RISE Working Paper 21/068
April 2021

Researching Socio-Emotional Learning, Mental Health and Wellbeing: Methodological Issues in Low-Income Contexts

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

Abstract

This paper explores methodological issues relating to research on children's socio-emotional learning (SEL), mental health and wellbeing in low- and lower-middle-income countries. In particular, it examines the key considerations and challenges that researchers may face and provides practical guidance for generating reliable and valid data on SEL, mental health and wellbeing in diverse settings and different cultural contexts. In so doing, the paper draws on the experience of recent research undertaken in Ethiopia to illustrate some of the issues and how they were addressed. The present study extends earlier 2018-2019 RISE Ethiopia research, expanding its scope to consider further aspects of SEL, mental health and wellbeing in the particular context of COVID-19. In particular, the research highlights that the pandemic has brought to the fore the importance of assessing learning, and learning loss, beyond academic learning alone.

Researching Socio-Emotional Learning, Mental Health and Wellbeing: Methodological Issues in Low-Income Contexts

Stephen Bayley
University of Cambridge

Darge Wole
Addis Ababa University

Louise Yorke
REAL Centre, University of Cambridge

Paul Ramchandani
University of Cambridge

Pauline Rose
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 into the design of this work, including Professor Tassew Woldehanna, Dr Belay Hagos, Professor Amare Asegdom, Professor Tirussew Tefera, and Professor Girma Lemma. Particular thanks to Chanie Ejigu for co-ordinating the research. We are also grateful to Professor Sileshi Zeleke and Professor Belay Kibret for their input. We are very grateful to Amy Jo Dowd, Eve Hadshar, and Celia Hsiao from the Lego Foundation for their helpful input into this paper, as well as members of the RISE Directorate for their feedback.

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., Yorke, L., Ramchandani, P. and 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

Contents

1. Introduction

This paper explores methodological issues relating to research on children's socio-emotional learning (SEL), mental health and wellbeing in low- and lower-middle-income countries. In particular, it examines the key considerations and challenges that researchers may face and provides practical guidance for generating reliable and valid data on SEL, mental health and wellbeing in diverse settings and different cultural contexts.

In so doing, the paper draws on the experience of recent research undertaken in Ethiopia to illustrate some of the issues and how they were addressed. This study was conducted as part of the Research on Improving Systems of Education (RISE) programme and funded with further support from the LEGO Foundation. This research in Ethiopia builds on previous data collection by the RISE programme, which examined aspects of students' SEL in 2018-2019 as part of the wider RISE Ethiopia research. It included approximately 4,000 pupils in Grade 4, across 168 government primary schools, using scales relating to: student effort; social skills; family support for learning; and teacher-student relationships. Further details regarding the initial sampling and findings are set out in Hoddinott et al. (2019) and Yorke, Wole & Rose (2020) respectively.

The present study extends the 2018-2019 research, expanding its scope to consider further aspects of SEL, mental health and wellbeing in the particular context of COVID-19. Specifically, it seeks to understand the experiences of children in Grades 3 and 6 of primary school (aged between 9 to 13 years) during the COVID-19 pandemic, including how their SEL, mental health and wellbeing have been affected, and to inform future policy and practice in Ethiopia. While a number of studies are assessing the effects of COVID-19 on learning with respect to literacy and numeracy specifically, this study highlights the importance of widening the conceptualisation of learning to include SEL, mental health and wellbeing. Yorke et al. (2021) provide a detailed rationale for the increased scope of the current study and the relevance of mental health and wellbeing. In deciding the appropriate approach for researching SEL, mental health and wellbeing, we draw on a variety of methodological and practical factors in a multi-stage process. A strength of the RISE Ethiopia research study is that our team is a collaboration between the Institute of Education Research, Addis Ababa University, the Policy Studies Institute, Addis Ababa and the REAL Centre, University of Cambridge. The RISE Ethiopia team members have a broad range of expertise in education, developmental psychology and psychometric analysis and therefore are well positioned to ensure that the research is grounded in the Ethiopian context.

The structure of the paper is as follows: section 2 provides an overview of SEL, mental health and wellbeing, and the interactions between them; section 3 discusses the limited research and knowledge gap regarding students' SEL in low-income countries; section 4 addresses the key issues and challenges pertaining to the measurement of SEL, mental health and wellbeing in different cultural settings; section

5 considers different research methods; section 6 highlights key factors in selecting appropriate tools; section 7 discusses the process of adapting and piloting instruments; and section 8 concludes the paper.

2. What are SEL, Mental Health and Wellbeing?

SEL is generally understood as the process of acquiring a wide range of personal and inter-personal competencies and attributes, which are believed to be important in and of themselves, but which also play a key role in the acquisition and development of other skills. For example, many formulations of SEL refer to individuals improving their abilities to manage their emotions, achieve their goals and work with other people (Jones & Doolittle, 2017; Zins & Elias, 2007). Numerous frameworks have been established for understanding and discussing SEL, with each identifying specific aspects of SEL which they suggest are important for students' learning and development. Examples include the 'Big Five' domains[1], and the Collaborative for Academic, Social and Emotional Learning (CASEL)[2] (CASEL, 2020; Organisation for Economic Co-operation and Development (OECD), n.d.).

Beyond learning, SEL shows important links with mental health and personal wellbeing. Mental health comprises "a state of well-being in which an individual realizes his or her own abilities, can cope with the normal stresses of life, can work productively and is able to make a contribution to his or her community" (WHO, 2018). It can be affected by a combination of social, biological and psychological factors, and there is a bidirectional relationship between mental health and SEL. SEL provides coping skills for improved wellbeing and emotional resilience, while poor mental health affects individuals' capacity for learning, as shown in Figure 1 (Diamond, 2014; Immordino-Yang & Damasio, 2007; UNESCO, 2020).

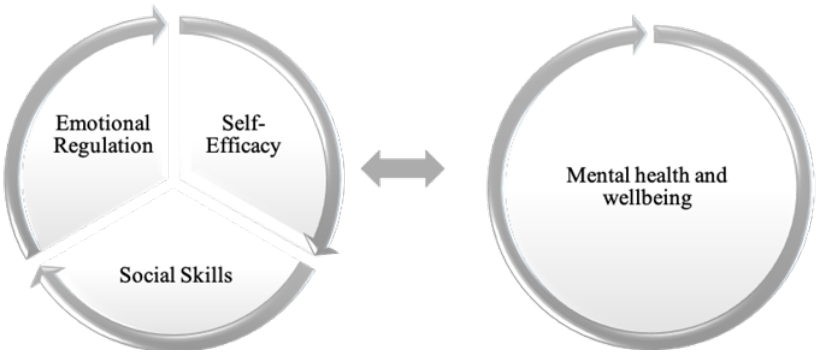

Figure 1 – Hypothesised Interactions between SEL, Mental Health and Wellbeing

[1] The 'Big Five' domains comprise task performance (conscientiousness), emotional regulation (emotional stability), collaboration (agreeableness), engaging with others (extraversion) and open-mindedness (openness to experience) (OECD, n.d.).
[2] CASEL describes SEL as comprising self-management, self-awareness, social awareness, relationship skills and responsible decision-making (CASEL, 2020).

Literature on the value of SEL, good mental health and positive wellbeing describes wide benefits for traditional academic attainment, employment and later life outcomes (Raikes et al., 2017). They may also be particularly important and relevant for learners in light of the COVID-19 pandemic, as discussed in Yorke, Wole and Rose (2020), and Yorke et al. (2021), or in conflict or emergency settings of heightened instability, insecurity and adversity. However, measuring children's SEL, mental health and wellbeing poses numerous methodological challenges for researchers generally, and especially for those working in difficult contexts or low-income settings, as the following sections highlight.

3. Existing Studies on SEL, Mental Health and Wellbeing in Low-income Countries

For many years, evidence and data on children's SEL, mental health and wellbeing in low- and lower-middle-income countries has been scant (Inter-Agency Network for Education in Emergencies, 2016). Following the global expansion of schooling systems to achieve universal primary education in the 1990s, learning indicators mainly focused on pupils' access and attendance, and then basic literacy and numeracy, but without any measure of their more holistic development or wellbeing. The limited data that are available estimate that up to 20% of children and adolescents in sub-Saharan Africa experience mental health problems, but the true extent and prevalence of these issues remain unknown. It is also likely to have increased during the COVID-19 pandemic (Atilola, 2017; Belfer, 2008; Cortina et al., 2012).

There are, however, various exceptions, and some examples of such studies in low-income countries are shown in Table 1. Indeed, the recent rise in international assessments and research to assess children's SEL, mental health and wellbeing has highlighted the value of expanding the definition of 'learning' to encompass broader education and developmental outcomes. In some cases, these comprise tools that have been used to ascertain psychosocial health among children affected by migration, conflict or trauma in contexts such as Syria and Palestine (Forsberg et al., 2019; D'Sa, 2019). Studies using such tools offer valuable insight regarding the impact of violence and instability, although the relevance of their findings to wider low-income settings remains questionable. Other research has focused on school readiness, including SEL among young children aged 3-6, notably using instruments such as the International Development and Early Learning Assessment (IDELA) and the Measuring Early Learning Quality and Outcomes (MELQO) (Pisani, Borisova & Dowd, 2018; Raikes et al, 2019; Wolf et al., 2017; Wolf & McCoy, 2019).

Table 1 – Studies of SEL, Mental Health and Wellbeing in Low-Income Countries

Focus	Instrument(s) Used	Age Range	Reference(s)	Country(ies)
SEL	International Development and Early Learning Assessment (IDELA)	4-8	Wolf, Halpin, Yoshikawa, Dowd, Pisani & Borisova (2017)	Ethiopia
	International Social and Emotional Learning Assessment (ISELA)	6-12	D'Sa & Krupar (2019)	Iraq
	Measuring Early Learning Quality and Outcomes (MELQO)	4-16	Raikes, Koziol, Janus, Platas, Weatherholt…& Sayre (2019)	Tanzania
	Social Emotional Response and Information Scenarios (SERAIS)	5-16	Kim & Tubbs Dolan (2019)	Lebanon/ Syria
	Student Learning in Emergencies Checklist (SLEC-26)	12-16	Forsberg, Schultz, Lodi & Tubbs Dolan (2019)	Palestine
Wellbeing	Cantril's Ladder	10-12	Camfield & Tafere (2009), Camfield (2012)	Ethiopia
	Children's Hope Scale	7-17	Haroz, Jordans, de Jong, Gross, Bass & Tol (2017)	Burundi, Indonesia, Nepal
Mental Health	Social Competence Scale (SCS) and others	5-9	Nakigudde, Bauta, Wolf & Huang (2016)	Uganda
	Spence Children's Anxiety Scale	13-17	Qadir, Maqsood, us-Sahar, Bukhtawer, Khalid…& Essau (2018)	Pakistan
	Strengths and Difficulties Questionnaire (SDQ)	5	Abera, Tesfaye, Hanlon, Admasu, Girma…& Andersen (2018)	Ethiopia

Perhaps the most valuable data regarding learners' SEL, mental health and wellbeing, over a larger age range of children and young people in low- and lower-middle-income contexts, is derived from the Young Lives[i] initiative. Established in 2002 as an international study of childhood poverty, Young Lives has collected longitudinal data on learners' lived experiences, including their SEL, mental health and wellbeing in Ethiopia, India, Peru and Vietnam. The Resilience Research Centre[ii] has also examined certain aspects of SEL, mental health and wellbeing among adolescents aged 12 to 19 in a range of high- to low-income contexts including in sub-Saharan Africa.

Overall, however, research on SEL, mental health and wellbeing among school-aged children in low- and lower-middle-income countries is patchy. While more research is available in high-income countries, the methods and tools used are not necessarily transferable or suitable for use in diverse cultural settings.

4. Key Issues in Measuring SEL, Mental Health and Wellbeing

Complexities around the measurement of SEL, mental health and wellbeing might explain some of the reasons for the research gap, especially in low- and lower-middle-income countries. To date, most studies have been conducted in high-income contexts, using instruments created and designed with those specific populations in mind. Researchers in low-income settings, however, cannot simply transfer measures from one context to another, and must adapt them to accommodate differences between the target groups. Figure 2 maps out some of the key stages involved in this process, building on work by Yorke, Wole and Rose (2020). In particular, it highlights the need to consider the specific cultural context within which students are located, which may affect the development of children's SEL, mental health and wellbeing, and the importance of engaging with local researchers.

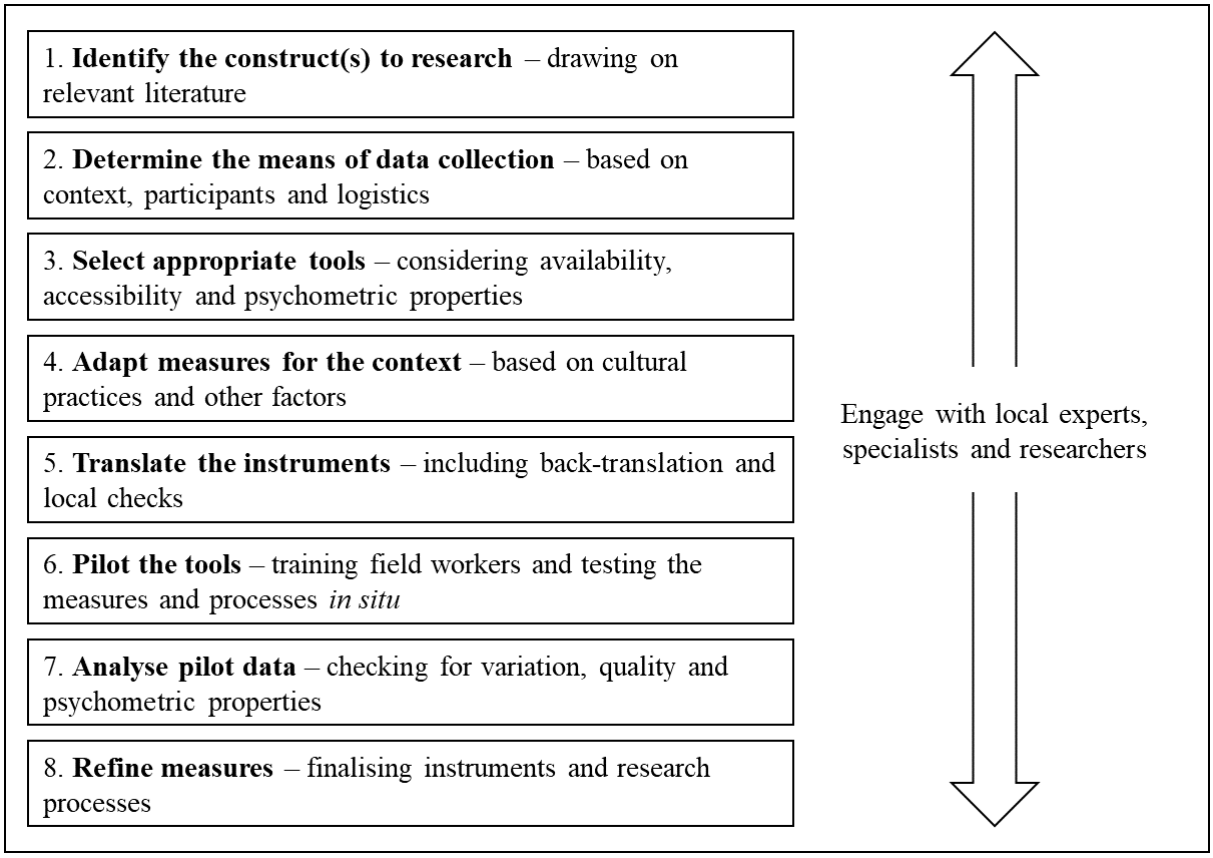

Figure 2 – Process for Conducting Research on SEL, Mental Health and Wellbeing (Adapted from Yorke, Wole and Rose (2020), Table 1, pp. 1-2)

First, across all contexts, SEL, mental health and wellbeing cover a broad range of overlapping skills, competencies and characteristics. The multiplicity of terms and frameworks has, to date, complicated the scope for the creation of appropriate scales and the generation of robust evidence that might promote meaningful discussion between stakeholders. Specifically, understanding and nurturing learners' SEL has been confused by the so-called 'jingle-jangle fallacy', where a single term is used to describe

numerous different things, or different terms are used to describe the same thing (Coleman & Cureton, 1954; Reeves & Venator, 2014).

SEL, mental health and wellbeing also comprise a mixture of both internal and external skills and processes, with implications of what needs to be included in measurement. Learners' outward social competencies may be easily observed, but many inward elements, for example how individuals understand emotions in themselves and others, are much harder to investigate. Researchers must therefore identify a set of responses, activities or visible behaviours which they believe capture and adequately represent the SEL, mental health and wellbeing construct(s) of interest (Miyake et al., 2000; Yorke, Wole & Rose, 2020).

Second, with respect to research in low- and lower-middle-income countries specifically, it is not possible or appropriate to simply transfer measures from high-income countries. For example, culture plays an important role in both the design and the use of research on learners' SEL, mental health and wellbeing. Across countries and settings, different cultural knowledge, norms, values, beliefs and preferences shape how key aspects of SEL and wellbeing are perceived and enacted (Doebel, 2020). These, in turn, may influence how social skills and coping strategies are conceived and demonstrated, and may include tendencies towards more individual or collective ways of thinking. These trends can affect how children learn to make decisions, solve problems and understand how to relate to others, which are all key aspects of SEL. More specifically, they shape how learners reconcile competing tensions, whether individual or community interests take precedence, and the extent to which social hierarchies, traditions or expectations govern final choices and outcomes (Roe, 1988).

Similarly, social norms around acceptable behaviours can affect attitudes and discipline practices, both at school and in the home (Talwar, Carlson & Lee, 2011). For example, studies of parents' interactions with their children in American and East Asian households have highlighted differences regarding expectations of children's compliance and engagement with older family members. While the former often invite children to participate in family conversations over the dinner table, the latter appear to encourage children to eat quietly and not talk while adults are speaking (Hsieh, 2004; Martini, 1996; Wanless et al., 2011).

Notions of childhood may also vary across culture and context (Burman, 2008). For example, in many low-income settings, children take on important household responsibilities, including cleaning chores, caring for younger siblings or contributing to family income, more frequently and at an earlier age than their counterparts in high-income countries. These chores often have a gender dimension and thus a disproportionate impact on girls, who spend between 30 and 50 per cent more time helping around the home than boys of the same age (UNICEF, 2016). Such realities arguably reflect different conceptualisations of what it means to be a child in diverse contexts, and the need to prioritise collective family requirements over individual child preferences, for example.

As such, efforts to understand and measure SEL, mental health and wellbeing must be grounded in the social, cultural, ecological and familial context of the specific developmental setting (Bronfenbrenner, 1979). This also highlights the need for caution in transferring global 'knowledge' between different contexts, as SEL, mental health and wellbeing may look very different and hold different associations in high-income countries compared with their lower-income counterparts.

Beyond the relevance of culture and context, studying SEL, mental health and wellbeing raises important ethical issues, particularly when conducting research with children in the Global South, which must be considered carefully. Asking questions about learners' internal experiences, their feelings, emotions and self-perceptions, admits researchers into a private space and must be treated with appropriate sensitivity. Similarly, investigations into children's family interactions or household relationships may create discomfort, give rise to safeguarding or disclosure obligations, or position the child as a gatekeeper for their private home environment (Homan, 2001). Researchers must therefore pay sufficient attention to ethical matters, not least by obtaining voluntary informed consent, protecting anonymity and confidentiality and, most importantly, avoiding any harm to the participants.

5. Review of Available Methods for Data Collection, and Approach for Ethiopian Study

Methodological challenges and cross-cultural considerations emphasise the importance of using appropriate research methods and of establishing measures relevant to the particular context. These comprise the focus of this and the next section, respectively.

Table 2 sets out the main methods that have been used to assess SEL, mental health and wellbeing, which include report scales, interviews, observations and direct assessments. The table further provides details of their suitability for different types of respondent, their key benefits and limitations, and examples of related tools. The use of particular approaches also depends on the construct(s) being measured, the age and maturity of the population of interest, the rationale for conducting the research and the intended analytical methods, as different approaches give rise to different types of data.

Table 2 – Methods for Measuring SEL, Mental Health and Wellbeing

Type	Description	Target population	Advantages	Limitations
Self-reporting scales e.g. Student Learning in Emergencies Checklist (SLEC-26), Children's Hope Scale	Participants rate their own abilities or wellbeing on an ordinal scale	• Older children and respondents	• Useful for collecting quantitative data on awareness, attitudes, voice and feelings which cannot be easily observed • Cost-efficient and easily constructed, scored and administered at scale • Unobtrusive and place minimal requirements or burden on respondents	• Require self-awareness of respondents and can be susceptible to biases including social desirability, faulty memory and/or references biases (comparisons with peers) • May require reading and so be unsuitable for young and/or illiterate participants

Teacher/ caregiver report scales e.g. Social Competence Scale (SCS), Strengths and Difficulties Questionnaire (SDQ)	Used by an external person, often a parent, teacher or clinician, to evaluate learners' behaviours, skills and strengths	• Respondents provide information on other people, which may include young children	• Useful for collecting quantitative data quickly and easily • Does not rely on participants for completion, so can be used with younger learners • Teacher ratings in particular often draw on strong reference group knowledge through which to compare learners	• Subject to reporting bias including reference, implicit and unconscious bias • Difficult to capture awareness and beliefs • Report limited to one setting and potential for misattributing behaviour
Semi-structured interviews	Participants are asked questions to prompt them to discuss a prescribed set of issues or topics	• Older children and respondents	• Enables complex, in-depth and qualitative responses • Can allow issues and themes to surface that may not be reflected in questionnaires, and to understand the causal factors • Does not rely on respondents' reading skills	• Difficult to conduct at scale, requiring time and resources for training, administration and coding the data • May also be subject to faulty memory, reference and/or social desirability bias
Observations e.g. World Bank *Teach*, Classroom Observation in Preschool and Teacher Observation in Preschool (COPTOP)	Used by an external person to record behaviours and activities using an agreed protocol	• Adult observer provides information on other people, which may include young children	• May be used to collect either quantitative or qualitative data • Does not rely on participants for completion, so can be used with younger learners • Potentially more objective depending on who completes the protocol and their expertise	• Potential that children/teachers adapt their behaviour in light of observation • May be subject to reporting, reference, implicit and/or unconscious bias • Difficult to capture awareness and beliefs
Performance assessments e.g. International Social and Emotional Learning Assessment (ISELA), Head-Toes-Knees-Shoulders task	Learners engage in complex, real-world or simulated tasks to directly measure their SEL skills	• All respondents (including young children)	• Created to approximate real-world conditions and generate quantitative data • Does not rely on subjective judgments of teachers or other raters	• Requires substantial investments in training, administration and scoring • May be more influenced by socio-cultural or -economic factors than other measures
Peer nominations e.g. Peer Nomination Inventory of Depression	Students nominate classmates who show particular behaviours or skills	• Respondents in formal education	• Techniques well-established in certain contexts • Potentially useful to access authentic behaviours away from authority figures	• Difficult to conduct with large groups and requires high levels of expertise to score and interpret

Sources: Adapted from Assessment Work Group (2019), Duckworth & Yeager (2015), Taylor et al. (2018).

5.1 Self-reporting Scales

Self-reporting scales are a common means of data collection for research on SEL, mental health and wellbeing. Individuals, typically adults or older children, rate their *own* skills or wellbeing, often on an ordinal Likert scale. Such scales comprise statements of what people feel, think and do, for example, the extent to which they would agree with the statements: '*I can control my temper*' and '*I can handle whatever comes my way*'. They generate quantitative data and are particularly useful for accessing information about unobservable aspects of people's internal lives such as their perceptions, beliefs and emotions.

Scales may be administered via numerous formats, including one-to-one interactions, focus groups or online surveys at scale. In low- and lower-middle-income contexts, individual and group administrations may be most appropriate, especially where rates of literacy among respondents are low. Many existing scales are widely available and cover a range of issues and constructs; however, their suitability, language and cultural relevance must always be considered when using them for the first time in a new context.

The main weakness of self-reporting scales is that they provide a single and subjective perspective on a person's skills or welfare. This undermines the scope for comparison between respondents, for example, to know whether two people scoring 3 on an anxiety scale really feel the same way. Sources of bias also include social desirability, that participants trying to present themselves in a favourable light, or participants reporting aspirational rather than actual behaviours. Similarly, respondents may give answers based on what they think the researcher wants to hear. Finally, self-reporting scales presume a degree of awareness, of oneself, of others and one's interactions with others, and so may be unsuitable for young children, or at the very least, require careful adaptation and piloting for use with such groups.

Notwithstanding these issues, we decided to use self-reporting scales within the Ethiopian study. Discussions with colleagues in Ethiopia suggested that the respondents, aged 9-13 years, in public primary schools, were old enough to report on their own self-efficacy, emotional regulation and social skills. Self-reporting scales were also used in the 2018/2019 data collection and therefore offered the advantage of comparing responses from the same pupils, using similar or identical items for measuring their SEL over time. Furthermore, such scales appeared to offer the best means for exploring learners' internal experiences, their feelings, wellbeing and wider mental health.

Regarding administration, the scale items and options were read aloud by the research field staff on a one-to-one basis with the younger pupils in Grade 3, thereby removing the need for the respondents' own literacy or reading skills. The older learners in Grade 6 were surveyed in groups, with one or two field researchers providing examples to the class with an opportunity to practise at the beginning, before the pupils completed their own responses on paper. These approaches helped reduce the risk of social desirability bias by keeping learners' responses private and confidential from their peers.

5.2 Teacher and Caregiver Report Scales

Teacher and caregiver report scales go some way to addressing the weakness of self-reporting scales, especially among children and adolescents. With such scales, teachers, parents and other guardians rate children's SEL, mental health and wellbeing using established ordinal scales. Depending on the purpose and use of the research, parents and teachers may also be asked to score children's behaviours against benchmarks or norms, which may be prescribed by the particular instrument, 'international'[iii] standards, or more local guidelines tied to national curricula. In each case, the ratings are more objective and can be compared across different respondents and in different settings, for example where a child's teacher and parent score them using the same scale regarding their behaviours at school and home respectively.

Teacher and caregiver reports can be effective for generating quantitative data on diverse aspects of learners' outward SEL, skills and competencies. Teacher ratings in particular may benefit from strong reference group knowledge, enabling them to draw comparisons and subtle distinctions between similarly aged children. However, caregiver and teacher reports also have certain limitations. They are less effective at capturing information regarding learners' internal experiences, their true beliefs, emotions and awareness. Reports by both teachers and parents may also give way to implicit or unconscious biases, over- or understating a child's abilities based on their own preferences or prejudices.

Similarly, using teacher and caregiver reports in low- and lower-middle-income contexts presents additional challenges. First, teachers who instruct large classes of 50 pupils or more may find it difficult to report on the skills of any one child. Second, parents and guardians may not have attended school themselves and may not be functionally literate, and therefore scales should be administered one-to-one or orally in small groups. Finally, and as with self-reports, both teachers and parents might be incentivised to provide particular responses if they think they might attract additional support or resources for the specific school or family. For these reasons, we did not use teacher or caregiver ratings for the Ethiopian study.

5.3 Semi-structured Interviews

Semi-structured interviews provide a flexible tool for capturing rich and in-depth qualitative data regarding SEL, mental health and wellbeing. In each case, they involve interviewers posing questions to prompt discussions on prescribed topics or issues with a wide range of respondents, from adults to children. Interviews enable researchers to investigate complex emotions, feelings and beliefs, to explore items that may have been missed in questionnaires, and to examine some of the relevant causal factors for good wellbeing or poor mental health.

The key challenge of using interviews is that they are particularly resource-intensive to analyse at scale, in both high- and low-income contexts compared with other research methods. For thorough analysis, interview responses should be transcribed verbatim before they can be coded and interpreted, which is

generally a time-consuming process, and can entail 5-6 hours of transcription for each hour of interview (Cohen, Manion & Morrison, 2017). Interviews can also be subject to biases, particularly the influence of the interviewer in terms of how the questions asked. In low-income countries or in interviews with children, power imbalances, and the participants' comfort levels (or lack thereof) in sharing their true views, may also affect the answers given. Given these constraints, the Ethiopian study did not include any interviews in its data collection.

5.4 Observations

Observations may help address some of the biases inherent in reports and interviews concerning children's SEL, mental health and wellbeing. While reports and interviews enable respondents to describe their skills and welfare in more or less favourable terms, observations allow researchers to record *actual* behaviours, activities and practices using predefined protocols. As such, they provide insights into externally displayed skills and emotions, but not the more internal aspects of learners' experiences.

Observations can be used to generate either quantitative or qualitative data across a range of respondents, including young children. They can involve one or more observers and can be used in different settings, not least classrooms and playgrounds when researching learners and adolescents. However, observations are not immune from bias, which can arise through the observer(s) paying particular attention to certain incidents and not others, or applying preconceptions to their interpretation of key events. Most importantly, the subjects, whether teachers or children, may adapt their ordinary behaviours while under observation, in a process known as 'reactivity', or the Hawthorne effect (Cohen, Manion & Morrison, 2017; Robson, 2011). This effect may be increased or reduced depending on how accustomed the participants are to the presence of external researchers or observers. Classroom observations in low-income countries with large class sizes may also be more difficult if they seek to focus on individual children's behaviours. Nevertheless, classroom observation of SEL can be particularly relevant to efforts to identify and improve teaching practices.

Therefore, to complement the self-report scales administered to students, we also carried out classroom observations in each of the schools included in the RISE Ethiopia research. We adapted the World Bank's *Teach* tool[iv], which has been used successfully in primary schools in other low-income contexts. Members of the RISE team have also adapted and used this tool (in collaboration with the World Bank) for secondary school classrooms in Rwanda (Carter et al., 2020). The *Teach* tool measures how well teachers foster SEL, which encourages students to succeed both inside and outside the classroom. Specifically, it measures time-on-task and quality of teaching practices, including classroom culture, instruction and socio-emotional skills. Each of these elements is linked to specific behaviours which are characterised as low, medium or high, based on the evidence collected during the observation. These behaviours are then translated into a 5-point scale that quantifies teaching practices. The information

gathered through this tool can then be used to understand and improve teachers' development of students' socio-emotional capacities (Carter et al., 2020).

5.5 Performance Assessments

Performance assessments involve research participants undertaking simulated tasks or activities, often designed to mirror real-life experiences, to measure particular aspects of their SEL. They can be used across a wide range of respondents, from adults to young children, and typically assess skills and competencies rather than attitudes, beliefs or emotions. The assessments tend to generate quantitative data and offer a greater degree of neutrality than reports, interviews and even observations, since the conditions are controlled to minimise the effects of any bias.

Some performance tools also include vignettes − short descriptions of hypothetical people or situations to elicit information about learners' behaviours and competencies. For example, the International Social and Emotional Learning Assessment (ISELA) created by Save the Children presents learners with imaginary scenarios, featuring angry or upset children to understand their levels of empathy, emotional awareness and how they would respond to cases of conflict. In ISELA, the learners' answers are rated to produce quantitative scores, but children's reactions to vignettes can also be recorded in detail to generate more qualitative data.

Direct assessments, including vignettes, can nevertheless be time-consuming and resource-intensive to set up, administer and score. Above all, there are questions around how transferable they are between different contexts, and whether certain aspects of SEL mean the same thing in diverse cultures. As discussed in section 4, different countries and regions hold diverse beliefs and expectations, for example the relevance of time for their daily lives, and how children should behave at home and around their elders (Brucki & Nitrini, 2008; Hsieh, 2004; Martini, 1996; Wanless et al., 2011). Each of these can affect how learners perform on assessments, which highlights the importance of their careful selection and adaptation in collaboration with local experts. For example, the classic test of supposed self-control, whereby children are offered one marshmallow immediately or two after a short wait, may function very differently in contexts of food insecurity or where children do not trust the researcher's promise (Mischel, 2014).

Given these reasons, the study in Ethiopia did not use any performance assessments to measure learners' SEL, mental health, or wellbeing. In the absence of established tasks that have already tested pupils in Ethiopian schools, it would have been onerous to develop, validate and then implement such assessments alongside the wider data collection. More importantly, direct assessments would not have been the most appropriate technique for measuring SEL, mental health or wellbeing in the current study, not least given its focus on skills such as self-efficacy and emotional regulation, which are better suited to other research methods.

5.6 Peer Nominations

Peer nominations provide a more participatory approach to reporting scales. They offer insight into children's social skills and welfare, including their friendships, acceptance or exclusion. They involve asking learners to nominate the peers that they like or dislike, or that display particular behaviours, traits or skills. Peer nominations produce quantitative data and can provide access to children's more private worlds and interactions, away from the oversight of parental or teacher supervision.

However, such an approach can be difficult in large groups or classes, particularly in the school settings that are common in low-income countries. Children's preferences and responses may also be skewed where participants speak different languages or hail from diverse ethnic groups, which can shape their interactions in both overt and subtle ways. Similarly, the reliability and validity of the results may be affected by social desirability bias, as well as fluctuating school attendance, where class compositions vary from one day to the next. For these reasons, peer nominations appeared to be inappropriate for measuring learners' SEL in the Ethiopian study.

5.7 Triangulation

As described above, different research methods have different advantages and disadvantages that affect their use, validity and reliability in diverse contexts. Often, the selection of one or more approach in a study will involve a trade-off or compromise between alternative options. However, collecting data using multiple techniques with the same respondents can help to triangulate the results and strengthen the rigour and validity of the final findings.

For example, children's self-control has historically been measured through parent or teacher ratings, direct assessments or observations of their behaviours in situations where they are exposed to certain challenges (Diamond, 2013). These provide relatively objective assessments of children's ability to ignore distractions or resist temptations, at least in the relevant setting(s). However, self-control can also be measured through self-reports, whereby the children themselves comment on how easy or difficult they find specific tasks or activities. Of course, children may feel inclined to over or understate their abilities, but using this additional method may shed further light on their true behaviour when teachers, parents or other adults are not physically present.

In the Ethiopian study, we used both self-reporting scales and lesson observations to help triangulate the findings in relation to pupils' SEL, mental health and wellbeing. On the one hand, the responses provided insight into the learners' internal and external experiences, their beliefs, attitudes and social interactions; while on the other hand, the lesson observations revealed the extent to which teachers were fostering key aspects of SEL within the classroom.

6. Identification of Measures and Selection for the Ethiopian Study

With the data collection methods identified, the next phase of the research process involved selecting appropriate tools and measures for conducting the study. The research in Ethiopia used both self-reporting scales and lesson observations, and so we sought to choose existing rating scales and observation protocols that had already been established and successfully validated. As discussed in section 4, assessing SEL, mental health and wellbeing draws on both internal and external skills, traits and behaviours, and it was therefore necessary to identify clear characteristics that could be related or observed, and which appeared to represent the key construct(s) of interest (Yorke, Wole & Rose, 2020). To aid this process, we referred to several websites[v] that offered databases or lists of tools for measuring SEL, some of which also included search functions to filter instruments according to particular research specifications.

For studies of SEL, mental health and wellbeing in low- and lower-middle-income contexts, two additional factors may further affect the identification of appropriate research measures. These concern where the tools were originally developed and previously used, and issues of age-appropriateness.

6.1 Origins and Prior Use

Sections 3 and 4 above noted the historical dearth of research on SEL, mental health and wellbeing in low- and lower-middle-income countries, and the relevance of culture in understanding and assessing them. Indeed, most of the international instruments to measure such aspects of personal welfare and development have been created in high-income contexts. As discussed above, this raises questions around transferability, whether tools can be applied across diverse settings and whether constructs have the same meaning and value in different cultures.

Evidence of prior use in environments similar to the specific research setting can go some way towards addressing these concerns. Many of the instruments reviewed during the Ethiopian study were established in Europe or North America, but had been successfully used and validated in numerous low-income contexts. Nevertheless, it was essential to consider the application of the measures in a new setting and the need for adaptation. This included adjustments to account for cultural sensitivities and translation into eight different languages[vi], which was undertaken in close collaboration with local knowledge experts and specialists. Details regarding the process of adaptation and translation in Ethiopia is set out in section 7.

6.2 Age-Appropriateness

Selecting age-appropriate instruments to measure children's SEL, mental health and wellbeing in low-income settings is also an important consideration. As discussed in section 4, conceptions of childhood vary between contexts, in which parents, teachers and policymakers in high- and low-income countries may hold different beliefs about at what age children should be able to perform certain tasks, take on

household responsibilities or contribute to family income (Burman, 2008). On the one hand, these may create opportunities for children to develop aspects of SEL like self-efficacy by cooking a meal or caring for a younger sibling; but on the other hand they may deprive learners of valuable time and space to play with peers and learn or practise other important skills. Similarly, financial pressures in low-income settings can expose young children to physical dangers or psychological strains that adversely affect their mental health and wellbeing, for example, if they are engaged in hazardous child work. Consequently, there may be wide differences in the skills and life experiences of children of the same age across diverse contexts, and thus instruments created and normed for learners of certain ages in wealthier contexts, may be wholly inappropriate for children in another.

Age-related factors may affect the means available for administering measures, whether using surveys, self-reports or direct assessments. In particular, years of schooling do not always equate with literacy and numeracy, and so simple questions that can be read by a 10-year-old in one context may pose a greater challenge for 10-year-olds elsewhere. Similarly, interpreting statements and tasks can place further mental and language demands on participants, to try and remember the different responses available, to articulate their position or to untangle the use of first and second person. As such, the timing and display of developmental milestones may vary based on certain cultural or cognitive factors in the relevant context (Raikes et al., 2017).

Existing tools that are available for low- and lower-middle-income contexts, such as IDELA and MELQO, have tended to focus on younger (pre-school-age) children. Both of these tools contain various measures of SEL, for example, using questions about how well children ignore distractions and understand other people's emotions. They were created for younger children aged 3-6, not least to gauge their readiness to start primary education (Pisani, Borisova & Dowd, 2018; Raikes et al, 2019; Wolf et al., 2017; Wolf & McCoy, 2019). Their relevance and application to older learners, whether primary-aged or adolescents for studies such as this one in Ethiopia, are therefore open to question.

6.3 Measures for the Ethiopian Research

Several additional practical factors drove the selection of measures in Ethiopia. These included (i) previous use; (ii) age-appropriateness; (iii) whether the instruments were freely available and accessible or required the payment of licence fees; and (iv) whether there were any restrictions on their adaptation or translation for the Ethiopian context. We also examined their psychometric properties such as validity and reliability, where reported. In each case, the decisions were led and guided by colleagues and experts at Addis Ababa University and the Policy Studies Institute. Table 3 lists the main measures that were considered, and the reasons for their subsequent inclusion or exclusion.

Table 3 – Key Tools Reviewed for the Ethiopian Study based on Identified Criteria

Domain	Scale	Focus	Reference	Age Range	Prior Use	Comments and Use
SEL	Academic Self-Concept Questionnaire (ASCQ)	Student Effort and Confidence	Liu, Wang and Parkins (2005)	12-13	Ethiopia, India, Singapore, Vietnam	Full scale available. Included in the previous RISE Ethiopia data collection and initial pilot but data showed little variation. *Excluded*.
	Big Five	Task performance, emotional regulation, collaboration, engaging with others, open-mindedness (also self-efficacy and -regulation)	Goldberg (1992)	Not specified	Worldwide, including Bolivia, China, India, and Iran	Scale available online. Items reviewed by Ethiopian specialists but deemed to be unsuitable. *Excluded*.
	Children's Self Report Social Skills Scale (CS4)	Social Skills	Danielson, & Phelps (2003)	9-12	Egypt, Estonia, Turkey	Scale available in journal. Included in previous data collection and initial pilot. Adapted items showed good properties. *Included*.
	Matson Evaluation of Social Skills with Youngsters (MESSY)	Social Skills	Matson, Rotatori & Helsel (1983)	4-18	China, India, Turkey	Scale free and available. Included in initial pilot. Adapted items showed good properties. *Included*.
	School Engagement Instrument (SEI)	Family Support and Teacher-Student Relationship	Appleton, Christenson, Kim & Reschly (2006)	11-18	Ethiopia	Scale free and available. Included in previous data collection and initial pilot. Adapted items showed good properties. *Included*.
	Self-Regulation of Learning Self-Report Scale (SRL-SRS)	Self-efficacy	Toering, Elferink-Gemser, Jonker, van Heuvelen & Visscher (2012)	11-17	Netherlands	Full scale available. Included in initial pilot. Data showed good properties. *Included*.
	Student Learning in	Emotional Regulation	Forsberg, Schultz, Lodi	12-16	Palestine	Full scale available.

	Emergency Checklist (26) (SLEC-26)		& Tubbs Dolan (2019)			Included in initial pilot. Data showed good properties. *Included*.
Wellbeing	Cantril's Ladder	Life Satisfaction	Levin & Currie (2014)	7-18	Ethiopia	Deemed to be too complex by Ethiopian specialists. *Excluded*.
	WHO Wellbeing Index	Wellbeing	WHO Collaborating Center for Mental Health (1998)	9-18	Japan, China, Denmark	Scale free and available. Included in initial pilot. Data showed good properties. *Included*.
Mental Health	Screen for Child Anxiety Related Emotional Disorders (SCARED)	Anxiety, Significant School Avoidance	Birmaher, Khetarpal, Brent, Cully, Balach, Kaufman & Neer (1997)	8-18	Global (but mainly North America)	Scale free and available. Included in initial pilot but data not normal or uni-dimensional. *Excluded*.
	Strengths and Difficulties Questionnaire (SDQ)	Emotional Symptoms, Peer Relationship Problems and more	Goodman, Meltzer & Bailey (1998)	4-16	Worldwide, including Cambodia, Ethiopia, India, Nigeria	Legal restrictions prevented timely adaptation or translation. *Excluded*.

For example, Cantril's Ladder invites participants to place themselves on a 9-point 'ladder of life', to indicate where they feel at the moment, and where they expect to be in 4 or 5 years' time (Camfield & Tafere, 2009). The measure had been previously used to gauge wellbeing and life satisfaction among Ethiopian respondents in the Young Lives study, but colleagues thought that it would be too complex for learners in the present research. Similarly, the Strengths and Difficulties Questionnaire (SDQ) is well established, freely available and has been used with learners around the world, including in low-income countries. An Amharic version of the teacher-report questionnaire already exists, which could have enabled its rapid deployment in Ethiopia. However, copyright restrictions limited the translation and adaptation of the SDQ instruments into other languages spoken in Ethiopia. They also prevented us from converting the existing Amharic tool from a teacher-reporting format to a self-reporting version, and thus the SDQ could not be used in the study.

7. Adaptation of Scales

Following the selection of suitable scales as described above, the measures needed to be reviewed and adapted for use in Ethiopia. This included: a) creating a pool of items; b) determining appropriate wording and response options; c) translation into Amharic and other languages; and d) piloting the tools.

7.1 Creating a Pool

As a first step in the adaptation process, we generated a pool of 91 items drawn from nine selected scales shown in Table 3. These were deemed to provide a comprehensive representation of SEL, mental health and wellbeing, that cut across the main international conceptual frameworks. Most importantly, they addressed aspects of SEL, mental health and wellbeing that were relevant to the specific Ethiopian context and that offered particular significance for learners during and after the COVID pandemic, as explained in Yorke et al. (2021).

To refine the pool, each specialist independently reviewed the items and considered both their relevance to the present study, and their cultural suitability for use with learners in Ethiopia. They then specified whether they thought each item should be retained in its original form, amended to make the wording clearer, shorter or more appropriate, or dropped altogether. For example, the experts queried the use of "*I get a lot of headaches, stomach-aches or sickness"* as an indication of emotional difficulties, suggesting that in Ethiopia, such physical issues can result from a host of different causal factors. Similarly, they challenged the item "*I look at people when I talk to them*" as being less relevant to social skills in the particular cultural context. In a few instances, the experts also proposed new items which they considered suitable and valuable, such as "*If I am hungry there is enough to eat at home*" and "*I am proud of my clothes*" to gauge children's levels of reported wellbeing.

Following the review, we compared areas of consensus and divergence across the items. In many cases, there was agreement regarding which items might be too complex or otherwise inappropriate for the Ethiopian pupils, which enabled us to refine the pool further.

7.2 Item Wording

The use of both positive and negative wording for scales has often been recommended to increase validity and identify participants who engage in certain response styles, irrespective of content (Dalal & Carter, 2015). This could include participants who provide the same or similar answers for all questions, which could indicate either an acquiescence response, an extreme response or social desirability bias (Van Vaerenbergh & Thomas, 2013; Van Sonderen, Sanderman & Coyne, 2013). Negative items include terms such as 'not' or 'never'), or the use antonyms for the other scale items (Van Sonderen, Sanderman & Coyne, 2013; Weems, Onwuegbuzie & Collins, 2006). Examples from the Ethiopian study include "*I do not have a good imagination*" and "*I always do poorly in tests*".

In the present research, however, we opted to use positive items only, such as, "*Other people like me*", and "*It is easy for me to stick to my aims*". Some researchers have cautioned against the use of negative items and raised a number of problems that are associated with mixing formats. For example, negative and positive items may not capture the same meaning and may generate different emotional responses from participants (Weems, Onwuegbuzie & Collins, 2006). They tend to display different psychometric

properties and are found to be statistically different from one another, which affects the factor structure, reliability and validity of the relevant scale. Negative words and items may be difficult to translate across languages, and in contexts like Ethiopia where over 80 languages are spoken, it is important to keep the wording as simple as possible (Yorke & Ogando, 2018). Mixing items may also introduce more measurement error, or require different mental processing skills, making them more burdensome for respondents with lower reading ability, and confusing for younger participants (Dalal & Carter, 2015; Roszkowski & Soven, 2010; Suárez-Alvarez et al., 2018; Van Sonderen, Sanderman & Coyne, 2013; Weems, Onwuegbuzie & Collins, 2006). Finally, while the inclusion of negative items may help to identify response patterns, they do not prevent such patterns from occurring in the first place.

In light of the above, the use of mixed format scales and negative items must be considered carefully when researching SEL, mental health and wellbeing, particularly with children. Any advantage by way of increased validity may be undermined if the scale items become too confusing or ambiguous for respondents to comprehend. In certain contexts, it is also culturally inappropriate to disagree with someone and therefore negative items may be unsuitable. Indeed, previous data collection in Ethiopia by RISE and Young Lives highlighted the difficulties of using negative items, so the scales in the current study retained positively worded items only.

7.3 Response Options

Similarly, the use of Likert scales to capture ordinal responses presents numerous issues which can affect the quality, validity and reliability of the resultant data. These include: a) the use of a neutral midpoint; b) the number of responses available; and c) if and how the points are labelled. Figure 3 shows some of the possible options and the differences between them. In the Ethiopian research, participants were presented with the first option, and each learner was invited to express their agreement or disagreement with a statement on SEL, mental health or wellbeing by selecting from 5-point scale, with all responses labelled and the options moving from negative to positive, from left to right.

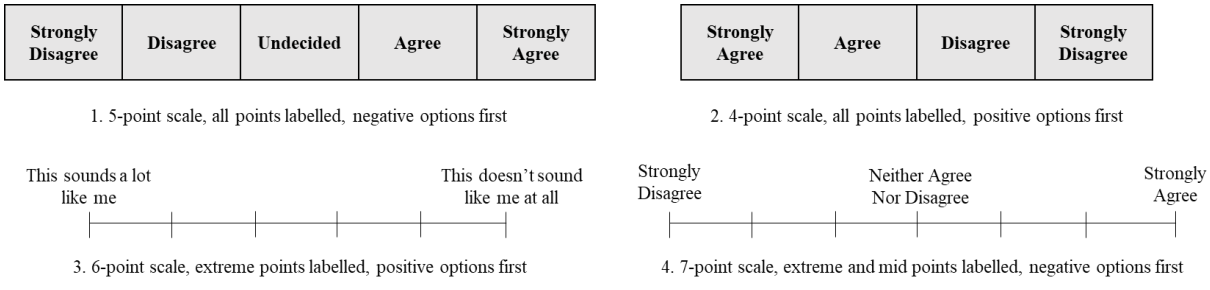

Figure 3 – Possible Scale Response Options

The inclusion or exclusion of a neutral midpoint on a scale relates closely to the number of response points used. Scales with an odd number of available responses (options 1 and 4 in Figure 3) typically allow for the selection of a middle option where participants do not wish to answer a particular item or are unable to do so, either due to difficulties in understanding the statement or as a result of a genuine

lack of opinion. Evenly numbered scales (options 2 and 3), however, force such ambivalent respondents to make a choice and position themselves towards either end of the scale. This can prompt greater reflection, but it can also distort results and increase the likelihood of non-response. Research has also shown that ambivalent participants are more likely to respond negatively out of frustration towards scales when forced to make a decision (Weijters, Cabooter & Schillewaert, 2010). For these reasons, scales with an odd number of responses that include a neutral midpoint are widely considered to be more accurate, and more reflective of participants' true positions.

Regarding the specific number of responses, various studies have examined the apparent effect of differently numbered scales on their research findings, for example, the use of scales with 5, 7 or 11 response points (each of which assumes the use of a midpoint). Factors concerning the selection of an appropriate scale can be considered both in relation to the quality of the resultant data, and the demands placed on respondents. Larger scales with more response points (for example, 7, 9 or 11) are considered to be more sensitive as they offer a better gradation of opinions, and may provide greater normality in terms of their distribution (Leung, 2011). However, Revilla, Saris and Krosnick (2014) argue that using more than 5 response points gives rise to a loss of data quality on agree-disagree scales.

Indeed, the quality of the data will likely depend on various respondent factors, not least their cognitive capacities, motivational dispositions and the circumstances in which they are providing their opinions. For example, participants may be less likely to provide considered responses on larger scales where they are not emotionally invested in the outcome of the survey, or are taking part for reasons of compliance (Krosnick & Presser, 2010). Similarly, larger scales place greater demands on participants' cognitive load to understand and remember the different response options. For these reasons, the study in Ethiopia limited the scales to 5 points, which was deemed to offer sufficient data quality and granularity. The earlier assessment of learners' SEL had used a 3-point scale to maximise simplicity and comprehension, but the resultant data were found to be highly skewed and offered little variation.

Related to the number of scale responses is the issue of labelling the different points, in particular whether all options should be labelled or only those at the extremes. Generally, labelling plays an important role in ensuring participants' understanding and accuracy in their answers. Ordinarily, it becomes more complex the larger the number of scale responses; but this also depends on the precise construct being assessed, for example, agreement ('*agree'/'disagree'*), or the extent to which a statement reflects the participant's position ('*this sounds a lot like me*'/'*this doesn't sound like me at all*'). There is also some evidence that labelling makes responses more reliable, more accessible, and therefore more attractive, which could distort responses where only the extreme points are labelled (Weijters, Cabooter & Schillewaert, 2010).

In light of these considerations, the Ethiopian research labelled each of the five response options to ensure that participants understood their meaning. In addition, the negative responses conveying

learners' disagreement were presented first to balance any propensity to select the initial options with the social desirability bias to agree with the items and statements.

7.4 Translation, Training and Piloting

With the measures, scale items and response options adapted, the next steps involved translating the tools into the relevant languages, training the field staff or enumerators to administer them, and piloting the instruments with a sub-sample of the intended population.

Effective translation is particularly important while researching SEL, mental health and wellbeing in a multilingual and multicultural setting like Ethiopia. Learners within the school may speak the same language, but come from different cultural groups, and so the translations may need to acknowledge or accommodate such variations. There is also the question of translation priorities, for example, whether word-for-word accuracy and equivalence is essential, or whether the focus should be on simplicity and intelligibility. In SEL research, and especially with young learners, the latter should take precedence.

The translations in the current study on SEL, mental health and wellbeing in Ethiopia were undertaken carefully by experienced translators. In each case, the translators were native speakers of the relevant language and had been involved in translating similar instruments during previous rounds of data collection. The whole process was also closely coordinated and monitored by bilingual members of the core team, who understood the purpose of the tools, and could provide detailed guidance in the case of any ambiguity.

Discussions during the training of field staff raised a few further issues concerning the translations. One field worker queried the appropriate response to "*I am proud of my shoes*" for children who were not wearing shoes. Another highlighted the ambiguity of the statement "*I trust my teacher*", since trust could arise between pupils and their teachings in several aspects of their interaction, such as grading and communications with parents. In both cases, the field staff encouraged the children to report their overall feeling, but the central midpoint on the response scales also provided an option for participants who were truly undecided.

The instruments were administered by a total of 47 enumerators, who received training from the RISE Ethiopia team to conduct the data collection. They were trained on each instrument orally, then on paper and finally with tablets. Each field worker was already proficient in using tablets for surveys but was given a full day to practise implementing the various tools, of which approximately 2.5 hours were dedicated to training on the SEL measures.

Piloting instruments is especially important where measures or scales have been adapted from one context for use in another, which is often the case for studies of SEL, mental health and wellbeing in low- and lower-middle-income countries. It also provides a valuable opportunity for enumerators trained

for the study to practise with real respondents and to raise any questions or challenges they faced in conducting the preliminary research.

The piloting in Ethiopia involved experienced enumerators from the wider RISE programme and helped to check that the tools functioned as planned. For example, the pilot data contained missing values which highlighted inconsistencies between enumerators, in terms of how certain follow-up questions were asked, on the basis of learners' prior responses. This issue was fixed for the main data collection through the introduction of an automated skip function on the pre-programmed tablets.

Further analyses of the pilot data examined whether they offered adequate variation and calculated psychometric properties such as their reliability and validity. The descriptive statistics showed that the items worked well, except for one scale (anxiety), which was not normally distributed and did not achieve uni-dimensionality. Similarly, the pupils had a tendency to agree with the statements in another scale (student effort), which showed little variability in their responses. In light of such findings, we removed those specific items and the main data collection proceeded using the remaining six scales.

Overall, careful adaptation of the selected measures was critical for the Ethiopian study to ensure the rigour and validity of its data collection. Throughout the process, Ethiopian colleagues guided the choice of particular scales and items, proposing amended wording in numerous instances to address cultural factors or linguistic ambiguities. For reasons of simplicity and consistency, we used positive statements only but also included neutral midpoint options for participants who felt undecided in their choices. Finally, thorough translation by experienced translators and piloting by trained field staff ensured that the tools functioned as intended, and captured valuable data on students' SEL, mental health and wellbeing.

8. Summary and Conclusion

This paper has explored methodological issues regarding research on children's SEL, mental health and wellbeing. In particular, it has highlighted important considerations, challenges and opportunities when measuring learners' personal welfare and development in low- and lower-middle-income countries. In so doing, the paper drew on lessons from our recent study among students attending primary schools in Ethiopia. This research used self-reporting scales with Grade 3 and 6 pupils to examine their a) self-efficacy; b) emotional regulation; c) social skills; d) important aspects of their SEL; as well as e) their overall mental health and wellbeing. The participants were further surveyed regarding their learning experiences during recent school closures, and Grade 6 mathematics classes were observed to understand how teachers promoted SEL during their lessons.

Together, these data will provide valuable insights regarding Ethiopian students' SEL and wellbeing, how they have been affected by the school closures, and how they interact with more traditional academic outcomes like literacy and numeracy. Future rounds of data collection may shed light on

Ethiopian learning trajectories now that schools have reopened, but they should also prioritise research on children's mental health. In the current study, the main measures for assessing mental health posed particular challenges, from the anxiety scale which failed to show uni-dimensionality during the pilot, to the SDQ whose adaptation and translation were legally restricted; and thus we used the short WHO scale as a combined measure of both mental health and wellbeing. Given estimates of mental health problems among children and adolescents in sub-Saharan Africa, this should be a priority for further data collection going forward (Atilola, 2017; Belfer, 2008; Cortina et al., 2012).

Beyond Ethiopia, increasing attention is being given to children's SEL, mental health and wellbeing, the importance of which has been further highlighted throughout the COVID-19 pandemic (Yorke et al., 2021). This paper has therefore also identified key lessons from the research in Ethiopia, including certain recommendations for investigating and measuring SEL, mental health and wellbeing in similar low- and lower-middle-income countries. These include:

- Considering the impact of culture on understanding the meaning, significance and measurement of SEL, mental health and wellbeing in diverse settings, including how they interact with conceptions of childhood and expected behaviours;
- Choosing appropriate research methods based on the construct(s) of interest, the characteristics of the target population and any factors in the local environment;
- Selecting suitable tools in light of where they were developed and previously used, the age of the intended participants and any logistical factors, which may include the availability and accessibility of the measures; and
- Carefully adapting, translating and piloting the instruments with experienced field staff to ensure they function as intended.

Throughout these steps, research teams should include local experts and specialists to help ensure the cultural sensitivity of the study, its methods and tools, and to maximise the validity and reliability of the resultant data. By taking such an approach, international research on SEL, mental health and wellbeing may offer greater nuance and understanding of children's holistic learning and development beyond their traditional academic outcomes alone.

References

Abera, M., Tesfaye, M., Hanlon, C., Admassu, B., Girma, T., Wells, J.C., Kaestel, P., Ritz, C., Wibaek, R., Michaelsen, K. F., Friis, H. & Andersen, G. S. (2018). Body Composition during Early Infancy and Mental Health Outcomes at 5 Years of Age: A Prospective Cohort Study of Ethiopian Children. *Journal of Pediatrics*, 200, 225-231. DOI: 10.1016/j.jpeds.2018.04.055.

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 .

Atilola, O. (2017). Child mental-health policy development in sub-Saharan African: broadening the perspectives using Bronfenbrenner's ecological model. *Health Promotion International*, 32, 380-391. DOI: 10.1093/heapro/dau065.

Belfer M.L. (2008). Child and adolescent mental disorders: the magnitude of the problem across the globe. *Journal of Child Psychology and Psychiatry*, 49(3), 226-236.

Birmaher, B., Khetarpal, S., Brent, D., Cully, M., Balach, L., Kaufman, J. & Neer, S.M. (1997). The Screen for Child Anxiety Related Emotional Disorders (SCARED): scale construction and psychometric characteristics. *Journal of the American Academy of Child and Adolescent Psychiatry*, 36(4), 545-53. DOI: 10.1097/00004583-199704000-00018.

Bronfenbrenner, U. (1979). *The Ecology of Human Development*. Cambridge MA: Harvard University Press.

Brucki, S. M. D. & Nitrini, R. (2008). Cancellation task in very low educated people. *Archives of Clinical Neuropsychology*, 23, 139-147.

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. (2020). Measuring teachers' encouragement of socioemotional skills in the secondary classroom. In Smart, A. & Sinclair, A. (Eds.) *NISSEM Global Briefs: Educating for the social, the emotional and the sustainable: Pedagogy, practices and materials* (pp.76-97). NISSEM.

Cohen, L., Manion, L. & Morrison, K. (2017). *Research Methods in Education* (8th ed.). Abingdon: Routledge.

Coleman, W., & Cureton, E. (1954). Intelligence and Achievement: The 'Jangle Fallacy' Again. *Educational and Psychological Measurement*, 14(2), 347–351.

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, accessed on 9 February 2021.

Cortina, M. A., Sodha, A., Fazel, M., & Ramchandani, P. G. (2012). Prevalence of Child Mental Health Problems in Sub-Saharan Africa: A Systematic Review. *Archives of Pediatrics & Adolescent medicine*, 166(3), 276-281.

D'Sa, N. (2019). *International Social and Emotional Learning Assessment: Adaptation and administration guidance*. Washington, DC: Save the Children.

D'Sa, N. & Krupar, A. (2019). *Psychometric Properties of the International Social and Emotional Learning Assessment among Syrian Refugee Children in Kurdistan, Iraq: Technical Working Paper*. Washington DC: Save the Children.

Dalal, D. K., & Carter, N. T. (2015). Negatively worded items negatively impact survey research. In C. E. Lance & R. J. Vandenberg (Eds.), *More Statistical and Methodological Myths and Urban Legends: Doctrine, Verity and Fable in Organizational and Social Sciences* (pp. 112–132). Routledge/Taylor & Francis Group.

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.

Diamond, A. (2013). Executive Functions. *Annual Review of Psychology*, 64(1), 135-168. DOI: 10.1146/annurev-psych-113011-143750.

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 from http://www.ncbi.nlm.nih.gov/pmc/articles/PMC4210770/, accessed on 31 March 2021.

Doebel, S. (2020). Rethinking Executive Function and its Development. *Perspectives on Psychological Science*, 15(4), 942-956.

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.

Forsberg, J.T., Schultz, J., 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*. Available at https://inee.org/resources/student-learning-emergency-checklist-slec?webform_id=toolkit_resources, accessed on 9 February 2021.

Goldberg, L. (1992). The Development of Markers for the Big-Five Factor Structure. *Psychological Assessment*, 4(1), 26-42.

Goodman, R., Meltzer, H. & Bailey, V. (1998). The Strengths and Difficulties Questionnaire: A pilot study on the validity of the self-report version. *European Child and Adolescent Psychiatry*, 7, 125-130.

Haroz, E. E., Jordans, M., de Jong, J. Gross, A., Bass, J. & Tol, W. (2017). Measuring Hope Among Children Affected by Armed Conflict: Cross-Cultural Construct Validity of the Children's Hope Scale. *Assessment*, 24(4), 528-539. DOI: 10.1177/1073191115612924.

Hsieh, M.-F. (2004). Teaching practices in Taiwan's education for young children: Complexity and ambiguity of developmentally appropriate practices and/or developmentally inappropriate practices. *Contemporary Issues in Early Childhood*, 5, 309-329. DOI:10.2304/ciec.2004.5.3.5.

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.

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.

Kim, H. Y., & Tubbs Dolan, C. (2019). *Social emotional response and information scenarios: Evidence on construct validity, measurement invariance, and reliability in use with Syrian refugee children in Lebanon: Technical working paper*. New York, NYU Global TIES for Children.

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.

Levin, K. A. & Currie C. (2014). Reliability and Validity of an Adapted Version of the Cantril Ladder for Use with Adolescent Samples. *Social Indicators Research*, 119, 1047-1063.

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.

Martini, M. (1996). 'What's new?' at the dinner table: Family dynamics during mealtimes in two cultural groups in Hawaii. *Early Development and Parenting*, 5, 23-34. DOI: 10.1002/(SICI)1099-0917(199603)5:1<23:: AID-EDP111>3.0.CO;2-D.

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.

Mischel, W. (2014). *The Marshmallow Test: Understanding self-control and how to master it*. London: Transworld Publishers.

Miyake, A., Friedman, N. P., Emerson, M. J., Witzki, A. H., Howerter, A., & Wager, T. D. (2000). The Unity and Diversity of Executive Functions and Their Contributions to Complex "Frontal Lobe" Tasks: A Latent Variable Analysis. *Cognitive Psychology*, 41(1), 49-100. DOI: 10.1006/cogp.1999.0734.

Nakigudde, J., Bauta, B., Wolf, S. & Huang, K-Y. (2016). Screening Child Socio-emotional and Behavioral Functioning in Low-Income African Country Contexts. *Jacobs J Psychiatry Behav Sci*, 2(2), 1-23.

Organisation for Economic Co-operation and Development (OECD). (n.d.) *Social and Emotional Skills: Well-being, connectedness and success*. Available at http://www.oecd.org/education/ceri/social-emotional-skills-study/, accessed on 31 March 2021.

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.

Qadir, F., Maqsood, A., us-Sahar, N., Bukhtawer, N., Khalid, A., Pauli, R., Gilvarry, C., Medhin, G. & Essau, C.A. (2018). Factor Structure of the Urdu Version of the Spence Children's Anxiety Scale in Pakistan. *Behavioral Medicine*, 44(2), 100-107, DOI: 10.1080/08964289.2016.1276427.

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. & Savre, R. (2019) Examination of school readiness constructs in Tanzania: Psychometric evaluation of the MELQO scales. *Journal of Applied Developmental Psychology*, 62, 122-134. DOI: 10.1016/j.appdev.2019.02.003.

Reeves, R.V. & Venator, J. (2014) *Jingle-Jangle Fallacies for Non-Cognitive Factors*. Available at https://www.brookings.edu/blog/social-mobility-memos/2014/12/19/jingle-jangle-fallacies-for-non-cognitive-factors/, accessed on 31 March 2021.

Revilla, M.A., Saris, W.E. & Krosnick, J.A. (2014) Choosing the Number of Categories in Agree-Disagree Scales, *Sociological Methods & Research*, 43(1), 73-97.

Robson, C. (2011). *Real World Research* (3rd ed.). Chichester: John Wiley & Sons.

Roe, E.M. (1988) Individualism versus Community in Africa? The Case of Botswana. *The Journal of Modern African Studies*, 26(2), 347-350.

Roszkowski, M. J., & Soven, M. (2010). Shifting gears: Consequences of including two negatively worded items in the middle of a positively worded questionnaire. *Assessment & Evaluation in Higher Education*, 35(1), 113-130.

Suárez-Alvarez, J., Pedrosa, I., Lozano Fernández, L. M., García-Cueto, E., Cuesta, M., & Muñiz, J. (2018). Using reversed items in Likert scales: A questionable practice. *Psicothema*, 30(2), 149-158. DOI: 10.7334/psicothema2018.33.

Talwar, V., Carlson, S. M., & Lee, K. (2011). Effects of a Punitive Environment on Children's Executive Functioning: A Natural Experiment: Punitive Environment. *Social Development*, 20(4), 805-824. DOI: 10.1111/j.1467-9507.2011.00617.x.

Taylor, J., Buckley, K., Hamilton, L., Stecher, B., Read, L., and Schweig, J. (2018) *Choosing and Using SEL Competency Assessments: What Schools and Districts Need to Know*. Available at https://measuringsel.casel.org/assessment-guide/considerations-for-sel-competency-assessment/, accessed on 31 March 2021.

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.

United Nations Children's Fund (UNICEF). (2016). *Harnessing the Power of Data for Girls: Taking stock and looking ahead to 2030*. New York: UNICEF. Available at https://data.unicef.org/resources/harnessing-the-power-of-data-for-girls/, accessed on 31 March 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, accessed on 31 March 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.

Van Vaerenbergh, Y., & Thomas, T. D. (2013). Response styles in survey research: A literature review of antecedents, consequences, and remedies. *International Journal of Public Opinion Research*, 25(2), 195-217.

Wanless, S. B., McClelland, M. M., Acock, A. C., Ponitz, C. C., Son, S.-H., Lan, X., … Li, S. (2011). Measuring behavioral regulation in four societies. *Psychological Assessment*, 23(2), 364-378. DOI: 10.1037/a0021768.

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.

Weijters, B., Cabooter, E. & Schillewaert, N. (2010). The effect of rating scale format on response styles: The number of response categories and response category labels. *International Journal of Research in Marketing*, 27, 236-247.

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. DOI: 10.1016/j.ecresq.2017.05.001.

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 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*. 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., Bayley, S., Wole, D. & Ramchandani, P. (2021). *Policy Brief: The Importance of Students' Socio-Emotional Learning and Mental Health in the time of COVID-19*. Available at https://riseprogramme.org/publications/importance-students-socio-emotional-learning-mental-health-and-wellbeing-time-covid-19, accessed on 31 March 2021.

Yorke, L., Wole, D. & Rose, P. (2020). *An Emerging Strategy for the Development of Culturally Relevant Scales to Capture Students' Socio-Emotional Learning*. Publication details to be available soon.

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.

[i] Further details can be found at https://www.younglives.org.uk/

[ii] The Resilience Research Centre examines pathways to resilience across different cultural contexts. Further information can be found at https://resilienceresearch.org/

[iii] Global benchmarks have typically relied on data from high-income countries, but recent research using tools such as IDELA and MELQO are helping to increase the representation of learners from low-income countries.

[iv] https://www.worldbank.org/en/topic/education/brief/teach-helping-countries-track-and-improve-teaching-quality

[v] Such websites include:
https://www.rand.org/education-and-labor/projects/assessments.html,
https://measuringsel.casel.org/assessmentguide/?accessform=true&position=Researcher%2Fprofessor and
https://educationendowmentfoundation.org.uk/projects-and-evaluation/evaluating-projects/measuring-essential-skills/spectrum-database/.

[vi] The surveys were ultimately conducted in seven languages as security issues prevented data collection in the Tigray region, and therefore use of the Tigriniya translations.