** (0.12)
	Somali	-	-	-	-	-	-0.58*** (0.11)
Constant		0.06 (0.03)	-0.80** (0.24)	-0.85** (0.24)	-0.92*** (0.25)	-0.68* (0.27)	-0.84** (0.28)
Observations		2,157	2,157	2,157	2,157	2,157	2,157
R-squared		.51	.54	.54	.55	.55	.58

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 16 – Comparison of Significant Variables for Grade 4/6 2021 Skills Models

Variable	2021 Social Skills model	2021 Numeracy model
2019 Social Skills	-	-
2021 Social Skills	-	✓
2019 Numeracy	-	✓
2021 Numeracy	✓	-
Gender	X	X
Age	X	X
Class Grade	X	✓
Caregiver Literacy	X	X
Family Wealth	✓	X
Household Size	X	X
Teacher Experience	X	✓
Teacher Qualifications	X	X
Location	X	X
Region	✓	✓

Notes: Table shows significant variables once all factors are taken into account. For categorical variables, such as region, the overall variable is shown as significant if at least one category significantly predicted the relevant outcome variable.

The results reveal several important findings. First, they show that pupils' 2019 social skills are not significantly associated with their 2021 social skills, whether or not other factors are taken into account (Table 14, model 1 to 5). By contrast, learners' 2021 numeracy is a significant and positive predictor of their 2021 social skills, even after controlling for different child, school and household factors (Table 14, models 2 to 5). Pupils' family wealth is also associated with their 2021 social skills, with Grade 4 to 6 children from the most affluent backgrounds reporting significantly higher social skills (Table 14, models 3b to 5).

Overall, the results shed further light on the trends shown in Figure 3 and the decline in learners' social skills between 2019 and 2021. The negative coefficients for gender indicate better social skills among girls, but these are not significant for the 2021 data, in line with the *t*-test results in Table 11. However, the coefficient for pupils' location in Table 14 (model 5) is also not significant in contrast to the significant difference between urban and rural pupils shown in Table 11.[xiii] This variation could be accounted for by the inclusion of attrition weights in the regressions, the clustering of standard errors at the school level or the relationship between learners' location and their family wealth, which *does* significantly predict their social skills.

Regarding changes in numeracy following school closures, the results in Table 15 indicate a strong and positive association between learners' scores in 2019 and their performance in 2021 (Table 15, model 1). This remains significant even after controlling for other factors (Table 15, models 2 to 5). Similarly, pupils' 2021 social skills and class grade positively predict their 2021 numeracy, again even after individual, household and school variables are taken into account (Table 15, models 2 to 5).

Beyond the child variables, relatively few wider factors significantly predict children's numeracy. Caregiver literacy (Table 15, model 3a), household size and family wealth (Table 15, model 3b) show initial associations but cease to be significant when controlling for learners' location and region (Table 15, model 5). In the case of caregiver literacy, this could reflect the low levels across the study sample, with just 26.2-26.8 percent of guardians capable of reading a basic sentence, as shown in Table 3. By contrast, pupils with more experienced teachers achieved higher numeracy scores even after taking other factors into account but the coefficient is very small (Table 15, model 5), and similar to social skills, the results do not identify differences between urban and rural learners but this variation may be captured by the effects of region, most of which are statistically significant (Table 15, model 5).

Across the models for social skills and numeracy, there are a couple of important findings. First, the association between pupils' social skills in 2019 and 2021 is weak and statistically insignificant. In practice, this means that children who reported strong social skills in 2019 may not necessarily have indicated good social skills in 2021. This lack of relationship potentially highlights the seemingly broken link between children's social skills before and after the 2020 school closures, and is consistent with the adverse effects of prolonged periods of isolation and distancing in response to the COVID-19 pandemic. For example, fewer learners in 2021 than 2019 indicated making friends easily, sharing things with others and helping people, which could reflect the lack of opportunities for social interaction and practising social skills while schools were closed and contacts were minimised. This contrasts with the results of pre-pandemic research, for example in Ghana by Wolf and McCoy (2019), who found that earlier SEL significantly predicted later SEL over multiple time points.

The relationship between learners' social skills and numeracy, on the other hand, is much stronger, even after taking other factors into account. Children achieving high scores in the 2021 numeracy assessments also reported good social skills. Similarly, the positive association between pupils' numeracy in 2019 and 2021 is significant and large, with strong links between learners' earlier performance and later attainment. Taken together, these findings could suggest that the pandemic has had a greater impact on children's social skills over time than on their numeracy.

In summary, the longitudinal data captured from Grade 4 to 6 pupils in Ethiopia show that social skills have declined between 2019 and 2021 and during the COVID-19 school closures, with the greatest reductions among rural children. During the same period, there have been increases in numeracy but these start from a low base, vary according to learner type with the smallest gains for rural pupils, and appear to fall short of the forecast trajectories for children's learning before the school closures (Kim et al., 2021).

4.2 The Relationship between Pupils' SEL and Numeracy

The second research theme focuses on the associations between learners' numeracy and SEL in 2021. Specifically, RQ2a asks '*What is the relationship between Ethiopian children's SEL and their numeracy after school closures?*' Such analyses are important to understand the interactions between numeracy and the wider set of SEL measures, specifically self-efficacy, emotional regulation and mental health and wellbeing, data on which were only collected once the schools had re-opened in 2021. By focusing on pupils' learning at one time point, we were able to include data from the younger cohort of Ethiopian children to investigate if and how SEL relationships may vary at the different levels of primary education.

In order to investigate these associations, we start by examining changes in numeracy among the younger cohort, as pupils moved from Grade 1 to 3. We then consider the correlations between each aspect of SEL, disaggregated by learner cohort, and the associations between pupils' SEL and their numeracy according to their gender, location and family wealth tercile. Finally, we use multivariate regressions to explore the extent to which different elements of pupils' SEL predict their numeracy, once individual, household and school factors are taken into account.

Before investigating the relationship between children's SEL and numeracy in 2021, it is useful to understand how their numeracy changed between 2019 and 2021, during the period of the school closures. Building on the analyses for Grade 4/6 numeracy set out above, which identified smaller than expected improvements between 2019 and 2021 (see Figures 4 and 5), Figure 6 charts the numeracy scores for the students tracked from Grade 1 to 3 between 2019 and 2021. Table 17 also sets out their IRT numeracy scores, disaggregated by gender and location, and the results of *t*-tests between them. Together they show overall gains in learning, but of much smaller magnitude than would have been expected in the absence of school closures (Kim et al., 2021). On average, boys achieved higher scores than girls at both time points, with such difference being significant in the 2021 data. There are also

significant differences according to pupils' location in both 2019 and 2021, and compared with the trajectory for learners moving from Grade 4 to 6 in Figure 4, the chart displays a widening gap between urban and rural learners, and a near stagnation in progress for the latter.

Figure 6 – Changes in Grade 1/3 IRT Numeracy Scores between 2019 and 2021

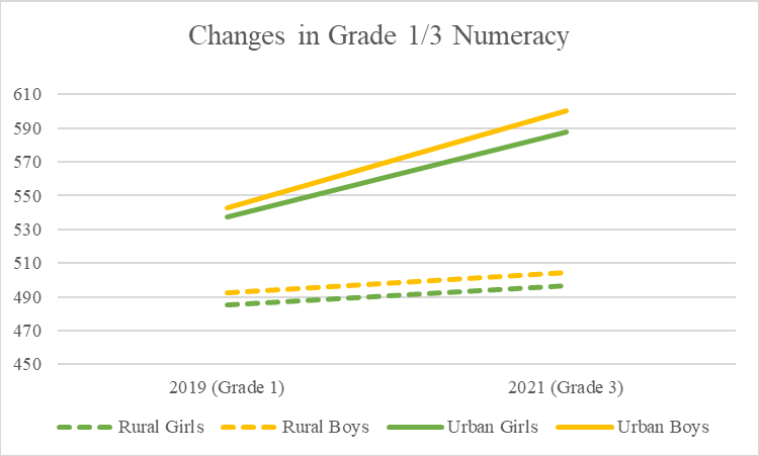


Table 17 – Mean IRT Scores and *T*-Test Results for Grade 1/3 Differences in Numeracy in 2019 and 2021

Measure	Girls	Boys	*t*-value	Rural	Urban	*t*-value
2019 Numeracy	504.88	511.48	-1.80	488.52	539.91	-14.28***
2021 Numeracy	531.06	541.16	-2.16*	500.13	593.98	-21.48***
Change in Numeracy	26.19	29.68	-0.78	11.61	54.07	-9.31***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. $n = 2{,}111$.

Regarding correlations, Table 18 sets out Pearson's coefficients for the SEL variables, disaggregated by cohort. In each case, the correlations are positive and statistically significant at the 0.10 percent level, but are generally stronger for the Grade 6 learners. Within measures, the correlations are largest between social skills and mental health and wellbeing, and between children's self-efficacy and emotional regulation. This could highlight important associations and groupings within different aspects of children's SEL which could have implications for their wider behaviours and support for learning.

Table 18 – Correlations between Different Aspects of SEL in 2021

SEL Measure	Social Skills	Self-Efficacy	Emotional Regulation	Mental Health and Wellbeing
Social Skills	1.00	.34	.36	.57
Self-Efficacy	.69	1.00	.71	.45
Emotional Regulation	.75	.75	1.00	.48
Mental Health and Wellbeing	.79	.65	.71	1.00

Notes: Coefficients in blue above the diagonal indicate correlations for Grade 3 learners, while coefficients in red below the diagonal show Grade 6 correlations. All correlations are significant at the 0.10 percent level.

To explore the relationship between learners' numeracy and different aspects of SEL, we examined correlations between their 2021 scores and responses on the various scales, disaggregated for learners' gender, location and family wealth. Table 19 shows the correlations and their significance for each cohort, adjusted using Bonferroni's correction to reflect the multiple tests (Field, 2009). Overall, the

figures show the strongest associations between numeracy and social skills for both Grades 3 and 6 pupils. Consistent with the high positive correlations between social skills and mental health and wellbeing in Table 18, the latter also display stronger associations with numeracy than either self-efficacy or emotional regulation. The relationships between SEL measures and numeracy are higher for boys than girls among the Grade 3 learners, but the reverse is true for Grade 6 where the associations between SEL and numeracy are stronger for girls. Regarding location and wealth, the positive correlations between SEL and numeracy are largely higher for Grade 3 children from rural settings and poorer households, than for learners in urban areas from wealthier families. By contrast, the patterns of association among pupils' responses at the Grade 6 level are less straightforward with stronger correlations between SEL and numeracy for urban pupils compared with rural learners, and variation by wealth according to the specific SEL measure.

Table 19 – Correlations between 2021 Numeracy and Different Aspects of SEL, Disaggregated by Gender, Location and Wealth

	Social Skills	**Self-Efficacy**	**Emotional Regulation**	**Mental Health and Wellbeing**
Grade 3				
Overall	.25***	.12***	.10***	.19***
Female	.22***	.09*	.04	.15***
Male	.29***	.15***	.15***	.23***
Rural	.23***	.16***	.16***	.21***
Urban	.11*	.01	-.02	.03
Wealth 1 (poorest)	.21***	.17***	.18***	.20***
Wealth 2	.25***	.15***	.15***	.20***
Wealth 3 (richest)	.14***	.04	.03	.09
Grade 6				
Overall	.32***	.20***	.19***	.26***
Female	.35***	.21***	.19***	.27***
Male	.29***	.18***	.19***	.25***
Rural	.26***	.18***	.17***	.20***
Urban	.34***	.20***	.22***	.29***
Wealth 1 (poorest)	.28***	.23***	.20***	.23***
Wealth 2	.27***	.16***	.15***	.17***
Wealth 3 (richest)	.32***	.15***	.17***	.27***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$

To investigate the relationships between numeracy and SEL further, we conducted multivariate regressions to explore the extent to which learners' SEL is associated with their performance in the numeracy assessments, once other factors are taken into account. Given the significant correlations shown in Table 18, we conducted two sets of regressions. We first investigated the extent to which each individual aspect of SEL relates to learners' numeracy in 2021, conditional on their numeracy in 2019 and controlling for other factors (identified as 'separate models'). We then used the same dependent variables but included *all* SEL components in the same regression (identified as the 'combined model'). This allowed us to evaluate the relative contribution of the different aspects of SEL when the influence of other components had been taken into account. Each model also included covariates concerning pupils' individual, household and school characteristics, as described in section 4.1. Table 20 summarises the key coefficients while Appendix 4 sets out the full regression results.

Table 20 – Multivariate Regressions for Grades 3 and 6 Numeracy in 2021

2021 Numeracy	Grade 3		Grade 6	
	Separate Models	Combined Model	Separate Models	Combined Model
2021 Social Skills	0.17*** (0.03)	0.12** (0.04)	0.17*** (0.02)	0.15*** (0.04)
2021 Self-Efficacy	0.10* (0.04)	0.04 (0.05)	0.12*** (0.02)	0.01 (0.04)
2021 Emotional Regulation	0.08* (0.03)	-0.02 (0.04)	0.11*** (0.02)	-0.05 (0.04)
2021 Mental Health and Wellbeing	0.14*** (0.03)	0.06* (0.03)	0.15*** (0.02)	0.06 (0.03)
Observations	2,086	2,086	2,157	2,157

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. The *z*-score for Grade 6 Social Skills in 2021 is derived using the full scale of 12 items, rather than the common seven items used in both 2019 and 2021, as was required for RQ1 longitudinal analyses.

Looking at the different measures of SEL separately, including pupils' mental health and wellbeing in accordance with Figure 1, the results for Grade 3 learners show that their responses on each aspect of SEL significantly predict their 2021 numeracy scores. When combining them into one model, however, we find that only social skills and mental health and wellbeing significantly relate to children's 2021 numeracy. The Grade 6 results show a similar pattern. In the separate models, each aspect of SEL predicts learners' 2021 numeracy, but only social skills show a significant association in the combined model.

Beyond the interactions between SEL and numeracy, these results also highlight nuances in the relationship between learners' social skills and their mental health and wellbeing, according to their level of education. In the combined Grade 6 model, mental health and wellbeing cease to be significant, indicating that all variation in pupils' numeracy can be explained by their social skills. Among Grade 3 children, however, mental health and wellbeing show some independent variation that remains significant in the combined model. As such, the results corroborate the figures in Table 18 and confirm a stronger correlation between social skills and mental health and wellbeing for Grade 6 than Grade 3 children.

More broadly, several other factors positively predict pupils' 2021 numeracy. In each model, pupils' class grade and numeracy scores in 2019 significantly predict their 2021 performance, with beta coefficients ranging 0.15-0.31 and 0.24-59 respectively (see Tables 4A and 4B in Appendix 4). Location and family wealth also predict numeracy among Grade 3 learners, 0.27-0.28 for urban children and 0.32-0.34 for the most affluent pupils, while Grade 6 learners taught by more experienced teachers attained slightly but significantly higher numeracy scores (0.01*).

To summarise, the data captured from Ethiopian children in Grades 3 and 6 following the COVID-19 school closures provide consistent evidence of a relationship between different aspects of SEL and numeracy, as measured within the same time point. Learners' social skills in particular show a significant and strong association with their numeracy, which accords with the findings for RQ1 regarding changes in pupils' skills over time. Children's mental health and wellbeing similarly show a relationship with numeracy at the Grade 3 level, but there is more limited evidence of any association between numeracy and either self-efficacy or emotional regulation once other factors are taken into account.

4.3 SEL and Numeracy Among Low Performers

The previous section investigated the interactions between children's SEL and their performance on the 2021 numeracy assessments. In particular, it identified a relationship between pupils' numeracy and their social skills. Given that SEL and mental health and wellbeing may be particularly relevant for lower-performing students, we extend the analysis to focus in particular on this sub-group of children, with the aim of answering RQ2b: '*To what extent does pupils' SEL predict their performance on numeracy following school closures, once other factors are taken into account?*'

To address the question, we identified the bottom quartile of performers in each cohort in 2019 and 2021. Specifically, we examined whether the SEL measures predicted the probability that a pupil was a low performer in 2021, conditional on whether he or she was also in the bottom 25th percentile of overall scores in 2019 (see Carter et al. (2020b) for a similar approach in Ghana). The focus of RQ2b is on learners' numeracy but for this analysis, we also used literacy test scores to mitigate the effects of 'regression to the mean', a statistical phenomenon whereby learners who scored high or low on a preliminary assessment are less likely to attain extreme results on subsequent assessments (for further explanation, see Carter et al. (2020b) and Crawford, Macmillan and Vignoles (2014)). Due to random error, the initial score may be attributed to good or bad 'luck', rather than true academic potential, but by categorising the learners in 2019 using their *literacy* assessments, we were able to minimise the effects of the correlation between numeracy in 2019 and 2021 to reduce any impact of regression to the mean.

Table 21 highlights the key characteristics of the low performers, relative to the remainder of the sample. As expected, children scoring in the bottom quartile commonly came from more rural settings, larger and poorer families, with caregivers who were less likely to have proficient literacy. Gender differences were nevertheless small and low performers tended to be slightly younger than the main sample group.

Table 21 – Descriptive Statistics for 2019 Low Performers and Remaining Sample

Grade 1/3	**Initial low performers n = 467**	**Remaining sample n = 1,379**
Female (%)	51.8	51.0
Average age	9.9 (1.6)	10.0 (1.5)
Pre-school (%)	48.6	58.2
Health (% health)	99.6	99.3
Food consumption (% acceptable)	50.1	54.6
Average time on domestic tasks	1.2 (1.2)	1.1 (1.2)
Wealth tercile – low (%)	33.8	25.7
Wealth tercile – medium (%)	36.2	29.7
Wealth tercile – high (%)	30.0	44.5
Caregiver literacy (% literate)	23.8	30.8
Average household size	5.7 (2.0)	5.6 (1.9)
Rural location (%)	68.5	55.4
Grade 4/6	**Initial low performers n = 501**	**Remaining sample n = 1,653**
Female (%)	48.3	47.4
Average age	12.9 (1.5)	13.1 (1.5)
Pre-school (%)	33.1	46.1
Health (% health)	98.6	99.3
Food consumption (% acceptable)	50.7	53.4
Average time on domestic tasks	1.8 (1.3)	1.6 (1.3)
Wealth tercile – low (%)	37.1	29.7
Wealth tercile – medium (%)	36.5	31.9
Wealth tercile – high (%)	26.4	39.4
Caregiver literacy (%)	22.2	29.1
Average household size	6.0 (2.0)	5.8 (2.0)
Rural location (%)	78.8	58.2

Notes: Standard deviations in brackets.

We used logistic regressions to examine the relationship between pupils' SEL and their poor performance over time. Table 22 presents the summary results of binary logit models for Grades 1 to 3 and 4 to 6 that indicate the probability that a learner was a low performer in the 2021 numeracy assessment, conditional on any low literacy performance in 2019 which was included as an additional explanatory variable (see Appendix 4 for the full results). Similar to the approach for RQ2a described above, we conducted regressions for each aspect of SEL separately and then together in a combined model. The regressions also included covariates for pupils' individual, household and school factors, and in the case of the Grade 4/6 learners, their teachers' years of experience and highest qualifications.

Table 22 – Logistic Regressions for Low Performance in 2021

2021 Low Performance	Grade 1/3		Grade 4/6	
	Separate Models	Combined Model	Separate Models	Combined Model
2021 Social Skills	-0.51*** (0.11)	-0.33* (0.13)	-0.51*** (0.09)	-0.42* (0.18)
2021 Self-Efficacy	-0.49*** (0.10)	-0.33* (0.15)	-0.38*** (0.09)	0.01 (0.16)
2021 Emotional Regulation	-0.37*** (0.09)	0.05 (0.14)	-0.40*** (0.09)	0.08 (0.15)
2021 Mental Health and Wellbeing	-0.41*** (0.08)	-0.12 (0.09)	-0.48*** (0.08)	-0.20 (0.16)
Observations	1,846	1,846	2,154	2,154

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. The *z*-score for Grade 6 Social Skills in 2021 is derived using the full scale of 12 items, rather than the common seven items used in both 2019 and 2021, as was required for RQ1 longitudinal analyses.

The results for the separate models show significant associations between learners' SEL, including their mental health and wellbeing, and their likelihood of low academic performance in 2021. The negative coefficients, in some cases quite large, indicate that children who reported higher SEL in 2021 were *less* likely to be low performers in the numeracy assessments.

The combined model results display fewer significant associations. In both cases, Grade 3 and Grade 6, children's social skills predict their likelihood of low numeracy achievement. Among the younger cohort, however, their reported self-efficacy also relates to their low performance with pupils who agreed with statements like '*I can usually find a way to get what I want*' and '*I can solve most problems if I make enough effort*' being less likely to rank in the bottom quartile for 2021 numeracy.

Beyond SEL, learners' numeracy performance in 2019 is the strongest predictor of their low numeracy in 2021. In each case, the coefficients are positive, significant and large, ranging from 0.69 to 0.78 for Grade 3 children and 1.04 to 1.10 for Grade 6 students (see Tables 4C and 4D in Appendix 4), which indicate that low performers in 2019 were likely to show low attainment again in 2021. The younger learners' gender, age, class grade and region also predict their low performance in all models, with boys, older pupils and students in higher grades less likely to be in the bottom performance quartile (Table 4C). For the Grade 6 learners, age, grade and region are also significantly related and in the combined model pupils whose teachers had university degrees were significantly less likely to be among the low performers in numeracy (-0.90*) (Table 4D).

In summary, the findings show that Ethiopian children's SEL, and particularly their social skills, predict their low attainment on numeracy following the school closures in 2020, even after taking their 2019 numeracy scores and other wider factors into account. Specifically, pupils in both cohorts with higher SEL and especially social skills were less likely to be low performers in 2021, whether or not they scored in the bottom quartile in 2019. This demonstrates the interrelated nature of children's SEL and academic skills and the importance of promoting them both to maximise outcomes in school and life beyond.

4.4 Interactions between Children's SEL and their Mental Health and Wellbeing

As the hypothesised model in Figure 1 shows, key aspects of children's SEL, namely their self-efficacy, emotional regulation and social skills, may also impact on and affect their mental health and wellbeing. Such interactions are particularly relevant following the COVID-19 school closures, when many learners have been isolated from their peers and evidence worldwide suggests growing mental health

issues among children (UNESCO, 2020). For this reason, the study examines the relationship between pupils' SEL and their mental health and wellbeing, taking into account other wider factors and variables, to address RQ3 which asks '*What is the relationship between Ethiopian children's SEL and their mental health and wellbeing, after COVID-19 school closures*?'

Building on the analyses described above, we conducted multivariate regressions to explore the extent to which learners' social skills, self-efficacy and emotional regulation predict their mental health and wellbeing, measured using statements like '*I do things that interest me*' and '*I am proud of my clothes*'. Similar to the models outlined in Tables 20 and 22, we used both separate and combined models, which examined the relationships between pupils' mental health and their SEL, both individually and collectively. We also included the covariates listed in section 4.1 to control for learners' individual, family and school factors. Table 23 summarises the key coefficients while Appendix 4 sets out the full regression results.

Table 23 – Multivariate Regressions for Grades 3 and 6 Mental Health and Wellbeing

2021 Mental Health and Wellbeing	Grade 3		Grade 6	
	Separate Models	Combined Model	Separate Models	Combined Model
2021 Social Skills	0.59*** (0.03)	0.44*** (0.03)	0.77*** (0.02)	0.55*** (0.03)
2021 Self-Efficacy	0.51*** (0.05)	0.16*** (0.03)	0.68*** (0.03)	0.13*** (0.03)
2021 Emotional Regulation	0.50*** (0.04)	0.24*** (0.03)	0.71*** (0.02)	0.19*** (0.03)
Observations	2,855	2,855	2,611	2,611

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$. The *z*-score for Grade 6 Social Skills in 2021 is derived using the full scale of 12 items, rather than the common seven items used in both 2019 and 2021, as was required for RQ1 longitudinal analyses.

In each case, pupils' 2021 social skills, self-efficacy and emotional regulation significantly predict their mental health and wellbeing for both Grade 3 and 6 learners. The high coefficients in the separate models reduced in the combined regressions, but each aspect of SEL individually, uniquely and positively relates to children's mental health and wellbeing. Between them, social skills again show the strongest relationship for both cohorts and across the different models.

Beyond pupils' SEL, none of the covariates display significant or strong associations with the Grade 3 learners' mental health and wellbeing. However, two household variables significantly predict the Grade 6 pupils' responses. In the combined model, Ethiopian children from the wealthiest families (0.10**) and those whose main caregiver could read (0.05*) reported higher overall mental health and wellbeing in 2021 (see Table 4F in Appendix 4). This accords with wider concerns on the growing social inequalities between rich and poor in light of the COVID-19 pandemic (Kim & Rose, 2020).

By way of summary, the findings support the hypothesised model in Figure 1 and indicate significant associations between all aspects of Ethiopian children's SEL as measured in the study and their mental health and wellbeing. Most notably, their social skills are a strong predictor of their affective wellbeing, how they feel about themselves and their lives. At a time when schools have re-opened but the ongoing pandemic requires some families and communities to maintain some distance, these findings highlight the potential relevance of social interactions for children's wider learning and holistic welfare, and the interconnected nature of their overall development.

4.5 Current Practices for Supporting SEL in Ethiopian Classrooms

In addition to the data collected on children's holistic learning, the study also included Grade 6 lesson observations in the Ethiopian schools to examine what teachers are actually doing that might support their pupils' SEL. We used the World Bank *Teach* tool to capture data and answer RQ4 which asks '*To what extent are teachers in Ethiopian classrooms supporting learners' SEL?*'

Teach specifically identifies lesson practices and behaviours that have been identified as nurturing children's autonomy, perseverance and social and collaborative skills. In total, the enumerators observed 138 teachers (21.7 percent female) and ranked them on a 3-point scale according to how much they displayed the relevant behaviours. This enabled us to examine which practices are more or less commonplace in Ethiopian classrooms. We also conducted regressions combining the observation data with the Grade 6 pupils' SEL responses to explore the relationship between teachers' behaviours and their learners' SEL. For these models, we used data from a subsample of the total children, namely 2,292 of the surveyed pupils who were taught by the 138 teachers observed.

Table 24 shows the relevant items and teachers' rankings by percentage, while Figures 7 and 8 present the average ratings for each item and domain respectively, disaggregated by teacher gender. Teachers were observed to be most likely to demonstrate behaviours to foster their pupils' perseverance, followed by autonomy, and then their social and collaborative skills. The majority of teachers, nearly 90 percent, displayed medium to high levels of incidence of acknowledging student efforts, encouraging learners to set goals and conveying a positive attitude towards their challenges. Less than three-quarters of teachers though demonstrated medium to high levels of behaviours to promote pupils' interpersonal skills, peer interactions and learner collaboration.

Female teachers were observed to show more SEL-related practices than their male counterparts and independent *t*-tests summarised in Table 25 reveal that these differences are statistically significant. This finding partially aligns with the results of lesson observations using *Teach* in Rwanda, where female teachers demonstrated significantly more behaviours to encourage learners' autonomy, however, the Rwandan teachers showed fewest practices to promote perseverance overall in contrast to the findings in Ethiopia (Carter & Rose, 2021).

Table 24 – Lesson Observations Demonstration of SEL-Related Practices

SEL Element	Description	Low (%)	Medium (%)	High (%)
Social and Collaborative Skills	Promotes students' collaboration through peer interaction	27.5	55.1	17.4
	Promotes students' interpersonal skills	22.5	58.7	18.8
	Students collaborate with one another through peer interaction	29.0	57.3	13.8
Perseverance	Acknowledges students' efforts	12.3	49.4	38.4
	Has a positive attitude towards students' challenges	7.3	46.4	46.4
	Encourages goal setting	11.6	62.3	26.1
Autonomy	Provides students with choices	18.2	56.5	24.6
	Provides students with opportunities to take on classroom roles	22.5	47.8	29.7
	Students volunteer to participate in the classroom	18.1	46.4	35.5

Figure 7 – SEL-Related Classroom Practices by Teacher Gender

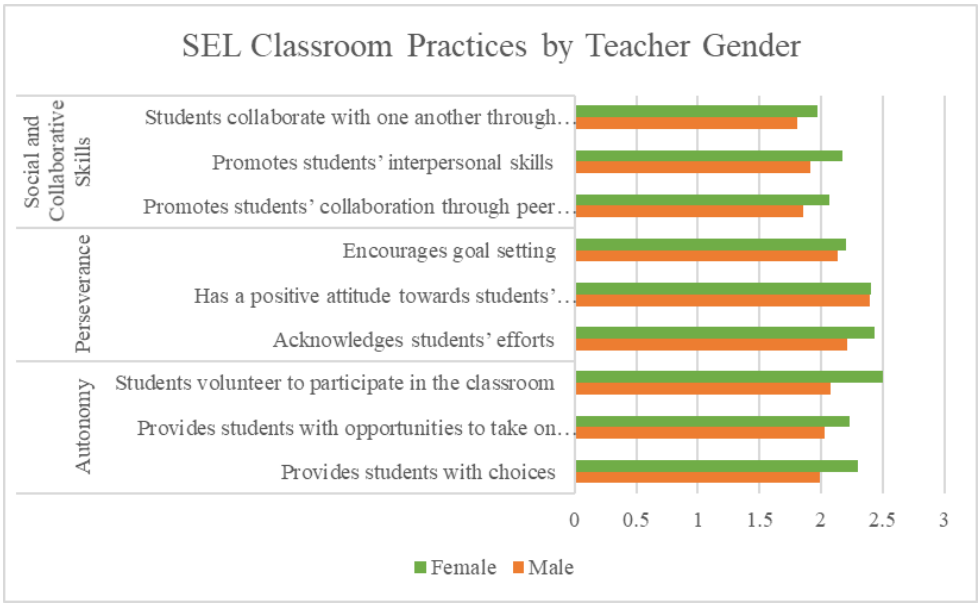


Figure 8 – Overall SEL-Related Classroom Practices by Teacher Gender

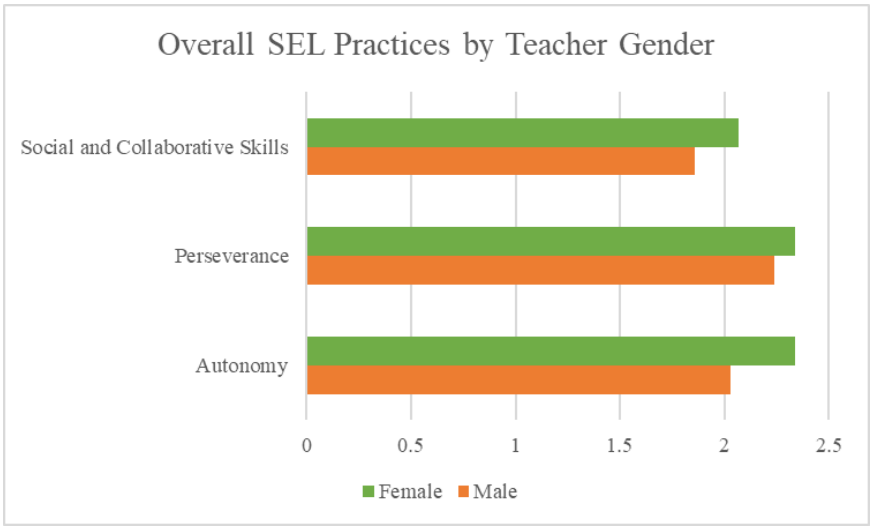


Table 25 – T-Test Results for Difference Between Male and Female Teacher Ratings

SEL Element	Mean Female Rating	Mean Male Rating	*T*-value
Social and Collaborative Skills	2.09	1.87	9.35***
Perseverance	2.37	2.26	5.37***
Autonomy	2.37	2.05	14.37***

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

To investigate the relationship between teachers' classroom behaviours and SEL among the learners' themselves, we conducted a further series of multivariate linear regressions. In each case, the dependent variable was the *z*-score for the average of pupils' responses across the SEL scales, including mental health and wellbeing, while the teachers' average ratings from the lesson observations comprised the main explanatory variable. As for the regressions in Tables 14 and 15 concerning Grade 6 social skills and numeracy, the models were built up step-by-step, to include child, family, teacher and school covariates, before including pupils' 2021 numeracy scores. Learners' grade was omitted because of

collinearity with the teacher and observation variables. The results of the regressions are set out in Table 26.

Overall, the results show no significant association between the teachers' observed practices and their pupils' SEL, including before controlling for other factors (Table 26, model 1). Similarly, even though there are significant differences between teachers' behaviours according to their gender, these do not translate into differences at the learner level (Table 26, model 4 to 6). Only pupils' region (models 5 and 6) and numeracy scores (model 6) (0.18***) significantly predict their average SEL when all factors are taken into account.

These results need to be considered in the light of the limitations of this aspect of the study. First, the observations were limited in terms of including only one teacher per school, typically in mathematics. Second, it is well-known that the 'Hawthorn' effect of being observed can affect teacher behaviours. Even so, we do find variation in the observations, with not all showing a high degree of engagement in SEL practices. Third, for the breakdown by gender of the teacher, only 30 teachers or 21.7 percent were female which could have been too low to offer adequate statistical power. Fourth and importantly, the *Teach* tool focuses on slightly different aspects of SEL from those addressed in the current study and was not adapted to the local context (unlike the measures of SEL used), which could account for the lack of association. Further exploration of the applicability of the *Teach* tool in this context would be valuable.

To summarise, the lesson observations provided valuable evidence of how Ethiopian teachers are fostering pupils' SEL in Grade 6 classrooms. They reveal stronger practices among female teachers and the highest frequency of positive behaviours associated with building children's perseverance, which may be particularly important for those facing increased educational challenges after the COVID-19 school closures (Carter & Rose, 2021). However, the multivariate regressions reveal no relationship between teachers' observed behaviours and their pupils' reported SEL. Further, support for social and collaborative skills was comparatively low, and given their importance for pupils' wider learning and life outcomes, should arguably be afforded greater attention in primary classrooms in the future.

Table 26 – Regressions for Grade 6 2021 SEL (Combined Variable)

2021 SEL		**(1) Basic Model**	**(2) + Child Variables**	**(3) + Family Variables**	**(4) + Teacher Variables**	**(5) + School Location**	**(6) + 2021 Numeracy**
2021 Teacher Observation SEL Score		0.07 (0.06)	0.07 (0.06)	0.05 (0.06)	0.03 (0.06)	0.05 (0.06)	0.06 (0.06)
Student gender (base group: female)		-	0.01 (0.05)	0.02 (0.05)	0.02 (0.05)	0.01 (0.05)	-0.01 (0.05)
Age		-	0.02 (0.02)	0.02 (0.02)	0.02 (0.02)	0.00 (0.02)	0.00 (0.02)
Caregiver Literacy (base group: no/low literacy)		-	-	-0.06 (0.07)	-0.05 (0.06)	0.00 (0.06)	-0.01 (0.05)
Family Wealth (base group: first tercile (poorest))	Second Tercile	-	-	0.16 (0.10)	0.15 (0.11)	0.11 (0.09)	0.13 (0.09)
	Third Tercile	-	-	0.34** (0.11)	0.30* (0.12)	0.21 (0.11)	0.19 (0.11)
Household Size		-	-	-0.02 (0.02)	-0.02 (0.01)	0.00 (0.01)	0.00 (0.01)
Teacher Experience		-	-	-	0.01 (0.01)	0.00 (0.01)	0.00 (0.01)
Teacher Qualification (base group: no training)	Certificate	-	-	-	-0.02 (0.27)	0.25 (0.21)	0.16 (0.22)
	Diploma	-	-	-	0.12 (0.09)	0.17** (0.05)	0.15 (0.08)
	Degree	-	-	-	0.08 (0.18)	0.07 (0.17)	0.03 (0.17)
Observed teacher gender (base group: female)		-	-	-	-0.16 (0.08)	-0.08 (0.07)	-0.07 (0.07)
Location (base group: rural)		-	-	-	-	-0.13 (0.16)	-0.18 (0.15)
Region (base group: Addis Ababa)	Amhara	-	-	-	-	-0.06 (0.15)	0.00 (0.15)
	Benishangul Gumuz	-	-	-	-	-0.85*** (0.24)	-0.58* (0.24)
	Oromia	-	-	-	-	-0.12 (0.15)	-0.02 (0.15)
	SNNP	-	-	-	-	-0.33 (0.25)	-0.18 (0.24)
	Somali	-	-	-	-	-0.33 (0.37)	0.56 (0.35)
2021 Numeracy		-	-	-	-	-	0.18*** (0.03)
Constant		0.04 (0.06)	-0.17 (0.32)	-0.29 (0.35)	-0.36 (0.34)	0.00 (0.28)	0.04 (0.29)
Observations		1,986	1,986	1,986	1,986	1,986	1,986
R-squared		.01	.01	.03	.04	.11	.14

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

5. Discussion

This study has examined key aspects of children's SEL in Ethiopia, including changes over time, its relationship with numeracy and behaviours by teachers in lessons to foster its development. Drawing on large scale data from across the country, the research offers valuable insights regarding SEL among primary-aged learners, particularly in light of the ongoing COVID-19 pandemic.

First, the longitudinal analyses revealed that Ethiopian pupils' social skills declined significantly between 2019 and 2021, over the period of the COVID-related school closures. As indicated in Figure 3 and Table 10, both boys and girls and learners in rural and urban locations showed a reduction in their social skills. Furthermore, there was no association between individual pupils' reported social skills in 2019 and their responses in 2021. Students who expressed agreement with statements like '*I feel confident talking to others*' and '*I make friends easily*' in 2019 did not necessarily agree with them in 2021. Taken together, these findings are consistent with concerns around learners' worsening social skills following the school closures, which could reflect the adverse effects of social distancing and isolation in response to the pandemic. These have potential implications for schools and education planning in Ethiopia and elsewhere, emphasising the need to promote social skills in addition to academic learning in responses to COVID-19 and other major crises.

Among learners, girls and especially girls living in urban locations reported significantly higher initial social skills. This was followed by steeper declines, although gender-based differences ceased to be significant in 2021 and taking wider child, household and school-related factors into account. Pupils in rural settings also show greater losses in social skills than learners in urban areas, which accords with evidence regarding the differences in support for Ethiopian children's continued learning during the school closures depending on their particular location (Yorke et al., 2020, 2021c). These factors may further intersect, for example, if rural girls were more likely to be kept at home to attend to domestic work responsibilities.

Related to this, pupils' family wealth shows a strong association and learners from the most affluent households indicated the highest social skills. This aligns with the results of a recent study by the OECD (2021) in nine countries[xiv] which found a positive relationship between students' socio-economic status and all aspects of SEL surveyed. Possible reasons for these differences include wealthier and more educated parents making greater investments in their children's SEL, whether through home resources, quality time and support, or encouraging participation in extracurricular activities, or lower-income caregivers having less capacity and fewer opportunities to nurture their children's SEL, especially through distance learning (OECD, 2021). Middle-class parents might also have impressed on their children the importance of diverse competencies beyond academic outcomes like literacy for success in school and their life beyond.

Regarding numeracy, pupils' trajectories between 2019 and 2021 look quite different from those for social skills, with small gains in learning despite the school closures and clear gaps between urban and rural children. Overall, Grade 3 and 6 boys scored significantly higher than girls in 2021, which accords with wider recent findings from the RISE programme in Ethiopia (Tiruneh et al., 2021). However, these differences in numeracy according to learners' gender disappear when controlling for other child, family and school-related factors (see Table 15).

Of particular interest, the research highlights a strong and positive relationship between learners' numeracy and their SEL, in particular their social skills. Such association emerges as significant over time, for both Grades 3 and 6, and even after the main covariates were taken into account. For example, pupils' social skills predict their numeracy performance in 2021 while controlling for other aspects of SEL, like self-efficacy and emotional regulation, and conditional on their performance in 2019.

Various studies have previously examined the relationship between children's social skills and their academic learning, including numeracy, and sought to identify the key causal mechanisms between them. The findings of the present research do not assert causality but highlight positive and potentially multiplicative effects between pupils' SEL and their academic outcomes. This aligns with literature reviewed above, for example, the study by Wolf and McCoy (2019) in Ghana which found that children's academic achievement predicted their social skills, perhaps because higher performance encourages greater self-confidence, -esteem and -efficacy and therefore more prosocial behaviours with their peers. In turn, learners with better social skills may be viewed more favourably by both peers and teachers alike, and therefore able to attract greater attention and support (Liew, 2012).

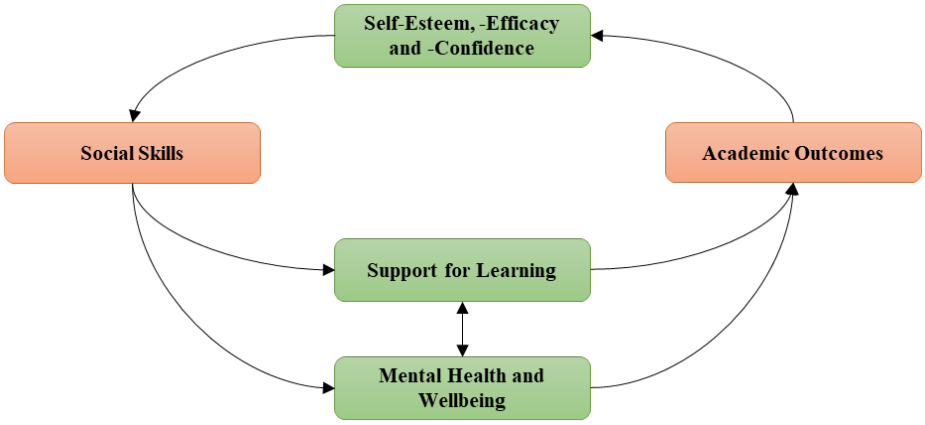
With particular reference to girls' learning, a review of evidence by the Girls' Education Challenge (2018) highlighted the challenge of establishing causality in the relationship between students' SEL and their academic outcomes. Instead, they suggest the relevance of a "softer middle layer of change" (p. 17) concerning learners' motivation, aspiration and ability to make decisions relating to their education. Specifically, they identify the importance of self-esteem for girls to adjust their attitudes, behaviours and capabilities through interventions around attendance, participation and creating an increased sense of belonging.[xv]

Between the literature and findings of the current study, the bidirectional relationship between social skills and academic outcomes like numeracy may therefore be understood as more of a virtuous circle. On the one hand, improvements in reading, mathematics and other school subjects may foster gains in learners' self-esteem, -efficacy and -confidence, their feeling of worth and belief in their capacity to succeed, and therefore their social skills and propensity to engage in prosocial behaviours. On the other hand, increased social skills may enhance children's abilities to build and navigate relationships with their teachers, peers and even family members to access support for their academic learning, both during everyday life and in times of particular crisis.

Findings from the present research and recent Young Lives and OECD studies regarding mental health and wellbeing may provide further insight on these interactions and offer an additional dimension to the hypothesised model. In particular, the cross-sectional analyses reported above revealed strong associations between Ethiopian children's social skills and their mental health and wellbeing, even after controlling for wider factors, which aligned with the results of the OECD (2021) research with 10- and 15-year-olds worldwide. Among the Grade 3 Ethiopian pupils, responses on the mental health and wellbeing scales also significantly predict their numeracy scores. Further, Porter et al. (2021) describe the importance of learners' positive social relationships as protective factors for their mental health during the COVID-19 school closures. Strong bonds with peers were especially significant to mitigate anxiety and depression for the respondents in Ethiopia.

Learners' mental health and wellbeing may therefore comprise another intermediate step between their social skills and academic performance as set out in Figure 9. Specifically, children with better SEL may have advanced coping skills to manage their individual mental health and ensure their capacity for learning (Diamond, 2014). Students with stronger social skills may be more adept at fostering family and social relationships, which can then protect their wellbeing and support their learning during times of crisis. Social skills may also help children and especially girls safeguard the conditions for their continued education, including both their psychological disposition and the avoidance of wider circumstantial threats to their schooling like early marriage, teen pregnancy or sexual exploitation (Ford, et al., 2021; Tiruneh, 2020).

Figure 9 – Emerging Model for the Relationship between Social Skills and Academic Outcomes



Beyond these interactions, the lesson observations shed some light on Ethiopian teachers' practices to foster SEL in the classroom. For instance, they revealed more frequent behaviours to nurture learners' perseverance than their social and collaborative skills. To date, there have been relatively few observation protocols developed to assess primary school teachers' practices regarding SEL in low-income environments. The *Teach* tool represents encouraging progress in this regard but remains limited to classroom behaviours that foster pupils' autonomy, perseverance, critical thinking and social and collaborative skills (Carter et al., 2020a; World Bank, 2019). Future iterations of *Teach* could however be expanded to include indicators for more aspects of SEL, for example, emotional regulation, self-efficacy and adaptability, perhaps providing a menu of options to accommodate research focused on different elements of children's holistic development.

The regression parameters for the longitudinal analyses for Grade 6 learners' social skills and numeracy also offer some insight on the relative ease and difficulty of understanding different aspects of pupils' development. In particular, the R-squared values for the numeracy models are much higher (.51-.58), with the variables explaining over half of the variance in children's 2021 numeracy, compared with those for the social skills models (.11-.15). This suggests that the regressions identified several significant factors that accounted for learners' numeracy, but far fewer that explained their social skills. In which case, there may be other variables that were omitted from the models that relate more closely to children's social skills than their numeracy, which could usefully be explored in future research.

In this regard, the study identified several opportunities to build on and extend the current research in Ethiopia. These could entail expanding the methods used to include direct assessment of SEL where feasible and appropriate, and further exploring the use of classroom observations to identify teaching practices associated with SEL. The latter could involve using multiple observers to investigate inter-rater reliability, conducting observations over several time points and observing more lessons with the same and different teachers in each school. Such an approach would capture wider information on pupils' learning experiences and might offer valuable insights on how Ethiopian education can nurture socio-emotional competencies in the classroom

6. Limitations

The study results and findings must nevertheless be understood in light of its limitations. First, the research design precludes any claims around direct causality. Although the analyses identified factors like wealth and class grade that significantly predict learners' social skills and numeracy respectively, the data prevent us from making assertions concerning their precise causal mechanisms. Related to this, pupils' scores and responses could also have been affected by seasonality and the data collection taking place during different months, May to June in 2019 and January to March in 2021. Based on seasonal trends, we would have expected lower SEL in 2019, as May and June fall in the rainy season when

many rural families and their children are busy with seeding and planting, than in January to March 2021 which include a festive period and typically much greater food availability. Other unobserved seasonal factors could potentially account for the variation but, on face value, the differences are consistent with the adverse effects of the school closures on children's learning and wellbeing.

Second, security concerns prevented the enumerators from returning to schools in Tigray and certain areas of Benishangul Gumuz and Oromia. A strength of the 2019 RISE data collection concerned the geographical spread of the respondents across most regions in Ethiopia. The omission of learners in Tigray and ten other schools raises questions around their SEL, mental health and wellbeing, not just during COVID school closures but also in light of the ongoing violence and conflict.

Several limitations also concern the use of self-reporting measures to assess children's SEL. For instance, more pupils agreed with the scale statements than disagreed with them. This gave rise to a negative skew in many of their distributions (shown in Appendix 2), which suggests the need for some caution in interpreting the results at face value. Learners' ratings also reflected how they *felt* about their SEL on the day of the survey, which could fluctuate over time or vary from their *actual* skills and demonstrable competencies.

Related to this is the question of whether and how children's perceptions and responses might change as they age and mature. The OECD (2021) study, for example, found *lower* SEL among 15-year-olds than 10-year-olds, which could reflect short-term dips and swings in attitudes and abilities, or the learners becoming more self-aware, self-conscious and therefore self-critical as they advance into adolescence (Soto, 2016). There is also the issue of whether they measure themselves against an internal benchmark or through comparison to their peers, which can give to reference group bias. Going forward, this could be addressed by combining self-reporting measures with performance assessments or observations of children's actual behaviours. An alternative would be to make the reference group more explicit in scale items with statements like '*I can control my temper better than my friends'* or '*I can manage my emotions better than other children in my class'*, however this would also have its limitations as pupils may have varying abilities to make such judgments.

Finally, various limitations relate to the use and administration of the *Teach* instrument to observe the Ethiopian classrooms. For example, the SEL components assessed by *Teach*, namely autonomy, perseverance and social and collaborative skills, differed from those addressed in the measures of children's SEL, mental health and wellbeing. Some of the indicators required a high level of inference with observers providing their judgment of teachers' behaviours to foster pupils' SEL. Each lesson was observed by one enumerator which prevented us from calculating any inter-rater reliability and the study only involved one observation per teacher per school.

7. Conclusions and Recommendations

Drawing on data from schools and classrooms across Ethiopia, this research has examined children's SEL, mental health and wellbeing and academic outcomes both before and after the COVID-19 school closures. The study advances current knowledge in several respects: by quantifying the decline in participating Ethiopian pupils' social skills over the period of the closures; by identifying a significant and strong relationship between their social skills and numeracy, even after taking other factors into account; and by revealing the significant association between learners' social skills and their mental health and wellbeing, suggesting the importance of interpersonal relationships and interactions to safeguard children's holistic welfare, even during times of crisis. The study and its findings also have various implications for education planning, policymaking and practice, as well as follow-up research on SEL in low-income settings, both in Ethiopia and elsewhere.

A key finding of the research concerns the interdependency and interconnectedness of children's SEL and their academic learning, as evidenced by the strong and positive relationship between Ethiopian

pupils' social skills and their numeracy. This indicates that SEL must not be addressed in addition to academic outcomes, but in conjunction with them. Education planners, teachers, headteachers and other stakeholders such as parents and guardians need to find solutions to embed and support SEL throughout children's school careers and beyond. Particular strategies could include: establishing dedicated offices within schools to develop approaches for mainstreaming SEL within education, including during crises and closures; designing and implementing specialised teacher training to share practical pedagogies and methods for promoting SEL; reviewing curricula, syllabi and related documents to ensure that they are aligned with teaching approaches to foster SEL throughout primary and secondary education, which could include specifying standards and strategies for integrating SEL within other academic subjects and disciplines, as well as the wider school culture, values and activities; committing additional resources to timetabled extra-curricular activities that nurture SEL, including clubs and safe spaces for girls, ideally also during the vacations; targeting funding to foster SEL and positive mental health and wellbeing among the most marginalised and disadvantaged children, not least given their relationship with socio-economic status and implications for their academic learning; and increasing support for teachers' mental health and wellbeing, particularly in light of COVID, so that they model effective SEL practices themselves and create safe and nurturing environments for their learners. In each case, any approaches or interventions should ensure that they are both gender sensitive and age-appropriate.

In Ethiopia specifically, the current guidelines identify pre-primary education as the main forum in which to build children's SEL (Teferra et al., 2018). Kindergarten scheduling for instance allocates up to 2 hours of class time for 'Personal and Socio-emotional Development' each week (Federal Ministry of Education, 2020b). However, this raises questions regarding learners who missed pre-primary schooling, how to weave SEL into more aspects of Ethiopian education and how to ensure that all students leave school with at least a minimum level of SEL. Indeed, the findings revealing the importance of SEL provide support for arguments in favour of 'self-contained classrooms' whereby children, at least in the lower primary years, learn all subjects with one teacher who can also monitor their broader growth and holistic development.[xvi]

Finally, the study and this paper propose a model for understanding the relationships between children's SEL and their academic learning, being set out in Figure 9. A key component of that model concerns the inclusion of some assessment of learners' self-esteem, self-efficacy or self-confidence, which allowed us to control for how they felt about themselves generally, in addition to how they viewed their SEL in any particular respect. Going forward, research using self-reporting scales to investigate learners' SEL should therefore include at least one measure of their self-esteem, -efficacy and -confidence, and future studies could usefully test the hypothesised model in Ethiopia or other international contexts.

To conclude, the research has highlighted the importance of children's SEL and social skills, and the role that schools can play in building valuable competencies over and above pupils' academic learning. Against the backdrop of COVID-19, children's mental health and wellbeing have also come into focus and emphasised the role of SEL to equip learners with the necessary tools to navigate any difficulties, crises or emergencies in the future. For now, however, as education systems contend with the ongoing challenges posed by the pandemic, the need for diverse skills and integrated learning for all children and at all levels of schooling is ever more apparent.

References

Abimpaye, M., Dusabe, C., Nzabonimpa, J. P., Ashford, R. & Pisani, L. (2019). Improving parenting practices and development for young children in Rwanda: Results of a randomized control trial. *International Journal of Behavioral Development*, 44(3), 205-215.

Appleton, J.J., Christenson, S.L., Kim, D. & Reschly, A.L. (2006). Measuring cognitive and psychological engagement: Validation of the Student Engagement Instrument. *Journal of School Psychology*, 44, 427-445. DOI: 10.1016/j.jsp.2006.04.002.

Assessment Work Group (2019). *Student social and emotional competence assessment: The current state of the field and a vision for its future*. Collaborative for Academic, Social, and Emotional Learning (CASEL). Chicago, IL. Available at https://measuringsel.casel.org/, accessed on 31 March 2021.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191–215. DOI: 10.1037/0033-295X.84.2.191.

Bayley, S., Wole, D., Yorke, L., Ramchandani, P. & Rose, P. (2021). *Researching Socio-Emotional Learning, Mental Health and Wellbeing: Methodological Issues in Low-Income Contexts. RISE Working Paper Series. 21/068.* https://doi.org/10.35489/BSG-RISE-WP_2021/068.

Borisova, I., Pisani, L., Dowd, A. J. & Lin, H.-C. (2017). Effective interventions to strengthen early language and literacy skills in low-income countries: comparison of a family-focused approach and a pre-primary programme in Ethiopia. *Early Child Development and Care*, 187(3-4), 655-671.

Borzekowski, D. L. G., Singpurwalla, D., Mehrotra, D. & Howard, D. (2019). The impact of *Galli Galli Sim Sim* on Indian preschoolers. *Journal of Applied Developmental Psychology*, 64, 101054.

Burman, E. (2008). *Developments: Child, Image, Nation*. London: Routledge.

Camfield, L. (2012). Resilience and Well-being Among Urban Ethiopian Children: What Role Do Social Resources and Competencies Play? *Social Indicators Research*, 107, 393-410. DOI: 10.1007/s11205-011-9860-3.

Camfield, L. & Tafere, Y. (2009). 'No, living well does not mean being rich': Diverse understandings of well-being among 11-13-year-old children in three Ethiopian communities. *Journal of Children and Poverty*, 15(2), 119-138. DOI: 10.1080/10796120903310889.

Carter, E., Molina, E., Pushparatnam, A. & Rose, P. (2020a). *Measuring teachers' encouragement of socioemotional skills in the secondary classroom*. https://drive.google.com/file/d/1brT5seQiPWfW82MtaZSEcG_4m7ULFN1W/view

Carter, E., Rose, P., Sabates, R. & Akyeampong, K. (2020b). Trapped in low performance? Tracking the learning trajectory of disadvantaged girls and boys in the Complementary Basic Education Programme in Ghana, *International Journal of Educational Research*, 101541. DOI: 10.1016/j.ijer.2020.101541.

Carter, E., & Rose, P. (2021). *Teacher practices in Rwandan secondary mathematics classrooms: Findings from classroom observations*. Leaders in Teaching Research and Policy Series, March 2021. Laterite, Rwanda and REAL Centre, University of Cambridge.

Cherinet, F. (2001). *Issues Surrounding Self-Contained Classroom in Primary Schools in Ethiopia*. [Unpublished Master's Thesis]. Addis Ababa University.

Collaborative for Academic, Social, and Emotional Learning (CASEL). (2020) *CASEL's SEL Framework: What Are the Core Competence Areas and Where Are They Promoted?* Available at https://casel.org/wp-content/uploads/2020/12/CASEL-SEL-Framework-11.2020.pdf.

Crawford, C., Macmillan, L. & Vignoles, A. (2014). *Progress made by high-attaining children from disadvantaged backgrounds: Research report*. London: Social Mobility and Child Poverty Commission.

Danielson, C.K. & Phelps, C.R. (2003). The Assessment of Children's Social Skills Through Self-Report: A Potential Screening Instrument for Classroom Use. *Measurement and Evaluation in Counseling and Development*, 35(4), 218-229. DOI: 10.1080/07481756.2003.12069068.

Dawis, R. (2000). 'Scale Construction and Psychometric Considerations', in H.E. Tinsley and S.D. Brown (eds.) *Handbook of Applied Multivariate Statistics and Mathematical Modeling*, 65-94, San Diego, CA: Academic Press.

Debebe, A. (2002). The Self-Contained Classroom Set-Up: What it is and What it is not. *IER Flambeau*, 10(1), 43-56.

Dercon, S., & Krishnan, P. (2009). Poverty and the socio-emotional competencies of children: evidence from the young lives sample in four developing countries. *Children Youth and Environments*, 19(2), 138-163.

Diamond, A. (2014). Want to Optimize Executive Functions and Academic Outcomes?: Simple, Just Nourish the Human Spirit. *Minnesota Symposia on Child Psychology*, 37, 205. Available at http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4210770/.

Duckworth, A. L., & Yeager, D. S. (2015). Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes. *Educational Researcher*, 44(4), 237-251.

Dunn, K. J. & McCray, G. (2020). The Place of the Bifactor Model in Confirmatory Factor Analysis Investigations Into Construct Dimensionality in Language Testing. *Frontiers in Psychology*, 11, 1357. DOI: 10.3389/fpsyg.2020.01357.

Espinoza Revollo, P. & Ogando Portela, M.-J. (2019). *Self-efficacy, Agency and Empowerment During Adolescence and Young Adulthood in Ethiopia, India, Peru and Vietnam*. Working Paper 184. Oxford: Young Lives.

Favara, M., Freund, R., Porter, C., Sanchez, A. & Scott, D. (2021). Young lives, interrupted: Short-term effects of the COVID-19 pandemic on adolescents in low- and middle-income countries. *Covid Economics*, 67, 172-188.

Federal Democratic Republic Government of Ethiopia. (1994). *Education and Training Policy*, 1st ed. Available at http://www.moe.gov.et/PoliciesStrategies, accessed on 28 September 2021.

Federal Democratic Republic of Ethiopia Ministry of Education. (2015). *Education Sector Development Program V (ESDP V): 2008-2012 E.C. 2015/6-2019/20 G.C. Program Action Plan*. Available at http://www.moe.gov.et/PoliciesStrategies, accessed on 28 September 2021.

Federal Democratic Republic of Ethiopia Ministry of Education. (2020a). *Education Statistics Annual Abstract: September 2019-March 2020*. Addis Ababa: Federal Ministry of Education.

Federal Ministry of Education. (2020b). *The Federal Democratic Republic of Ethiopia General Education Curriculum Framework*. Addis Ababa: Ministry of Education.

Field, A. (2009). *Discovering Statistics Using SPSS* (3rd ed.). London: SAGE Publications Ltd.

Ford, K., Porter, C. & Pankhurst, A. (2021). *A Lost Year of Learning for Girls in Ethiopia: Evidence From the Young Lives at Work COVID-19 Phone Survey*. Policy Brief 44. Young Lives: Oxford.

Forsberg, J. T., Schultz, J-H., Lodi, C. & Tubbs Dolan, C. (2019). *Development of the Student Learning in Emergency Checklist (26) (SLEC-26): A measurement tool of promoters and barriers for learning among conflict-affected students in Palestine*. Technical working paper. Tromsø, Norway: University of Tromsø.

Girls' Education Challenge (GEC). (2018). *Thematic Review: Girls' Self-Esteem*. Available online at: https://girlseducationchallenge.org/media/hl2d344o/thematic-review-girls-self-esteem.pdf, accessed on 8 September 2021.

Goldschmidt, T. & Pedro, A. (2019). Early childhood socio-emotional development indicators: Pre-school teachers' perceptions. *Journal of Psychology in Africa*, 29(5), 474-479.

Himaz, R. (2009). *Working Paper No. 44: The Impact of Parental Death on Schooling and Subjective Well-being: Evidence from Ethiopia using Longitudinal Data*. Oxford: Young Lives. Available at https://ora.ox.ac.uk/objects/uuid:7ee11828-d172-4289-b1f5-40cbb125f978, accessed on 8 September 2021.

Hoddinott, J., Iyer, P., Sabates, R., & Woldehanna, T. (2019). *Evaluating Large-Scale Education Reforms in Ethiopia*. Available at https://riseprogramme.org/publications/evaluating-large-scale-education-reforms-ethiopia, accessed on 31 March 2021.

Homan, R. (2001). The Principle of Assumed Consent: The Ethics of Gatekeeping. *Journal of Philosophy of Education*, 35(3), 329-343.

Hu, L. & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal*, 6(1), 1-55.

Huang, Y., & Bentler, P. M. (2015). Behaviour of asymptotically distribution free test statistics in covariance versus correlation structure analysis. *Structural Equation Modelling: A Multidisciplinary Journal*, 22(4), 489-503.

Immordino-Yang, M. H., & Damasio, A. (2007). We feel, therefore we learn: The relevance of affective and social neuroscience to education. *Mind, brain, and education*, 1(1), 3-10.

Inter-Agency Network for Education in Emergencies (INEE). (2016). *INEE Background Paper on Psychosocial Support and Social and Emotional Learning for Children and Youth in Emergency Settings*. Available at https://resourcecentre.savethechildren.net/node/12312/pdf/334._inee_background_paper_pss_and_sel_for_children_and_youth_in_emergency_settings.pdf, accessed 9 February 2021.

Jones, S. M. & Doolittle, E. J. (2017) Social and Emotional Learning: Introducing the Issue. *Spring*, 27(1), 3-11.

Jukes, M. C. H., Gabrieli, P., Loti Mgonda, N., Nsolezi, F. S., Jeremiah, G., Tibenda, J. J. & Bub, K. L. (2018). "Respect is an investment": Community Perceptions of Social and Emotional Competencies in Early Childhood from Mtwara, Tanzania. *Global Education Review*, 5(2), 160-188.

Kaiser, H. F. (1960) The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, 20(1), 141-151.

Kim, J. & Rose, P. (2020, May 4). *The Threat of COVID-19 on Ethiopia's Recent Gains in Pre-primary Education*. [Web blog post] Available at https://riseprogramme.org/blog/COVID-19-ETH-Pre-Primary-Ed accessed on 6 September 2021.

Kim, J., Rose, P., Tiruneh, D., Sabates, R. & Woldehanna, T. (2021, May 4). *Learning Inequalities Widen Following COVID-19 School Closures in Ethiopia*. [Web blog post] Available at https://riseprogramme.org/blog/learning-inequalities-widen-COVID-19-Ethiopia, accessed on 15 July 2021.

Krosnick, J. & Presser, S. (2010) Question and Questionnaire Design. In J.D. Wright and P.V. Marsden (Eds.), *Handbook of Survey Research* (pp. 263-313). San Diego, CA: Elsevier.

Leung, S. (2011) A Comparison of Psychometric Properties and Normality in 4-, 5-, 6-, and 11-Point Likert Scales. *Journal of Social Science Research*, 37(4), 412-421.

Liew, J. (2012). Effortful Control, Executive Functions, and Education: Bringing Self-Regulatory and Social-Emotional Competencies to the Table. *Child Development Perspectives*, 6(2), 105-111.

Liu, W.C., Wang, C.K.J. & Parkins, E.J. (2005). A longitudinal study of students' academic self-concept in a streamed setting: The Singapore context. *British Journal of Educational Psychology*, 75, 567-586. DOI:10.1348/000709905X42239.

Liu, Y. & Zumbo, B. D. (2012). Impact of Outliers Arising From Unintended and Unknowingly Included Subpopulations on the Decisions About the Number of Factors in Exploratory Factor Analysis. *Educational and Psychological Measurement*, 72(3), 388-414.

Matson, J.L., Rotatori, A.F. & Helsel, W.J. (1983). Development of a Rating Scale to Measure Social Skills in Children: The Matson Evaluation of Social Skills with Youngsters (MESSY). *Behaviour, Research and Therapy*, 21(4), 335-340.

Matsunaga, M. (2010). How to Factor-Analyze Your Data Right: Do's, Don'ts, and How-To's. *International Journal of Psychological Research*, 3(1), 97-110.

Melese, S. & Tadege, A. (2019). The Ethiopian curriculum development and implementation vis-à-vis Schwab's signs of crisis in the field of curriculum. *Cogent Education*, 6(1), 1633147.

National Scientific Council on the Developing Child. (2014). *Excessive Stress Disrupts the Architecture of the Developing Brain: Working Paper No. 3*. Updated Edition. Available at www.developingchild.harvard.edu, accessed on 8 September 2021.

Organisation for Economic Co-operation and Development (OECD). (2021). *Beyond Academic Learning: First Results from the Survey of Social and Emotional Skills*. OECD Publishing: Paris. DOI: 10.1787/92a11084-en.

Pisani, L., Borisova, I., & Dowd, A. J. (2018). Developing and validating the International Development and Early Learning Assessment (IDELA). *International Journal of Educational Research*, 91, 1-15. DOI: 10.1016/j.ijer.2018.06.007.

Porter, C., Favara, M., Hittmeyer, A., Scott, D., Sanchez Jimenez, A., Ellanki, R., Woldehanna, T.,& Stein, A. (2021). Impact of the COVID-19 pandemic on anxiety and depression symptoms of young

people in the global south: evidence from a four-country cohort study. *BMJ Open*, e049653. DOI: 10.1136/bmjopen-2021-049653.

Raikes, A., Yoshikawa, H., Britto, P. R., & Iruka, I. (2017). Children, Youth and Developmental Science in the 2015-2030 Global Sustainable Development Goals. *Social Policy Report*, 30(3), 1-23. DOI: https://doi.org/10.1002/j.2379-3988.2017.tb00088.x.

Raikes, A., Koziol, N., Janus, M., Platas, L., Weatherholt, T., Smeby, A. & Sayre, R. (2019). Examination of school readiness constructs in Tanzania: Psychometric evaluation of the MELQO scales. *Journal of Applied Developmental Psychology*, 62, 122-134.

Rossiter, J., Hagos, B., Rose, P., Teferra, T. & Woldehanna, T. (2018). *Early Learning in Ethiopia: Equitable access and Learning. System Diagnostic Report for World Bank Early Learning Program*. https://doi.org/10.5281/zenodo.3371317.

Schulz, W. (2005, April). *Measuring the socio-economic background of students and its effect on achievement in PISA 2000 and PISA 2003*. Paper presented at Annual Meetings of the American Educational Research Association. Available at https://eric.ed.gov/?id=ED493510.

Soto, C. J. (2016). The Little Six Personality Dimensions from Early Childhood to Early Adulthood: Mean-Level Age and Gender Differences in Parents' Reports. *Journal of Personality*, 84(4), 409-422.

Teferra, T., Asgedom, A., Oumer, J., Woldehanna, T., Dalelo, A. & Assefa, B. (2018). *Ethiopian Education Development Roadmap (2018-30): An integrated Executive Summary*. Addis Ababa: Ministry of Education Education Strategy Center (ESC).

Tiruneh, D. T. (2020). *COVID-19 School Closures May Further Widen the Inequality Gaps between the Advantaged and the Disadvantaged in Ethiopia*. [Web blog post] Available at https://riseprogramme.org/blog/COVID-19_ethiopia_school_closures, accessed on 8 September 2021.

Tiruneh, D. T., Hoddinott, J., Rolleston, C., Sabates, R. & Woldehanna, T. (2021b). *Understanding Achievement in Numeracy Among Primary School Children in Ethiopia: Evidence from RISE Ethiopia Study*. *RISE Working Paper Series. 21/071*. https://doi.org/10.35489/BSG-RISE-WP_2021/071.

Toering, T., Elferink-Gemser, M.T., Jonker, L., van Heuvelen, M.J.G. & Visscher, C. (2012). Measuring self-regulation in a learning context: Reliability and validity of the Self-Regulation of Learning Self-Report Scale (SRL-SRS). *International Journal of Sport and Exercise Psychology*, 10(1), 24-38, DOI: 10.1080/1612197X.2012.645132.

Tran, T. D., Holton, S., Nguyen, H. & Fisher, J. (2019). Physical growth: is it a good indicator of development in early childhood in low- and middle-income countries? *BMC Pediatrics*, 19(1), 276.

Tubbs Dolan, C. (2019). *Psychometric Analysis of the Pilot Holistic Assessment for Learning (HAL): Validity and Reliability*. New York: Global TIES for Children, NYU.

United Nations Development Programme. (2020). *Human Development Report 2020: The next frontier – Human development and the Anthropocene*. Available at http://hdr.undp.org/sites/default/files/hdr2020.pdf, accessed on 28 September 2021.

United Nations Educational, Scientific and Cultural Organization (UNESCO) (2020). *UNESCO COVID-19 Education Response: Education Sector issue notes. Issue note no 1.2 – April 2020:*

Nurturing the social and emotional wellbeing of children and young people during crises. Available at https://unesdoc.unesco.org/ark:/48223/pf0000373271.

United Nations Educational, Scientific, and Cultural Organization (UNESCO). (n.d.). *Education: From disruption to recovery*. Available at https://en.unesco.org/covid19/educationresponse, accessed on 3 February 2021.

Van Sonderen, E., Sanderman, R., & Coyne, J. C. (2013). Ineffectiveness of reverse wording of questionnaire items: Let's learn from cows in the rain. *PloS one*, 8(7), e68967.

Watts, S. E., Oburu, P., Lah, S., Hunt, C. J. & Rhodes, P. (2016). Psychosocial Stimulation: A Qualitative Study on Kenyan Mother's Motives and Challenges to Promote Children's Development. *Journal of Child and Family Studies*, 25, 1840–1847.

Weems, G. H., Onwuegbuzie, A. J., & Collins, K. M. (2006). The role of reading comprehension in responses to positively and negatively worded items on rating scales. *Evaluation & Research in Education*, 19(1), 3-20.

Weldesilassie, A., Hoddinott, J. & Woldehanna, T. (2020). *Survey Attrition in the RISE Ethiopia Baseline Quantitative Surveys*. Unpublished.

Wolf, S., Halpin, P., Yoshikawa, H., Dowd, A.J., Pisani, L. & Borisova, I. (2017). Measuring school readiness globally: Assessing the construct validity and measurement invariance of the International Development and Early Learning Assessment (IDELA) in Ethiopia, *Early Childhood Research Quarterly*, 41, 21-36.

Wolf, S. & McCoy, D.C. (2019). The role of executive function and social-emotional skills in the development of literacy and numeracy during preschool: a cross-lagged longitudinal study. *Developmental Science*, 22:e12800. DOI: 10.1111/desc.12800.

World Bank. (2019). *Teach: Helping Countries Track and Improve Teaching Quality*. Online brief available at https://www.worldbank.org/en/topic/education/brief/teach-helping-countries-track-and-improve-teaching-quality?cid=EXT_WBEmailShare_EXT, accessed on 14 July 2021.

World Bank. (2021). *Ethiopia Overview* [web page]. Available at https://www.worldbank.org/en/country/ethiopia/overview, accessed on 28 September 2021.

World Health Organization (WHO) Collaborating Centre in Mental Health. (1998). *WHO (Five) Well-Being Index (1998 Version)*. Available at https://www.psykiatri-regionh.dk/who-5/Documents/WHO-5%20questionaire%20-%20English.pdf, accessed on 9 February 2021.

World Health Organization (WHO). (2018). *Mental Health: Strengthening Our Response*. World Health Organisation. Available at https://www.who.int/news-room/fact-sheets/detail/mental-health-strengthening-our-response, accessed on 2 February 2021.

Yorke, L. & Ogando, M. J. (2018). *Psychosocial scales in the Young Lives Round 4 Survey: Selection, adaptation and validation*. Available at https://assets.publishing.service.gov.uk/media/5b9a89a240f0b6786cf0f4fa/YL-TN45.pdf, accessed on 9 February 2021.

Yorke, L., Rose, P., Hagos, B. & Woldehanna, T. (2020). *The effects of COVID-19 on primary education in Ethiopia: Perspectives of school principals and teachers*. Research and Policy Paper No. 20/10. REAL Centre, University of Cambridge.

Yorke, L., Wole, D. & Rose, P. (2021a). *An Emerging Strategy for the Development of Culturally Relevant Scales to Capture Students' Socio-Emotional Learning*. 2021/031. https://doi.org/10.35489/BSG-RISE-RI_2021/031.

Yorke, L., Rose, P., Bayley, S., Wole, D. and Ramchandani, P. (2021b). *The Importance of Students' Socio-Emotional Learning, Mental Health and Wellbeing in the time of COVID-19. 2021/025*. https://doi.org/10.35489/BSG-RISE-RI_2021/025.

Yorke, L., Rose, P., Woldehanna, T. & Hailu, B. H. (2021c). Primary School-Level Responses to the COVID-19 Pandemic in Ethiopia: Evidence from Phone Surveys of School Principals and Teachers. *Perspectives in Education*, 39(1), 189-206.

Zins, J. E. & Elias, M. J. (2007). Social and Emotional Learning: Promoting the Development of All Students. *Journal of Educational and Psychological Consultation*, 17:2-3, 233-255, DOI: 10.1080/10474410701413152.

Appendix 1: Final Instruments Used

These questions are about things you did when schools were closed. There are no right or wrong answers.

1. What was the <u>main</u> activity that you did during the day when schools were closed because of COVID-19? (*Choose only one answer and put a "√" mark in front of it.)*		
Staying at home playing		
Staying at home reading textbooks or other books		
Staying at home watching educational programmes on TV		
Staying at home listening to educational programmes on the radio		
Staying at home helping		
Playing outside home		
Working outside home		
2. Did you continue learning at home when schools were closed because of COVID-19?	Yes	No
[IF YES]		
2a. What type of learning activities did you do? (C*hoose as many as you like*)		
I met with my teacher	Yes	No
I listened to educational programmes on the radio	Yes	No
I watched educational TV programmes	Yes	No
I used online/mobile learning apps	Yes	No
I completed work given to me by the teacher	Yes	No
I completed schoolwork given to me by my parents/caregiver	Yes	No
I read books that I got from school	Yes	No
3. Did any of your teachers contact you about schoolwork when schools were closed?	Yes	No
4. Did you receive any books or learning materials from your school when schools were closed?	Yes	No
5. Did you find it difficult to concentrate on your schoolwork when schools were closed?	Yes	No
6. Did anyone in your family help you with your schoolwork when schools were closed?	Yes	No

Please read each sentence carefully and think about how you feel about it. Then respond to each statement. There are no right or wrong answers.

Scale A – Self-Efficacy	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
1. If I persist on a task, I'll eventually succeed.					
2. I know how to handle new situations.					
3. I can usually find a way to get what I want.					
4. I am confident that I could handle unexpected events.					
5. If I am in a difficult situation, I can usually think of something to do.					
6 I remain calm when facing difficulties.					
7. I can solve difficult problems if I try.					
8. It is easy for me to concentrate on my goals and to achieve them.					
9. I can solve most problems if I make enough effort.					
10. When I am confronted with a problem, I usually find solutions.					
11. No matter what comes my way, I'm usually able to handle it.					

Scale B – Emotional Regulation	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
1. I can control my temper.					
2. When I am scared, I can calm myself down.					
3. When I feel angry, I can calm myself down.					
4. I can manage my anger (e.g. I don't hit other people or things).					
5. It is easy for me to stick to my aims.					
6. It is easy for me to achieve my goals.					
7. If I am in trouble, I can think of a solution.					
8. I can handle whatever comes my way.					
9. If something I did not expect happens, I know how to react					
10. I can concentrate when doing schoolwork.					

Scale C – Social Skills	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
1. I feel confident talking to others.					
2. Other people like me.					
3. I like to share things with others.					
4. I help others when they need help.					
5. I make friends easily.					
6. If I hurt someone, I say sorry.					
7. I am polite towards others (e.g. greetings, saying thank you)					
8. I feel good if I help someone.					
9. Other children like to play with me.					
10. I take care of other people's things as if they were my own.					
11. I enjoy co-operating with others.					
12. I respect the opinion of others.					

In answering the following questions, please think about how you have felt over the last two weeks.

Scale D – Mental Health and Wellbeing	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
1. I have felt cheerful and in good spirits					
2. I have felt calm and relaxed					
3. I have felt active and vigorous					
4. I woke up feeling fresh and rested					
5. I do things that interest me					
6. If I am hungry there is enough to eat at home					
7. I am proud of my clothes					
8. I am proud of my shoes (or having shoes)					

Scale E – Family Support	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
1. My family supports my education.					
2. There is at least one person in my family who helps me with my schoolwork.					
3. There is at least one person in my family who encourages me to do well at school.					
4. There is at least one person in my family who asks me about my results in school.					
5. My family encourages me to learn.					

Scale F – Teacher Relationship	Strongly disagree	Disagree	Undecided	Agree	Strongly Agree
1. I feel comfortable asking my teacher questions.					
2. My teacher cares about me.					
3. I trust my teacher.					
4. My teacher treats me the same as other students.					
5. My teacher asks me questions.					
6. My teacher answers my questions.					

Appendix 2: SEL and Numeracy Variable Distributions

Grade 3

Figure 2A – Grade 3 Social Skills Distribution

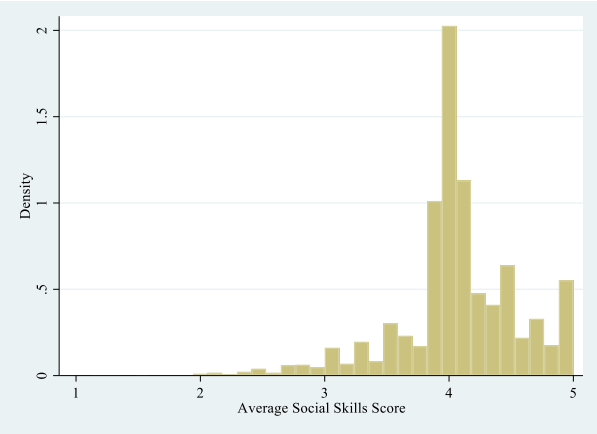


Figure 2B – Grade 3 Self-Efficacy Distribution

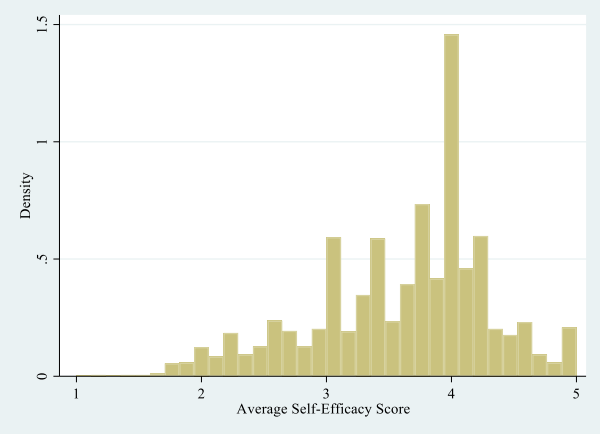


Figure 2C – Grade 3 Emotional Regulation Distribution

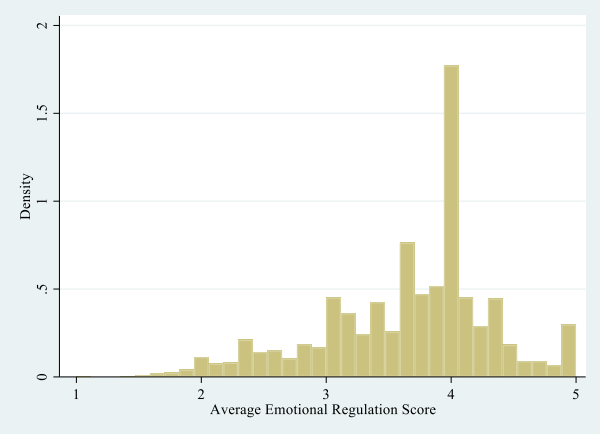


Figure 2D – Grade 3 Mental Health and Wellbeing Distribution

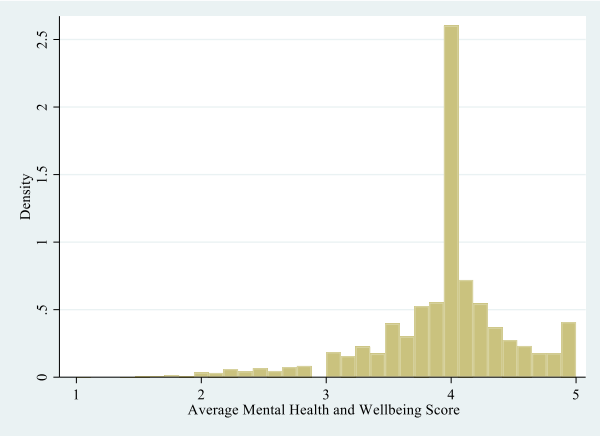
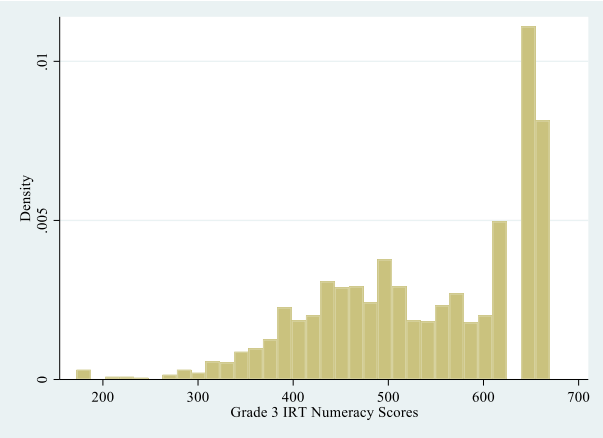


Figure 2E – Grade 3 IRT Numeracy Score Distribution



Grade 6

Figure 2F – Grade 6 Social Skills Distribution

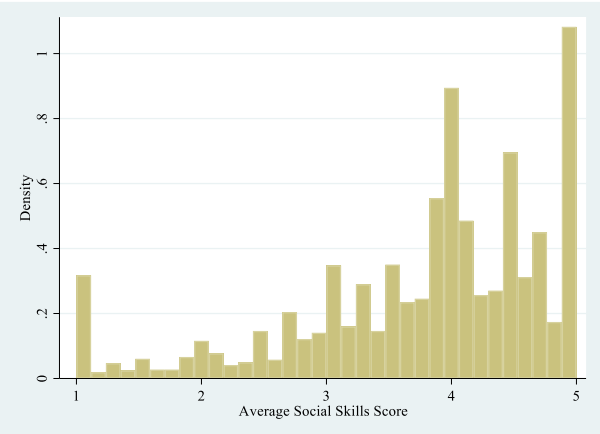


Figure 2G – Grade 6 Self-Efficacy Distribution

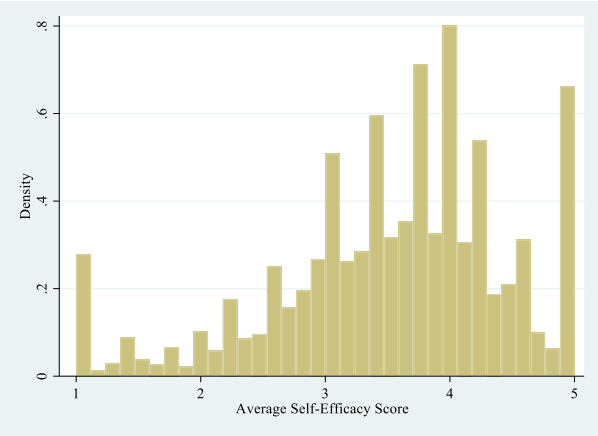


Figure 2H – Grade 6 Emotional Regulation Distribution

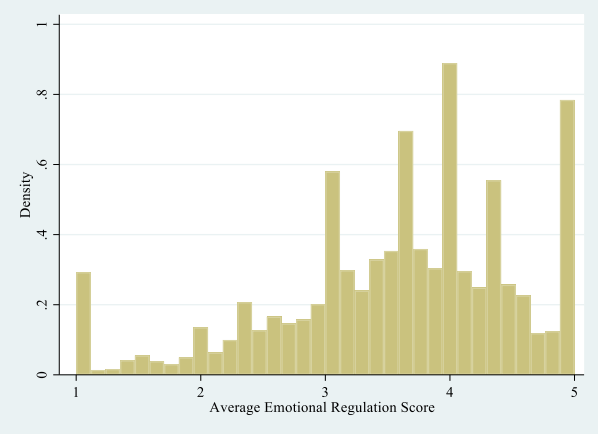


Figure 2I – Grade 6 Mental Health and Wellbeing Distribution

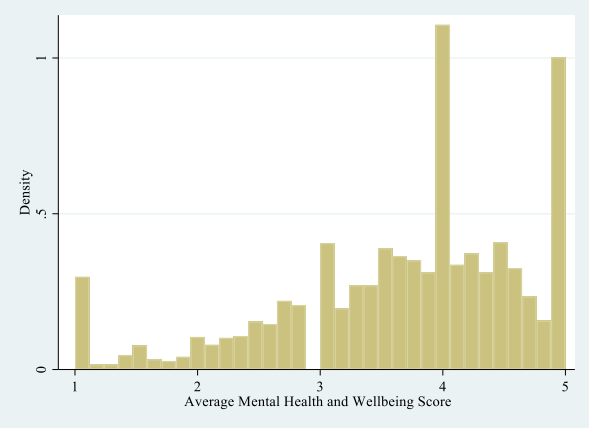
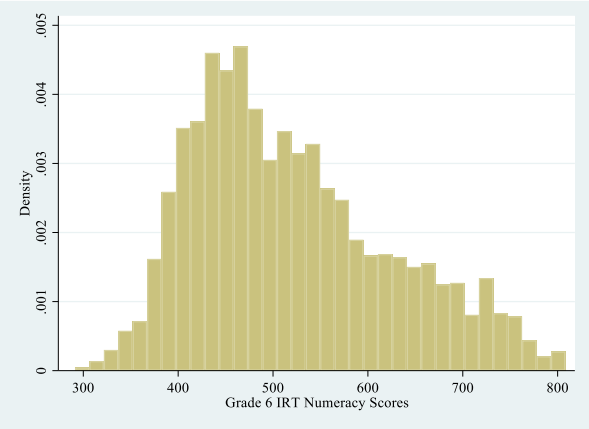


Figure 2J – Grade 6 IRT Numeracy Score Distribution



Appendix 3: Validating the Self-Reporting Scales

Exploratory Factor Analysis

Exploratory factor analysis (EFA) was undertaken with the datasets for the two cohorts to investigate the reliability of the SEL scores and the extent to which items on each scale contributed to the underlying trait. Tables 3A and 3B set out the results of the EFAs for pupils in Grades 3 and 6 respectively. The statistics for the older cohort show that all scales typically loaded onto one factor, and the eigenvalues for such factor were well above the suggested cut-off of 1, indicating that each scale primarily measured one main trait (Kaiser, 1960). The EFA results for the younger cohort also indicate similar findings for the scale concerning pupils' self-efficacy.

Table 3A: Grade 3 Pupils' EFA Results

Scale	No. Items	Cronbach's Alpha (α)	Factor 1		Factor 2	
			Eigen-value	Standard. Loadings	Eigen-value	Standard. Loadings
Social Skills	12	.91	6.08	.64-.79	1.00	.53-.78
Self-efficacy	11	.91	5.87	.58-.79	-	-
Emotional Regulation	10	.89	5.14	.45-.83	1.16	.70-.78
Mental Health and Wellbeing	8	.88	4.34	.60-.80	1.03	.54-.89

Table 3B: Grade 6 Pupils' EFA Results

Scale	No. Items	Cronbach's Alpha (α)	Eigen-value	Standard. Loadings
Social skills	12	.93	6.73	.64-.80
Self-efficacy	11	.89	5.14	.62-.72
Emotional regulation	10	.89	4.95	.67-.73
Mental health and wellbeing	8	.88	4.37	.69-.76

However, the EFA statistics for the remaining three scales, social skills, emotional regulation and mental health and wellbeing, show that the Grade 3 pupils' responses did not load onto one factor, although in each case the eigenvalue for the second factor only just exceeded 1 (1.00-1.16). Tables 3C, 3D and 3E show the full factor loadings for these scales. Reviewing the distribution of the Grade 3 data for these scales revealed no outliers that could account for the difference, whether resulting from pupils' language difficulties, misunderstanding the items or becoming fatigued by certain points in the scale (Liu & Zumbo, 2012). The pattern of ratings is nevertheless slightly different to that for the Grade 6 cohort, with fewer responses at the extremes (i.e. 'Strongly agree' or 'Strongly disagree') and therefore a higher proportion of learners choosing 'Agree' to rate their SEL.

Table 3C – Grade 3 Rotated Factor Loadings for Social Skills Scale

Social Skills – Scale Item	Factor 1 Loading	Factor 2 Loading	Unique Variance
1. I feel confident talking to others.	.2098	.6984	.4682
2. Other people like me.	.1865	.7807	.3557
3. I like to share things with others	.3577	.7271	.3433
4. I help others when they need help.	.4781	.5915	.4216
5. I make friends easily.	.3773	.5267	.5803
6. If I hurt someone, I say sorry.	.6356	.3959	.4393
7. I am polite towards others (e.g. greetings, saying thank you).	.7006	.3354	.3966
8. I feel good if I help someone.	.7588	.2462	.3635
9. Other children like to play with me.	.6859	.2773	.4527
10. I take care of other people's things as if they were my own.	.7925	.2341	.3171
11. I enjoy co-operating with others.	.7420	.2471	.3884
12. I respect the opinion of others.	.7458	.2263	.3926

Table 3D – Grade 3 Rotated Factor Loadings for Emotional Regulation Scale

Emotional Regulation – Scale Item	Factor 1 Loading	Factor 2 Loading	Unique Variance
1. I can control my temper.	.1974	.7875	.3409
2. When I am scared, I can calm myself down.	.3438	.6985	.3939
3. When I feel angry, I can calm myself down.	.2609	.7375	.3880
4. I can manage my anger (e.g. I don't hit other people or things).	.2510	7513	.3726
5. It is easy for me to stick to my aims.	.7446	.3212	.3424
6. It is easy for me to achieve my goals	.8024	.2346	.3012
7. If I am in trouble, I can think of a solution.	.7749	.2613	.3313
8. I can handle whatever comes my way.	.8271	.1989	.2763
9. If something I did not expect happens, I know how to react.	.7745	.2247	.3497
10. I can concentrate when doing schoolwork.	.4538	.4392	.6012

Table 3E – Grade 3 Rotated Factor Loadings for Mental Health and Wellbeing Scale

Mental Health and Wellbeing – Scale Item	Factor 1 Loading	Factor 2 Loading	Unique Variance
1. I have felt cheerful and in good spirits.	.7635	.1932	.3797
2. I have felt calm and relaxed.	.7986	.2047	.3203
3. I have felt active and vigorous.	.7877	.2131	.3341
4. I woke up feeling fresh and rested.	.6913	.3568	.3948
5. I do things that interest me.	.5966	.4783	.4154
6. If I am hungry there is enough to eat at home.	.5048	.5441	.4491
7. I am proud of my clothes.	.1888	.8925	.1678
8. I am proud of my shoes (or having shoes).	.1913	.8875	.1757

Confirmatory Factor Analysis

In parallel, we conducted confirmatory factor analysis (CFA) with the other randomised datasets for the two cohorts. The purpose of the CFA was to test the hypothesised model for each scale, and the extent to which the model explained the pupils' responses. Specifically, we used the asymptotically

distribution free (ADF) estimation method, which is preferable to maximum likelihood estimation where the sample size is large and the data violate the assumption of normal distribution (Huang & Bentler, 2015). As indicated above, the SEL data were negatively skewed with more learners conveying their agreement with the scale statements than disagreeing with them.

The models were constructed according to the anticipated relationships, namely each scale item being driven by the respective latent variable (Matsunaga, 2010). The specified models were then estimated, and yielded standardised coefficients ranging from .55 to .82, indicating that every item was contributing to the SEL scales. Tables 3F and 3G below show the loadings for basic models for each scale for Grades 3 and 6 respectively. They also set out various 'goodness of fit' statistics for the different models, which indicate whether each scale shows uni-dimensionality with all items contributing to the relevant underlying factor. These include the chi-squared test, comparative fit index (CFI), Tucker-Lewis index (TLI), the root mean square of error of approximation (RMSEA) and the standardised root mean square residual (SRMR). Multiple indicators of fit are preferable for CFAs because some are particularly sensitive to certain characteristics of the data, for example, the chi-squared test which often shows statistically significant results for large samples.

Table 3F: Grade 3 Pupils' CFA Loadings and Goodness of Fit Indices

Scale	Coefficient Loadings	Chi-squared	CFI	TLI	RMSEA	SRMR
	-	p>.05	>.95	>.90	<.06	<.08
Social skills	.55-.76	p<.05	.65	.57	.04	.09
Self-efficacy	.67-.77	p<.05	.78	.73	.05	.09
Emotional regulation	.60-.78	p<.05	.75	.68	.06	.10
Mental health and wellbeing	.70-.79	p<.05	.68	.55	.08	.10

Table 3G: Grade 6 Pupils' CFA Loadings and Goodness of Fit Indices

Scale	Coefficient Loadings	Chi-squared	CFI	TLI	RMSEA	SRMR
		p>.05	>.95	>.90	<.06	<.08
Social skills	.64-.82	p<.05	.92	.91	.02	.03
Self-efficacy	.62-.72	p<.05	.93	.91	.03	.03
Emotional regulation	.65-.73	p<.05	.92	.89	.03	.03
Mental health and wellbeing	.73-.78	p<.05	.86	.80	.06	.05

The tables reveal several key results of interest, especially relating to the models' goodness of fit, namely that:

- The models show a better fit for the Grade 6 data, than the Grade 3 responses. This accords with the findings of the EFA whereby more of the older cohorts' rating scores loaded onto single factors;
- The chi-squared test shows lower fit for all models but this may be because such a large sample size was used. Similarly, the CFI and TLI indicate reduced levels of fit, at least for the Grade 3 data; and
- By contrast the RMSEA and, in some instances, the SRMR show better model fit, the former being more robust with large sample sizes (Matsunaga, 2010).

Investigation of some of the models showing lower fit revealed evidence of covariance between the items' errors. For example, data for both cohorts showed a correlation between pupils' responses to the statements '*I am proud of my clothes*' and '*I am proud of my shoes (or having shoes)*'. Where there was a strong theoretical basis for doing so, we therefore adjusted the models to include these covariances, as shown in Table 3H. In each case, the revisions improved the overall fit of the scales. We also calculated factor scores for the models before and after the modifications and assessed their correlations

to ascertain whether the model had been substantially changed. In each case, the correlations exceed .97 and so it appears that the adjustments did not affect the overall structure.

Table 3H – Modified Models for Confirmatory Factor Analyses

Scale	Related Items	Basic Model					Revised Model					Factor Correlation
		Loadings	CFI >.95	TLI >.90	RMSEA <.06	SRMR <.08	Loadings	CFI >.95	TLI >.90	RMSEA <.06	SRMR <.08	
Grade 3												
Self-efficacy	'When I am confronted with a problem, I usually find solutions.' 'No matter what comes my way, I'm usually able to handle it.'	.67-.77	.78	.73	.05	.09	.67-.77	.80	.75	.05	.09	.99
Emotional Regulation	'It is easy for me to stick to my aims.' 'It is easy for me to achieve my goals.'	.60-.78	.75	.68	.06	.10	.60-.77	.78	.71	.06	.09	.99
Mental Health and Wellbeing	'I am proud of my clothes.' 'I am proud of my shoes (or having shoes).'	.70-.79	.68	.55	.08	.10	.59-.79	.88	.83	.05	.05	.97
Grade 6												
Self-efficacy	'I can solve difficult problems if I try.' 'I can solve most problems if I make enough effort.'	.62-.72	.93	.91	.03	.03	.62-.72	.95	.93	.02	.03	.99
Emotional Regulation	'I can control my temper.' 'When I feel angry, I can calm myself down.'	.65-.73	.92	.89	.03	.03	.63-.71	.93	.91	.03	.03	.99
Mental Health and Wellbeing	'I am proud of my clothes.' 'I am proud of my shoes (or having shoes).'	.73-.78	.86	.80	.06	.05	.64-.78	.96	.94	.03	.02	.99

Additional Analyses for Grade 3 Data

In addition to the inclusion of covariances, we examined several further models to explain the Grade 3 data loadings. First, we removed the items that loaded onto the second factor and re-ran the CFAs. Second, we structured the models to include two separate but correlated latent factors, using the loading classifications shown in Tables 3C, 3D and 3E. Finally, we explored the fit of bifactor models, which include a general factor which loads onto all observable variables as well as two unrelated factors that load onto the sub-groups (Dunn & McCray, 2020). Figure 3A highlights the differences between the models while Table 3I below compares the goodness-of-fit indices for each of the scale models.

Figure 3A– Hypothesised Models for Confirmatory Factor Analyses

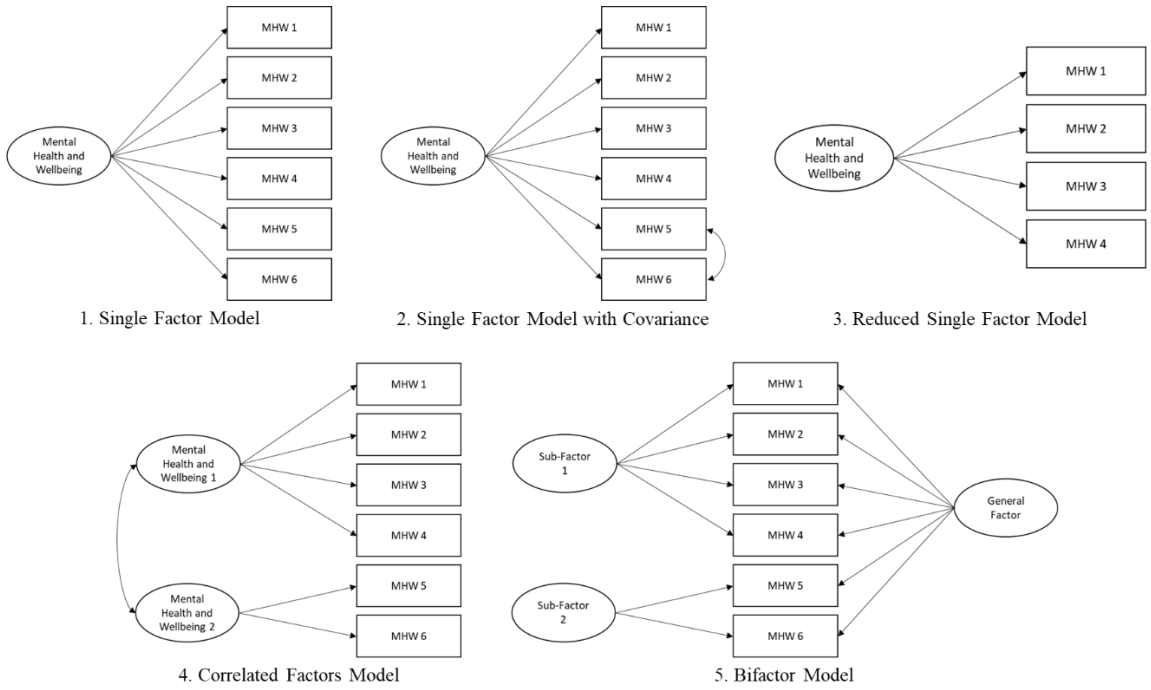


Table 3I – Goodness of Fit Indices for Different Models for Grade 3 Data

Scale	Coefficient Loadings	Chi-squared p>.05	CFI >.95	TLI >.90	RMSEA <.06	SRMR <.08
Social Skills						
Basic one-factor model	.55-.76	p < .05	.651	.574	.042	.093
Basic model with covariances	.55-.75	p < .05	.684	.607	.040	.087
Reduced one-factor model	.71-.78	p < .05	.872	.808	.043	.053
Correlated factors model	.58-.79	p < .05	.790	.732	.035	.063
Bifactor model	-.20-.77	p < .05	1.000	-	-	.023
Emotional Regulation						
Basic one-factor model	.60-.78	p < .05	.754	.684	.061	.096
Basic model with covariances	.60-.77	p < .05	.783	.712	.059	.092
Reduced one-factor model	.45-.82	p < .05	.942	.904	.050	.030
Correlated factors model	.58-.81	p < .05	.861	.816	.047	.067
Bifactor model	.00-.80	p < .05	1.000	-	-	.026
Mental Health and Wellbeing						
Basic one-factor model	.70-.79	p < .05	.678	.549	.075	.101
Basic model with covariances	.59-.79	p < .05	.881	.825	.046	.048
Reduced one-factor model	.72-.79	p < .05	.972	.945	.033	.021
Correlated factors model	.73-.82	p < .05	.804	.711	.060	.063
Bifactor model	.07-.77	p < .05	.987	.969	.020	.015

Notes: The bifactor models did not converge for either the emotional regulation or social skills data.

Based on both theoretical considerations and the above indices, we used the reduced one-factor models to analyse the Grade 3 learners' social skills, emotional regulation and mental health and wellbeing data. Such approach offered the advantages of simplicity and parsimony, and has been adopted in other studies in low-income countries (Forsberg et al, 2019; Raikes et al, 2019; Wolf et al, 2017).

Various considerations regarding the Grade 3 pupils' age and maturity, and the administration of the scale items might account for the complexity of the final analyses. These include:

- Fatigue, which could be evidenced by the sequential loading onto factors shown in Tables 3C, 3D and 3E, with pupils becoming tired or bored as they reached the later items in each scale. This potentially points at the value of using short scales for young children in future SEL research going forward;
- Cultural or language factors, whereby despite careful efforts to develop and translate the scales, some of the children could have found certain items difficult to grasp, for example the use of conditional or hypothetical statements starting with 'if' like '*If I am in trouble, I can think of a solution*'; and
- Distractions, with the younger learners being more susceptible to disruptions or disturbances if they were not interested in the task, or rushing to complete the survey without understanding the items or considering their responses to compete with their peers or so that they could return to lessons.

Appendix 4: Full Regression Tables

Table 4A– Grade 3 Multivariate Regressions for 2021 Numeracy

2021 Numeracy		Separate Models				Combined Model
		(1) Social Skills	(2) Self-Efficacy	(3) Emotional Regulation	(4) Mental Health and Wellbeing	(5) All SEL Variables
2019 Numeracy		0.24*** (0.03)	0.25*** (0.03)	0.24*** (0.03)	0.24*** (0.03)	0.24*** (0.03)
2021 Social Skills		0.17*** (0.03)	-	-	-	0.12** (0.04)
2021 Self-Efficacy		-	0.10* (0.04)	-	-	0.04 (0.05)
2021 Emotional Regulation		-	-	0.08* (0.03)	-	-0.02 (0.04)
2021 Mental Health and Wellbeing		-	-	-	0.14*** (0.03)	0.06* (0.03)
Gender (base group: female)		0.08* (0.03)	0.07* (0.03)	0.07* (0.03)	0.08* (0.03)	0.08* (0.03)
Age		0.03 (0.02)	0.03 (0.02)	0.03 (0.02)	0.03* (0.02)	0.03 (0.02)
Class Grade		0.31*** (0.06)	0.30*** (0.06)	0.30*** (0.06)	0.31*** (0.06)	0.31*** (0.06)
Caregiver Literacy (base group: no/low literacy)		0.03 (0.04)	0.02 (0.05)	0.02 (0.05)	0.02 (0.05)	0.03 (0.05)
Family Wealth (base group: first tercile (poorest))	Second Tercile	0.09 (0.06)	0.09 (0.06)	0.09 (0.06)	0.09 (0.06)	0.09 (0.06)
	Third Tercile	0.32*** (0.08)	0.33*** (0.08)	0.34*** (0.08)	0.33*** (0.08)	0.32*** (0.08)
Household Size		-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)	0.00 (0.01)
Location (base group: rural)		0.28** (0.09)	0.28** (0.09)	0.28** (0.09)	0.28** (0.09)	0.27** (0.09)
Region (base group: Addis Ababa)	Amhara	-0.11 (0.11)	-0.18 (0.11)	-0.18 (0.11)	-0.16 (0.11)	-0.12 (0.11)
	Benishangul Gumuz	-0.90*** (0.17)	-0.99*** (0.17)	-0.99*** (0.17)	-0.97*** (0.17)	-0.92*** (0.17)
	Oromia	-0.12 (0.11)	-0.22* (0.11)	-0.20 (0.11)	-0.17 (0.10)	-0.14 (0.11)
	SNNP	-0.16 (0.12)	-0.21 (0.13)	-0.21 (0.13)	-0.20 (0.13)	-0.17 (0.13)
	Somali	-0.05 (0.13)	-0.05 (0.13)	-0.07 (0.13)	-0.08 (0.14)	-0.06 (0.14)
Constant		-1.31*** (0.23)	-1.20*** (0.24)	-1.22*** (0.24)	-1.28*** (0.23)	-1.31*** (0.23)
Observations		2,086	2,086	2,086	2,086	2,086
R-squared		.39	.38	.38	.38	.39

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4B– Grade 6 Multivariate Regressions for 2021 Numeracy

2021 Numeracy		**Separate Models**				**Combined Model**
		(1) Social Skills	**(2) Self-Efficacy**	**(3) Emotional Regulation**	**(4) Mental Health and Wellbeing**	**(5) All SEL Variables**
2019 Numeracy		0.58*** (0.02)	0.59*** (0.02)	0.59*** (0.02)	0.59*** (0.02)	0.58*** (0.02)
2021 Social Skills		0.17*** (0.02)	-	-	-	0.15*** (0.04)
2021 Self-Efficacy		-	0.12*** (0.02)	-	-	0.01 (0.04)
2021 Emotional Regulation		-	-	0.11*** (0.03)	-	-0.05 (0.04)
2021 Mental Health and Wellbeing		-	-	-	0.15*** (0.02)	0.06 (0.03)
Gender (base group: female)		0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)	0.00 (0.03)
Age		0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)	0.02 (0.01)
Class Grade		0.15*** (0.03)	0.15*** (0.03)	0.15*** (0.03)	0.15*** (0.03)	0.15*** (0.03)
Caregiver Literacy (base group: no/low literacy)		0.01 (0.04)	0.01 (0.04)	0.02 (0.04)	0.01 (0.04)	0.01 (0.04)
Family Wealth (base group: first tercile (poorest))	Second Tercile	0.03 (0.04)	0.04 (0.04)	0.04 (0.04)	0.03 (0.04)	0.03 (0.04)
	Third Tercile	0.09 (0.05)	0.11* (0.05)	0.11* (0.05)	0.09 (0.05)	0.09 (0.05)
Household Size		0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Teacher Experience		0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)	0.01* (0.01)
Teacher Qualification (base group: no training)	Certificate	0.10 (0.15)	0.04 (0.14)	0.06 (0.14)	0.06 (0.15)	0.09 (0.15)
	Diploma	-0.06 (0.13)	-0.10 (0.12)	-0.08 (0.12)	-0.09 (0.14)	-0.06 (0.13)
	Degree	-0.06 (0.13)	-0.11 (0.13)	-0.09 (0.13)	-0.08 (0.14)	-0.06 (0.14)
Location (base group: rural)		0.06 (0.06)	0.06 (0.06)	0.06 (0.06)	0.06 (0.06)	0.06 (0.06)
Region (base group: Addis Ababa)	Amhara	-0.18 (0.11)	-0.19 (0.11)	-0.19 (0.11)	-0.18 (0.11)	-0.18 (0.11)
	Benishangul Gumuz	-0.63*** (0.13)	-0.66*** (0.13)	-0.68*** (0.13)	-0.65*** (0.13)	-0.62*** (0.13)
	Oromia	-0.27** (0.09)	-0.30** (0.09)	-0.30** (0.09)	-0.27** (0.09)	-0.26** (0.09)
	SNNP	-0.38** (0.12)	-0.41** (0.12)	-0.42*** (0.12)	-0.39** (0.11)	-0.37** (0.12)
	Somali	-0.57*** (0.11)	-0.61*** (0.11)	-0.61*** (0.11)	-0.58*** (0.11)	-0.56*** (0.11)
Constant		-0.86** (0.28)	-0.77** (0.29)	-0.81** (0.30)	-0.84** (0.29)	-0.87** (0.28)
Observations		2,157	2,157	2,157	2,157	2,157
R-squared		.58	.57	.57	.58	.58

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4C– Grade 3 Logistic Regressions for 2021 Low Performance

2021 Low Performance		**Separate Models**				**Combined Model**
		(1) Social Skills	**(2) Self-Efficacy**	**(3) Emotional Regulation**	**(4) Mental Health and Wellbeing**	**(5) All SEL Variables**
2019 Low Performance		0.78*** (0.16)	0.75*** (0.16)	0.69*** (0.16)	0.72*** (0.16)	0.77*** (0.17)
2021 Social Skills		-0.51*** (0.11)	-	-	-	-0.33* (0.13)
2021 Self-Efficacy		-	-0.49*** (0.10)	-	-	-0.33* (0.15)
2021 Emotional Regulation		-	-	-0.37*** (0.09)	-	0.05 (0.14)
2021 Mental Health and Wellbeing		-	-	-	-0.41*** (0.08)	-0.12 (0.09)
Gender (base group: female)		-0.33** (0.11)	-0.34** (0.11)	-0.34** (0.11)	-0.33** (0.12)	-0.34** (0.11)
Age		-0.10* (0.05)	-0.09* (0.05)	-0.11* (0.05)	-0.10* (0.05)	-0.10* (0.05)
Class Grade		-0.68*** (0.14)	-0.66*** (0.14)	-0.66*** (0.14)	-0.67*** (0.14)	-0.69*** (0.14)
Caregiver Literacy (base group: no/low literacy)		-0.06 (0.19)	-0.06 (0.19)	-0.04 (0.19)	-0.02 (0.19)	-0.05 (0.19)
Family Wealth (base group: first tercile (poorest))	Second Tercile	-0.34 (0.19)	-0.32 (0.20)	-0.35 (0.20)	-0.35 (0.19)	-0.32 (0.19)
	Third Tercile	-0.58* (0.30)	-0.59* (0.29)	-0.67* (0.29)	-0.63* (0.29)	-0.57 (0.30)
Household Size		0.01 (0.04)	0.00 (0.03)	0.01 (0.04)	0.00 (0.04)	0.00 (0.03)
Location (base group: rural)		-0.62 (0.36)	-0.61 (0.35)	-0.59 (0.35)	-0.59 (0.35)	-0.60 (0.36)
Region (base group: Addis Ababa)	Amhara	0.82 (0.49)	1.02* (0.51)	1.03* (0.50)	0.95 (0.50)	0.88 (0.51)
	Benishangul Gumuz	2.70*** (0.63)	3.00*** (0.63)	3.00*** (0.62)	2.90*** (0.63)	2.82*** (0.64)
	Oromia	0.88 (0.50)	1.26* (0.50)	1.18* (0.50)	1.06* (0.50)	1.03* (0.52)
	SNNP	0.96 (0.53)	1.10* (0.55)	1.10* (0.55)	1.09* (0.55)	0.98 (0.53)
	Somali	1.00 (0.59)	0.97 (0.61)	1.08 (0.60)	1.08 (0.61)	0.96 (0.61)
Constant		1.14 (0.77)	0.85 (0.77)	0.92 (0.76)	0.99 (0.76)	1.08 (0.78)
Observations		1,846	1,846	1,846	1,846	1,846
Pseudo R-squared		.22	.22	.22	.22	.23

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4D– Grade 6 Logistic Regressions for 2021 Low Performance

2021 Low Performance		Separate Models				Combined Model
		(1) Social Skills	(2) Self-Efficacy	(3) Emotional Regulation	(4) Mental Health and Wellbeing	(5) All SEL Variables
2019 Low Performance		1.04*** (0.16)	1.10*** (0.17)	1.10*** (0.17)	1.07*** (0.16)	1.04*** (0.16)
2021 Social Skills		-0.51*** (0.09)	-	-	-	-0.42* (0.18)
2021 Self-Efficacy		-	-0.38*** (0.09)	-	-	0.01 (0.16)
2021 Emotional Regulation		-	-	-0.40*** (0.09)	-	0.08 (0.15)
2021 Mental Health and Wellbeing		-	-	-	-0.48*** (0.08)	-0.20 (0.16)
Gender (base group: female)		-0.07 (0.12)	-0.06 (0.12)	-0.07 (0.12)	-0.08 (0.13)	-0.08 (0.12)
Age		-0.09* (0.04)	-0.09* (0.04)	-0.10* (0.04)	-0.10** (0.04)	-0.09* (0.04)
Class Grade		-0.33* (0.15)	-0.31* (0.15)	-0.32* (0.15)	-0.32* (0.15)	-0.33* (0.15)
Caregiver Literacy (base group: no/low literacy)		-0.12 (0.16)	-0.12 (0.16)	-0.15 (0.16)	-0.11 (0.16)	-0.11 (0.16)
Family Wealth (base group: first tercile (poorest))	Second Tercile	-0.06 (0.17)	-0.08 (0.17)	-0.08 (0.17)	-0.06 (0.17)	-0.06 (0.17)
	Third Tercile	-0.06 (0.20)	-0.14 (0.20)	-0.11 (0.20)	-0.06 (0.20)	-0.05 (0.20)
Household Size		0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)	0.04 (0.03)
Teacher Experience		-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)	-0.01 (0.02)
Teacher Qualification (base group: no training)	Certificate	-0.78 (0.44)	-0.59 (0.45)	-0.64 (0.44)	-0.70 (0.46)	-0.78 (0.45)
	Diploma	-0.42 (0.38)	-0.27 (0.37)	-0.30 (0.36)	-0.37 (0.40)	-0.43 (0.39)
	Degree	-0.87 (0.45)	-0.69 (0.44)	-0.74 (0.43)	-0.86 (0.47)	-0.90* (0.46)
Location (base group: rural)		-0.43 (0.25)	-0.40 (0.25)	-0.41 (0.25)	-0.43 (0.25)	-0.44 (0.25)
Region (base group: Addis Ababa)	Amhara	0.99* (0.39)	0.98* (0.39)	1.01** (0.39)	0.97* (0.39)	0.97* (0.39)
	Benishangul Gumuz	2.62*** (0.43)	2.71*** (0.43)	2.76*** (0.43)	2.69*** (0.43)	2.61*** (0.43)
	Oromia	1.30*** (0.36)	1.39*** (0.37)	1.40*** (0.36)	1.32*** (0.37)	1.28*** (0.36)
	SNNP	1.49*** (0.41)	1.59*** (0.40)	1.65*** (0.40)	1.55*** (0.40)	1.47*** (0.40)
	Somali	2.27*** (0.40)	2.39*** (0.41)	2.41*** (0.40)	2.30*** (0.41)	2.24*** (0.40)
Constant		0.69 (1.01)	0.41 (1.03)	0.51 (1.02)	0.65 (1.03)	0.73 (1.03)
Observations		2,154	2,154	2,154	2,154	2,154
Pseudo R-squared		.21	.19	.20	.20	.21

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4E– Grade 3 Multivariate Regressions for 2021 Mental Health and Wellbeing

2021 Mental Health and Wellbeing		**Separate Models**			**Combined Model**
		(1) Social Skills	**(2) Self-Efficacy**	**(3) Emotional Regulation**	**(5) All SEL Variables**
2021 Social Skills		0.59*** (0.03)	-	-	0.44*** (0.03)
2021 Self-Efficacy		-	0.51*** (0.05)	-	0.16*** (0.03)
2021 Emotional Regulation		-	-	0.50*** (0.04)	0.24*** (0.03)
Gender (base group: female)		-0.02 (0.03)	-0.02 (0.03)	-0.01 (0.03)	-0.02 (0.02)
Age		0.00 (0.01)	-0.01 (0.01)	-0.01 (0.01)	-0.01 (0.01)
Class Grade		0.00 (0.03)	-0.01 (0.04)	0.00 (0.04)	0.00 (0.03)
Caregiver Literacy (base group: no/low literacy)		0.03 (0.03)	0.03 (0.03)	0.01 (0.03)	0.03 (0.03)
Family Wealth (base group: first tercile (poorest))	Second Tercile	0.00 (0.04)	-0.02 (0.05)	0.01 (0.05)	-0.01 (0.04)
	Third Tercile	-0.01 (0.05)	-0.01 (0.05)	0.08 (0.05)	0.01 (0.04)
Household Size		0.00 (0.01)	0.00 (0.01)	0.00 (0.01)	0.00 (0.01)
Location (base group: rural)		0.06 (0.05)	0.08 (0.06)	0.05 (0.05)	0.03 (0.04)
Region (base group: Addis Ababa)	Amhara	0.08 (0.08)	-0.15 (0.09)	-0.16 (0.10)	0.02 (0.06)
	Benishangul Gumuz	0.11 (0.09)	-0.23* (0.10)	-0.24* (0.10)	0.01 (0.07)
	Oromia	0.11 (0.08)	-0.24** (0.09)	-0.19* (0.09)	-0.01 (0.05)
	SNNP	0.02 (0.12)	-0.12 (0.14)	-0.12 (0.14)	0.00 (0.08)
	Somali	0.14 (0.09)	0.12 (0.09)	0.03 (0.09)	0.12* (0.06)
Constant		-0.12 (0.16)	0.22 (0.16)	0.15 (0.16)	0.04 (0.12)
Observations		2,855	2,855	2,855	2,855
R-squared		.34	.24	.26	.43

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

Table 4F– Grade 6 Multivariate Regressions for 2021 Mental Health and Wellbeing

2021 Mental Health and Wellbeing		**Separate Models**			**Combined Model**
		(1) Social Skills	**(2) Self-Efficacy**	**(3) Emotional Regulation**	**(5) All SEL Variables**
2021 Social Skills		0.77*** (0.02)	-	-	0.55*** (0.03)
2021 Self-Efficacy		-	0.68*** (0.03)	-	0.13*** (0.03)
2021 Emotional Regulation		-	-	0.71*** (0.02)	0.19*** (0.03)
Gender (base group: female)		-0.01 (0.02)	-0.02 (0.02)	-0.01 (0.02)	-0.01 (0.02)
Age		-0.01 (0.01)	-0.01 (0.01)	0.00 (0.01)	-0.01 (0.01)
Class Grade		0.01 (0.03)	-0.01 (0.04)	-0.01 (0.03)	0.00 (0.03)
Caregiver Literacy (base group: no/low literacy)		0.04 (0.02)	0.04 (0.03)	0.08** (0.02)	0.05* (0.02)
Family Wealth (base group: first tercile (poorest))	Second Tercile	0.02 (0.03)	0.03 (0.03)	0.03 (0.03)	0.01 (0.03)
	Third Tercile	0.11** (0.04)	0.15** (0.05)	0.13** (0.04)	0.10** (0.03)
Household Size		0.00 (0.00)	0.00 (0.01)	0.00 (0.01)	0.00 (0.00)
Teacher Experience		0.00 (0.00)	0.00 (0.00)	0.00 (0.00)	0.00 (0.00)
Teacher Qualification (base group: no training)	Certificate	0.14 (0.12)	-0.08 (0.11)	0.02 (0.11)	0.08 (0.11)
	Diploma	0.11 (0.11)	-0.07 (0.09)	0.02 (0.10)	0.06 (0.10)
	Degree	0.07 (0.12)	-0.13 (0.11)	-0.03 (0.13)	0.01 (0.12)
Location (base group: rural)		-0.08* (0.03)	-0.05 (0.05)	-0.06 (0.05)	-0.06 (0.03)
Region (base group: Addis Ababa)	Amhara	-0.03 (0.04)	-0.05 (0.05)	-0.08* (0.04)	-0.03 (0.03)
	Benishangul Gumuz	-0.05 (0.05)	-0.25** (0.08)	-0.30*** (0.06)	-0.05 (0.05)
	Oromia	-0.04 (0.04)	-0.21*** (0.05)	-0.20*** (0.04)	-0.08* (0.04)
	SNNP	-0.02 (0.07)	-0.18* (0.08)	-0.26** (0.07)	-0.05 (0.06)
	Somali	-0.04 (0.06)	-0.21* (0.09)	-0.22* (0.08)	-0.07 (0.05)
Constant		-0.06 (0.20)	0.38 (0.24)	0.20 (0.22)	0.07 (0.19)
Observations		2,611	2,611	2,611	2,611
R-squared		.62	.44	.51	.65

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$.

[i] This section is not intended as an exhaustive review of the literature regarding SEL in low- and lower-middle-income countries. Rather it builds on our earlier papers to clarify the key concepts used throughout the study and to highlight the most relevant research in similar contexts.

[ii] The use of different terms for SEL used in different papers can make it challenging to identify all work in this area. This literature review primarily focuses on papers including the terms 'socio-emotional', 'psychosocial' and 'non-cognitive'. Although relevant, literature regarding 21st century learning also covers topics like digital literacy, which are beyond the scope of this paper.

[iii] Young Lives also works in India, Peru and Vietnam. For further details, see www.younglives.co.uk.

[iv] During the initial piloting, additional instruments were used to assess pupils' resilience, motivation and peer support for learning. However, these were excluded from the final data collection based on psychometric analyses of their functioning, for example, whether they offered sufficient variation or uni-dimensionality.

[v] Students were also assessed on their literacy and reading abilities. However, linguistic and cultural differences between the various languages spoken in Ethiopia and used in the assessments means that they are not directly comparable (Yorke & Ogando, 2018). As such, we have focused on numeracy in these analyses.

[vi] These databases included: https://www.rand.org/education-and-labor/projects/assessments.html, https://measuringsel.casel.org/assessmentguide/?accessform=true&position=Researcher%2Fprofessor, https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/measuring-essentialskills/spectrum-database/ and https://inee.org/measurement-library.

[vii] We also piloted scales for anxiety and student effort, however, these were omitted from the final data collection. The anxiety scale did not achieve uni-dimensionality and was not normally distributed, while the student effort scale showed little variation.

[viii] Further questions examined pupils' activities and experiences during the COVID-19 school closures, however these data are not reported in this paper.

[ix] Ethiopian children are supposed to enter Grade 1 at age 7 so should be aged 9-10 during Grade 3 and 12-13 in Grade 6.

[x] The attrition rates were calculated using the full datasets, not just pupils' SEL responses, not least because the earlier SEL data were only captured for the older students. These figures and the subsequent analyses also exclude 2019 participants in Tigray who, as noted, were omitted from the 2021 surveys due to ongoing conflict, violence and insecurity.

[xi] The SEL scales contained positively worded items only to ensure that they captured the same outcomes, to minimise cognitive demands on respondents and to avoid linguistic ambiguity when translating the statements into multiple languages (Van Sonderen, Sanderman & Coyne, 2013; Weems, Onwuegbuzie & Collins, 2006; Yorke & Ogando, 2018). See Bayley et al. (2021) for a summary on the merits and limitations of using positive and negative items.

[xii] Initial versions of the models also included a variable for whether or not a child had attended any pre-primary education. Although some of the results were significant, we found that this variable was related to other child and household factors, such as family wealth, and so it was omitted from the final regressions to identify maximum variance among the other related variables.

[xiii] Many of the variables for learners' region are statistically significant, for example showing lower social skills and numeracy in areas like Benishangul Gumuz. However, variations in sampling between the regions, for reasons including conflict and insecurity, undermine the direct comparability of these results and so they are not reported.

[xiv] These comprised Canada, China, Colombia, Finland, Portugal, Russia, South Korea, Turkey and the USA.

[xv] The value of SEL and related life skills for girls does not detract from the particular structural challenges they face which must be addressed to address and improve girls' learning outcomes.

[xvi] The use of 'self-contained classrooms' is both controversial and mixed in Ethiopia. Arguments in favour highlight the integration of children's learning across subjects and outcomes, while arguments against emphasise teachers' lack of competence to instruct all subjects, the challenge of large classes and the constraints of inadequately resourced classrooms. For further detail of the debate, including its use and history in Ethiopia, see Cherinet (2001), Debebe (2002) and Melese and Tadege (2019).