OECD Education Working Papers No. 207

Assessment framework of the OECD Study on Social and Emotional Skills

Miloš Kankaraš,

Javier Suarez-Alvarez

https://dx.doi.org/10.1787/5007adef-en

OECD

Organisation for Economic Co-operation and Development

EDU/WKP(2019)15

Unclassified **English text only**

22 October 2019

DIRECTORATE FOR EDUCATION AND SKILLS

Cancels & replaces the same document of 22 October 2019

ASSESSMENT FRAMEWORK OF THE OECD STUDY ON SOCIAL AND EMOTIONAL SKILLS

OECD Education Working Paper No. 207

Miloš Kankaraš and Javier Suarez-Alvarez (OECD)

This working paper has been authorised by Andreas Schleicher, Director of the Directorate for Education and Skills, OECD.

Miloš Kankaraš (milos.kankaras@oecd.org) and
Javier Suarez-Alvarez (javier.suarez-alvarez@oecd.org)

JT03453280

Acknowledgements

The authors would like to thank Rose Bolognini, Elizabeth Cooksey, Filip de Fruyt, Fiona Parsons, Andreas Schleicher and William Thorn for valuable feedback on earlier drafts of this paper. The authors would also like to thank members of the technical advisory group and delegates of the SSES network for their valuable contributions, which helped to consolidate this assessment framework. Translation (in French) was provided by Valentine Bekka. Rose Bolognini edited the initial draft of this paper, and Deborah Fernandez provided editorial support for the final draft.

Abstract

The OECD's new Study on Social and Emotional Skills aims to provide policy makers, educators, families and communities with a comprehensive set of tools to foster students' social and emotional learning. The Study's assessment framework – presented in this paper – is a result of an extensive literature review of previous research, existing frameworks and assessment approaches in the area of social and emotional skills. The framework, developed by a team of experts in various domains, incorporates evidence from psychology, education, organisational sciences, sociology, economy, and public policy. This framework describes the objectives, characteristics and expected outcomes of the Study. It presents the conceptual model of social and emotional skills assessed in the Study, their development, malleability and predictive value. The framework also discusses how factors in students' family, school and peer environment influence their social and emotional skills' development along with the contextual questionnaires designed to gather this information. The framework also presents the Study's design, assessment approach, instrument development process, sampling procedures and data collection methods.

Résumé

La nouvelle enquête de l'OCDE sur les compétences sociales et émotionnelles a pour objectif de fournir aux décideurs politiques, aux professionnels de l'éducation, aux familles et aux communautés un ensemble complet d'outils favorisant l'apprentissage social et émotionnel des élèves. Ce document présente le cadre d'évaluation de cette enquête qui est le résultat d'une analyse approfondie de recherches précédentes, de cadres d'evaluations existants et d'approches d'évaluation dans le domaine des compétences sociales et émotionelles. Le cadre, conçu par des experts de divers domaines, intègre des preuves issues de la psychologie, de l'éducation, des sciences organisationnelles, de la sociologie, de l'économie et des politiques publiques. Ce cadre d'évaluation décrit les objectifs, les caractéristiques et les résultats attendus de l'enquête. Il présente le modèle conceptuel des compétences sociales et émotionnelles évaluées dans l'enquête, leur développement, leur malléabilité et leur valeur prédictive. Le cadre – combiné avec le questionnaire contextuel construit pour rassembler ces informations – examine également comment l'environnement familial, scolaire et le rôle des pairs influencent le développement des compétences sociales et émotionnelles de l'élève. Le cadre présente également la conception de l'enquête, son approche d'évaluation, le processus de développement des instruments, les procédures d'échantillonnage et les méthodes de collecte de données.

Table of contents

Tables

Figures

Boxes

1. Introduction

Developing social and emotional skills not only helps people adjust to their environment and determines their success, but they also shape the larger communities and societies we live in. Resourceful, respectful and tolerant citizens who work well with others, and take personal and collective responsibility, are increasingly becoming the foundation of a society working towards the common good (OECD, 2017[1]; OECD, 2015[2]).

Social and emotional skills are malleable, and they can be shaped by a variety of individual and contextual factors, including direct policy interventions (Chernyshenko, Kankaraš and Drasgow, 2018[3]). Children continuously develop these skills through interactions with family, friends and teachers, at home and at school. As their social networks expand during adolescence and early adulthood, new learning environments open up within the community and workplace. Although social and emotional skills can be developed at a later age, early and continuous development achieves the best results (Shuey and Kankaraš, 2018[4]).

The OECD initiated the Study on Social and Emotional Skills (SSES or the Study) with the goal of gathering empirical evidence on the social and emotional skills of young people in school. By gathering a comprehensive set of information on students' families, schools and community learning contexts, the Study aims to provide policy-makers and educators with relevant information about the conditions and practices that foster or hinder the development of social and emotional skills in schools and other settings.

In this paper, we will present the Study's assessment framework, describing its objectives, characteristics and outcomes. This assessment framework is organised as follows:

- Section 1 begins with a narrative of the Study's background, providing context to the underlying goals and rationale for the Study. An overview of the Study follows, describing how it builds on prior OECD studies and how it will fill gaps in our current knowledge base.
- Section 2 presents the conceptual model of social and emotional skills used in this Study. The section focuses on research results about the development, predictive validity, malleability and cross-cultural comparability of each skill being assessed.
- Section 3 discusses the potential influence of contextual factors, such as family, school and peer environment on students' social and emotional skills development. Questions about many of these factors are included in four Contextual Questionnaires aimed at parents, teachers, school principals and the students themselves. This section outlines each of the questionnaire's content including a description of concepts and their relevance for child development.
- Section 4 focuses on how this survey assesses social and emotional skills and related contextual factors. It presents multiple aspects of the Study's design, including an overview of sampling procedures, data collection methods, the development of survey instruments, assessment approaches, and a timeline for the development of study instruments.

1.1. Study background

Today's world is made up of interwoven and continuously shifting economic, social and environmental conditions, changing how individuals and families live, and the ways that communities and economies operate. The increasing rate of technological innovations is also exponentially changing our environment, necessitating constant adjustments in the way we interact with the world around us and with each other. Such transformations mean that today's students are growing up in a far more diverse and less easily defined world than what their parents and teachers experienced when they were the same age. This is why it is crucial for educational leaders and decision-makers to consider how well their educational systems are preparing today's students for tomorrow's world, and what needs to be changed in order to stay relevant to our children's futures. Enabling education systems to better promote children's social and emotional skills development might be one of the best ways to fulfil such a role.

Interest in social and emotional skills has long been rooted in psychological and educational research. The large body of accumulated evidence shows that social and emotional skills have powerful consequences for many important life outcomes, such as educational achievement, employment, health or personal well-being (Chernyshenko, Kankaraš and Drasgow, 2018[3]; Kankaraš, 2017[5]; Kautz et al., 2014[6]; OECD, 2015[2]). Empirical evidence also shows that in many situations, these skills also play a role in improving educational attainment, employability and work performance, and civic engagement. Social and emotional skills have even been found to correlate more strongly with a wide range of quality of life outcomes, such as mental health and subjective well-being than with IQ and other cognitive skills (Chernyshenko, Kankaraš and Drasgow, 2018[3]). They also help to reduce anti-social and criminal behaviours and increase safety (Heckman and Kautz, 2012[7]; Heckman and Kautz, 2014[8]; Heckman and Kautz, 2014[9]; Kankaraš, 2017[5]; Kautz and Zanoni, 2014[10]; OECD, 2015[2]; Roberts et al., 2007[11]). Furthermore, research has identified the *inter-related* nature of cognitive and social and emotional skills (Cunha and Heckman, 2007[12]; Cunha, Heckman and Schennach, 2010[13]). Intersecting cognitive and socio-emotional skills empower children to succeed both within and outside of school. Evidence that social and emotional skills are instrumental in increasing cognitive skills underscores their importance as part of a model for children to lead meaningful and prosperous lives (Chernyshenko, Kankaraš and Drasgow, 2018[3]).

Social and emotional skills are also seen as crucial components of 21st century and employability skills (De Fruyt, Wille and John, 2015[14]; Trilling and Fadel, 2009[15]), because they are increasingly crucial for an individual's personal and career development, and being able to contribute productively to society (National Academy of Sciences, 2012[16]).

The term *"social and emotional skills"* therefore refers to individual characteristics manifested as consistent patterns of thoughts, emotions and behaviours, which can transform throughout life and influence important outcomes (Kankaraš, 2017[5]; Chernyshenko, Kankaraš and Drasgow, 2018[3]). The role and impact of social and emotional skills are increasingly critical for individuals to successfully navigate diverse and changing economies and societies. In conclusion, hindering social and emotional development negatively influences educational attainment, the transition from school into the labour market, productivity and job satisfaction, mental and physical health and overall well-being.

Box 1.1. Defining social and emotional skills

The domain of social and emotional skills is the subject of interdisciplinary research by researchers, academics, educators and practitioners, all from very different backgrounds. In consequence, there are many terms used to describe social and emotional skills and their broader conceptual frameworks (Kankaraš, 2017[5]). Terminology also differs across countries, time, and research and social contexts. For example, the vast scope of literature on the subject uses terms that have similar meanings, such as 21st century skills, life skills, essential skills, behavioural skills, non-cognitive skills, youth development assets, workplace or work readiness competencies, social-emotional learning, and character skills or strengths (Lippman et al., 2015[17]). Moreover, the social context shapes the terminology used: for example, employers refer to soft skills. Moreover, within particular fields terminology changes as well; in psychology, personality psychologists describe sub-dimensions of the Big Five model[1] (described in more detail in Section 2) as sub-domains, sub-elements, or facets. Developmental psychologists refer to these constructs as assets such as developmental assets (Benson, Scales and Syvertsen, 2010[18]). Within the field of education, social and emotional learning (SEL) is a widely-used term to describe social and emotional competences, non-cognitive or non-academic skills, behaviours and mind-sets (Lippman et al., 2015[17]). Within economics, the Nobel Prize winner, James Heckman, uses the term non-cognitive to distinguish social and emotional skills from cognitive skills commonly measured by IQ or standardised academic tests (Heckman, Stixrud and Urzua, 2006[19]). The term non-cognitive implies the absence of any cognitive activities. However, every aspect of mental functioning is based on some form of information processing and cognition (Duckworth and Yeager, 2015[20]). Further, prime examples of non-cognitive skills are social competencies which are fundamentally dependent on perception, memory and reasoning abilities – so much so that they are often seen as a form of intelligence (Murphy and Hall, 2011[21]; Kankaraš, 2017[5]).

In this Study, we use the term **social and emotional skills**. The OECD defines social and emotional skills as:

"…individual capacities that can be (a) manifested in consistent patterns of thoughts, feelings and behaviours, (b) developed through formal and informal learning experiences, and (c) important drivers of socio-economic outcomes throughout the individual's life" (OECD, 2015, p. 35[2]).

We favour the term skills rather than traits as the former indicates the possibility of change and development. The skills we have chosen to assess in this Study are malleable, representing potential targets for policy interventions. For clarity and consistency, we also use the term **sub-domains** rather than sub-elements or facets when referring to the sub-dimensions of the Big Five.

Developing these kinds of skills are also increasingly important for communities and nations as a whole, as they have been linked to increased levels of civic engagement, volunteering and social integration, better interpersonal trust and tolerance, and a decrease in anti-social and criminal behaviours. Despite their importance, large-scale international studies on social and emotional skills are still scarce. However, a growing realisation of their importance, especially in regards to future living and working environments has led to increased attention to this topic among researchers, policy-makers and practitioners. OECD studies such as the Programme for International Student Assessment (PISA), the Survey of Adult Skills (PIAAC) and the International Early Learning and Child Well-being Study (IELS) primarily focus on cognitive skills, such as reading/literacy and mathematics/numeracy. However, in recent years, PISA is now broadening its scope by assessing a growing set of social and emotional skills, such as academic self-efficacy (belief in one's ability to successfully complete the task at hand), perseverance, openness to knowledge, curiosity and civic engagement. PISA results show that these skills are related to important life outcomes and that they can be compared within and across cultural and linguistic boundaries. The OECD is taking this work further with a comprehensive international assessment of the social and emotional skills of school-age children, through the Study on Social and Emotional Skills (SSES).

1.1.1. How the SSES builds on prior OECD studies

The OECD's Directorate for Education and Skills recognises the importance of social and emotional skills and is broadening the metrics beyond traditional academic domains, such as reading and writing. OECD studies, such as PISA, PIAAC and IELS are covering a growing range of social and emotional skills.

The International Early Learning and Child Well-being Study (IELS) is a survey that assesses five-year-old children, identifying key factors that drive or hinder early learning development. The first five years of children's lives are crucial to their development. During this period, children learn at a faster rate than at any other time in their lives, developing basic cognitive and socio-emotional skills that are fundamental for their future achievements in school and later on as adults. These skills are also the foundation for their general well-being – how they cope with future successes and failures, professionally and in their personal lives. The survey produces scales on empathy, trust, pro-social behaviour and disruptive behaviour.

The Programme for International Student Assessment (PISA) is a triennial survey – since 2000 – that emphasises the functional skills that 15-year-old students have acquired in reading, mathematics and science literacy as they near the end of compulsory schooling. PISA also includes measures of general or cross-curricular competencies, such as collaborative problem-solving. In particular, the latest PISA cycles include scales on various types of academic self-efficacy, persistence, intellectual curiosity, meta-cognition and achievement motivation. In 2015, PISA gathered international evidence on the importance of social and emotional skills, publishing a volume detailing key positive and negative well-being indicators, such as life satisfaction, bullying and anxiety (OECD, 2017[1]). In this volume, *PISA 2015 Results (Volume III): Students' Well-Being*, PISA also explores the positive characteristics that promote healthy development such as interest,

[1] The "Big Five" model comprises five broad personality dimensions: Conscientiousness, Extraversion, Agreeableness, Emotional Stability (also called Neuroticism), and Openness to Experience. Each represents a cluster of related thoughts, feelings, and behaviours.

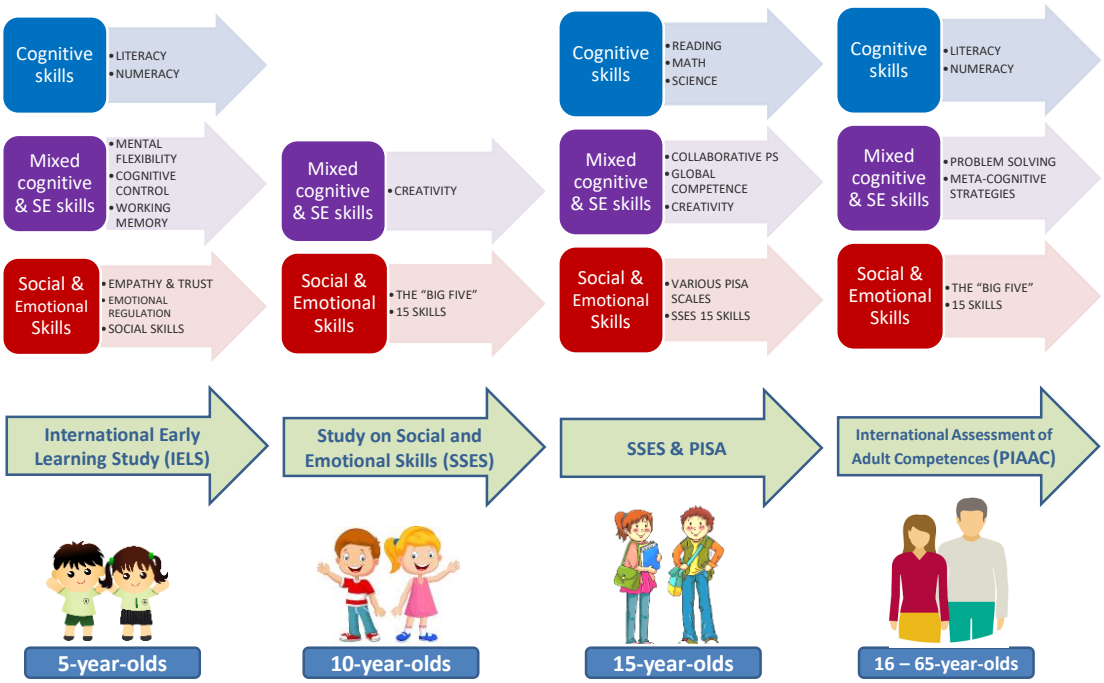
engagement, and motivation to achieve. Moreover, for the first time in 2018, PISA assessed students' global competence, providing information on their abilities to examine local, global and intercultural issues, understand and appreciate different perspectives and world views, interact successfully and respectfully with others, and take responsible action towards sustainability and collective well-being (OECD, 2018[22]).

The Programme for the International Assessment of Adult Competencies (PIAAC) has conducted two rounds of the Survey of Adult Skills to date (2011-12 and 2014-15). Approximately 250 000 adults between the ages of 16 and 65 have participated in over 40 countries. The study's primary focus is on literacy, numeracy and problem-solving. The first round of PIAAC also included scales on social trust and intellectual curiosity. The study takes this work further with a more comprehensive assessment of adults' social and emotional skills in the forthcoming third round of PIAAC in 2021-2022. Moreover, gathering information and data on how adults use these skills at home, at work and in the broader community will guide policy-makers on how to help individuals participate in society and for economies to prosper.

The Teaching and Learning International Survey (TALIS) focuses on teachers, teaching practices and how students learn. Twenty-four countries took part in the initial TALIS survey conducted in 2008 and 34 in the second round in 2013. Forty-eight countries are signed up to take part in the upcoming 2018 survey. TALIS develops and administers questionnaires to teachers and school principals to determine how countries can prepare teachers to face diverse challenges in today's schools and teach effectively to produce high-performing students. Among other topics, the survey also gathers information about teaching practices, such as active learning pedagogies, that are especially conducive to the promotion of the development of social and emotional skills.

The OECD is continuing to build on this work with a comprehensive international assessment of the social and emotional skills of school-age children, through the Study on Social and Emotional Skills. The Study together with other OECD surveys across the Directorate for Education and Skills are gathering international evidence of the importance in fostering the development of cognitive but also social and emotional skills throughout life (Figure 1).

Figure 1.1. Life-long learning: How the OECD measures skill development throughout life



1.1.2. How the Study on Social and Emotional Skill fills gaps in our current knowledge on these topics

The OECD Secretariat recently reviewed empirical evidence extensively to better understand the dynamics of skills development and the impact of social and emotional skills on individuals' social and economic outcomes (OECD, 2015[2]). Not only did they conclude that social and emotional skills are critical for the well-being of individuals, families and communities and that these skills influence overall levels of social cohesion and prosperity, but their review also uncovered **gaps** in our current knowledge base concerning children's social and emotional skills development:

- Large-scale studies mainly evaluate *cognitive skill development* among children at many different ages, within countries and internationally. And even though there are a growing number of large-scale international studies that collect data on the social and emotional skills of adults, *large-scale studies aimed at collecting internationally comparative data on young people's social and emotional skills development* are still scarce. The Study on Social and Emotional Skills aims to fill this gap by gathering empirical data on 10- and 15-year-olds that are robust, reliable and valid across different cultures, languages and diverse educational, national and local settings.
- Information still lacks on what types of family environments, social relations and educational policies and practices might work to enhance social and emotional skills. Moreover, the accumulated evidence in this regard is mostly limited to particular cultural settings in a limited number of countries. By gathering information from parents, teachers and school principals, as well as from students

themselves, the Study can compare the predictive value of different skills and contextual factors in students' environments relative to their social and emotional skills. In addition, the Study examines varying effects of the same family and school factors across a wide range of cultural and social settings.

- The Study is designed to provide policy-makers, education practitioners, parents and researchers a more comprehensive knowledge-base on where and how to improve systems, policies and practices in order to better support students' social and emotional skills development. By mapping a very comprehensive set of contextual factors, the Study can identify policy targets that are important for the development of these skills and contribute to improving targeted outcomes.
- It is unclear which individual social and emotional skills (and their combinations) help students achieve success in school and later on in life. Most of the research in this area relates to the overarching Big Five dimensions, representing broad measures of individuals' characteristics, and thus making it difficult to translate into policy action. This Study provides an opportunity to examine the predictive values of individual social and emotional skills while controlling for the potential effects of a wide range of other skills. As a result, the Study sheds light on the incremental value of individual social and emotional skills in relation to the broad set of life outcomes.
- There is also relatively little understanding of how social and emotional skills develop throughout childhood and adolescence; research is needed to shed light on what practices might enhance positive social and emotional skills development, and under what conditions these practices should be implemented to best support students. The Study can address these issues by evaluating two critical age-groups of students as they transition from primary to lower and then to upper secondary education.

1.2. Objectives of the Study

1.2.1. Purpose and objectives of the Study on Social and Emotional Skills

The overall goal of the Study on Social and Emotional Skills is to assist cities and countries to better support the development of social and emotional skills of their students. The Study builds on the premise that a holistic approach, promoting both cognitive and non-cognitive development, is best suited to enable children to fulfil their full potential. Thus, as school systems usually focus on traditional academic knowledge and skills, the Study aims to expand the scope of education policies to include the domain of social and emotional skills, while remaining aligned with traditional academic domains and cognitive skills.

More specific objectives of the Study are to:

- Provide participating cities and countries with robust and reliable information on their students' levels of social and emotional skills.
- Provide insights on individual, family, peer and school characteristics that foster or hinder the development of these skills.
- Provide evidence of the predictive value of social and emotional skills for life outcomes in education, conduct, health and personal well-being.

- Demonstrate that valid, reliable and comparable datasets on social and emotional skills can be produced across diverse student populations and settings.

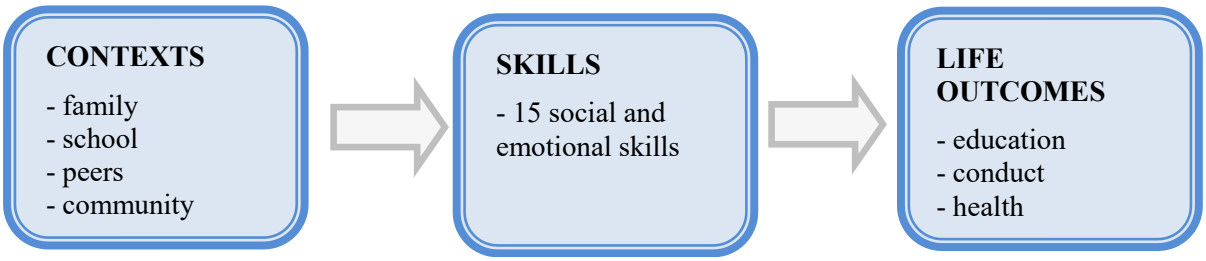
The Study can identify policies, practices and environmental conditions that are associated with the development of these critical skills. It can also provide an assessment tool for policy-makers and education practitioners to use to monitor the development of students' social and emotional skills and could measure the impacts of changes in policies or practices on these skills.

1.2.2. Analytical model and key research questions

The general analytical approach of the Study is based on a simple model aimed at accommodating complex interactions, while highlighting the key policy questions of interest.

General analytical model of the SSES:

Figure 1.2. Analytical model of the Study on Social and Emotional Skills



The study is specifically designed to treat multiplicity of skills, wide range of contextual factors and diverse set of life outcomes. In the most consise form, it has two broad topics of interest, i.e. two sets of structural relations, as indicated in Figure 1.2:

- Identification of contextual factors that promote or hinder skill development;

- Examination of relevance of different skills for various life outcomes.

First, we tackle the issue of drivers of skill development by investigating structural relationships between different aspects of students' home, school and peer environments and their social and emotional skills. Then we explore the wider benefits of social and emotional skills by incorporating a variety of life outcomes. Importantly, we shed light on the multi-dimensional nature of social and emotional skills by analysing a broad and comprehensive set of these skills and their structural relations with contextual factors and life outcomes.

In this way, the Study can help address numerous questions that are considered important for policy-makers, researchers, educators and parents.

These include:

- Which family learning contexts – such as parenting styles, quality of parent-child relationships and learning resources available at home – influence students' social and emotional development?

- Which school learning contexts – such as teachers' pedagogies, school climate and school practices – influence students' social and emotional development?
- Which social contexts – such as quality and quantity of relationships with friends and classmates – influence students' social and emotional development?
- Which community learning contexts – such as sport and cultural resources, community support or safety – influence students' social and emotional development?
- How do students' social and emotional skills differ across demographic groups, e.g. across students' gender, immigrant and socio-economic background?
- Which social and emotional skills influence students' learning and social outcomes, such as academic achievement, health, as well as their general well-being?
- What are the similarities and differences in levels of social and emotional skills, factors that influence them and their outcomes, between 10- and 15-year-old students?
- What are the similarities and differences in factors related to social and emotional skills across different cities and countries?

1.2.3. Policy relevance of the Study

Governments around the world recognise the importance of developing young people's social and emotional skills through education in order to prepare them for the future. National curricula often target skills that include autonomy, responsibility, tolerance, critical thinking and intercultural understanding (OECD, 2015[2]). These may be developed in multiple ways through:

- providing opportunities to learn about social and emotional skills
- teaching students how they can develop a fact-based and critical worldview
- equipping students with the means to analyse a broad range of cultural practices and meanings
- engaging students in experiences that facilitate intercultural relations
- promoting the value of diversity
- physical education in which students learn how to set goals, work towards improvement and work with others
- health education which generally aims to develop students' self-esteem and emotional stability
- civic education where objectives often include developing students' skills in conflict resolution as well as their capacity to think independently
- education in ethics which endeavours to instil values of fairness and respect for others, and promote skills such as self-control or willpower (Lapsley and Yeager, 2012[23]).

The ultimate aim of the Study is to provide an empirical basis for policies that cities and countries can follow in order to promote students' social and emotional skills development. It can provide valuable information that cities/countries can use to assess how developed

students' skills are, and what factors might be related to skill acquisition. This information can help policy-makers and educators discern if there are gaps or lags in skill development that could be addressed through curricula, through changes made to the schools' cultural context, or by encouraging schools to work more closely with parents.

As stated in *Skills for Social Progress: The Power of Social and Emotional Skills:*

"Although there may be no one-size-fits-all solution, given children's diverse social and cultural backgrounds, identifying and expanding promising strategies on a larger and wider scale could improve the effectiveness and efficiency of educational systems in raising social and emotional skills." (OECD, 2015, p. 109[2])

When considering the distribution of skills, the Study provides information on:

- How are social and emotional skills distributed among subgroups of the overall populations studied? For example, do they differ by gender, socio-economic status or cognitive ability?

When thinking about factors that might be related to social and emotional skill acquisition, the Study provides material to address questions, such as:

- How are the characteristics of the family environment related to children's social and emotional skills?
- In what ways are parents, teachers and schools invested in promoting the development of social and emotional skills among their children?
- How do official school policies regarding the importance of developing social and emotional skills in their students translate into the acquisition of those skills?
- What role does school climate (as measured by the tolerance of bullying, promotion of collaboration, etc.) play in the acquisition of particular social and emotional skills?
- How much do children's skills depend on the close alignment of parents' views with those of the school?

Participating cities and countries will be able to compare structural relations between skills and contextual factors in their jurisdiction to those in other participating jurisdictions. Likewise, participating cities and countries will be able to compare relevance of different social and emotional skills with varous life outcomes with the international average and other individual participants. Thus, the Study will enable policy-makers to answer questions, such as:

- How does observed distribution of social and emotional skills across gender, immigration or socio-economic categories in my city differ from that found in other international settings? What policy actions could be possible in order to reduce skill disparities across disadvantaged groups of students, based on observed international evidence?
- What family characteristics and activities are important drivers of social and emotional skills in other participating cities and countries? How do they differ from the ones found in my city?
- What aspects of school environment are found to be the most congruent for skill development across all sites? How do these findings compare to those found in my constituence?

- How do peer and community networks affect skill development across different sites? How does it compare with the situation in my city?
- Are student's life outcomes differently influenced by their social and emotional skills across the cities? What could be reasons for this? How to improve life outcomes of students using these insights from the international comparative perspective?

Thus, the international perspective allows policy-makers to go beyond observation of structural relations between skills and their antecedents and outcomes in their local context. It allows them to gain important insights about differences in these structural relations across the sites, international trends, as well as about possible reasons for such differences. In other words, apart from describing situations in their own context, they get the opportunity to observe the same situation in a number of other contexts and to compare how the same structural relations vary across these contexts. This allows for a more comprehensive and in-depth examination of interrelations between relevant aspects and consequently for better policy insights. Such insights can then be used for creation of more adequate and fine-tuned policy interventions.

1.3. Main features of the Study

The Study on Social and Emotional Skills is complex and ground-breaking, involving tens of thousands of students, parents and teachers from all around the world, and gathers information on a large set of personal and contextual factors. This section briefly outlines the Study's key aspects.

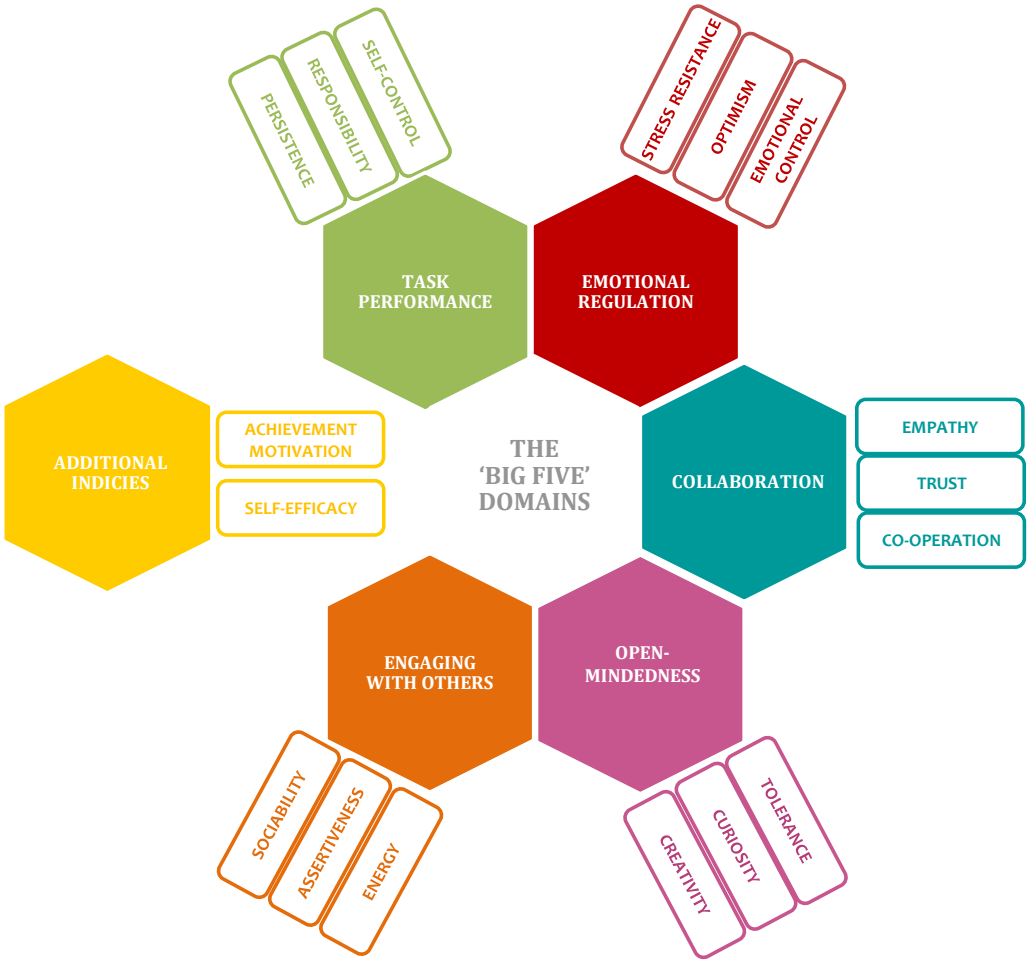
1.3.1. Social and emotional skills assessed in the Study

The Field Test collected information on 19 selected social and emotional skills of students. Following the Field Test, 15 social and emotional skills were selected for the assessment in the Main Study. The Study includes the following 15 social and emotional skills in the Big Five model domains (see Section 2):

- task performance: self-control; responsibility; persistence
- emotional regulation: stress resistance; emotional control; optimism
- engaging with others: energy; assertiveness; sociability
- open-mindedness: curiosity; creativity; tolerance
- collaboration: empathy; co-operation; trust

In addition, the Main Study will include two additional skills in form of indices that are calculated from a selection of items that belong to scales of related skills:

- Skills estimated from indices: self-efficacy; achievement motivation.

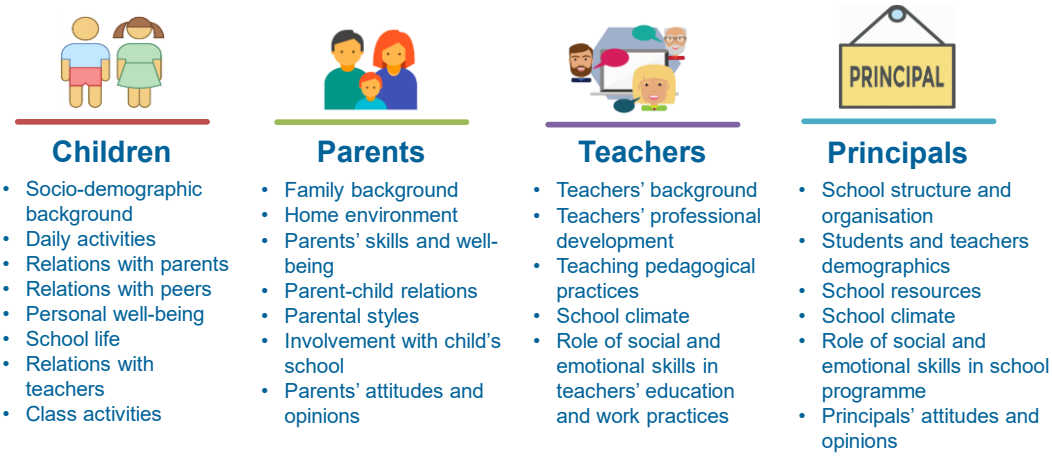


1.3.2. Contextual information collected in the Study

Apart from assessing students' social and emotional skills, the Study examines a wide scope of contextual factors (see Section 3). These can be divided into five broad groups:

- socio-demographic background of students
- family environment
- school environment
- peer environment
- wider community environment.

Collecting contextual information is critical in helping to understand how students' social and emotional skills have developed and how these skills may be improved. Students learn in many different settings, including in their families, schools and communities, with each context playing an important role throughout childhood and adolescence. Contextual information gives us a better understanding of what helps and what hinders social and emotional skills development, including the policies and practices that support them.



1.3.3. Participating cities

The following 10 cities from 9 countries are participating in the Study:

- Bogota, Colombia
- Daegu, South Korea
- Helsinki, Finland
- Houston, Texas, United States
- Istanbul, Turkey
- Manizales, Colombia
- Moscow, Russian Federation
- Ottawa, Ontario, Canada
- Sintra, Portugal
- Suzhou, People's Republic of China

1.3.4. Study respondents

The Study assesses social and emotional skills of two groups (cohorts) of students:

- 10-year-olds (younger cohort)
- 15-year-olds (older cohort)

In each of the participating cities, 3 000 students per cohort is randomly selected for participation in the study (thus, 6 000 students in total per participating city).

The study also collects data from three additional groups of respondents:

- parents of selected students
- teachers of selected students (those who know selected students best)
- school principals of selected students' schools.

1.3.5. Data collection instruments

The Study collects information through four questionnaires developed for students, parents, teachers and schools principals.

There are two main types of questionnaires in the Study:

  - Part A: Scales for assessment of students' social and emotional skills
  - Part B: Contextual questionnaires – used for collection of information about students' home, school and peer environment.

Students' social and emotional skills were assessed using three separate sources of reports:

- students' self-reports
- parents' reports on students
- teachers' reports on students.

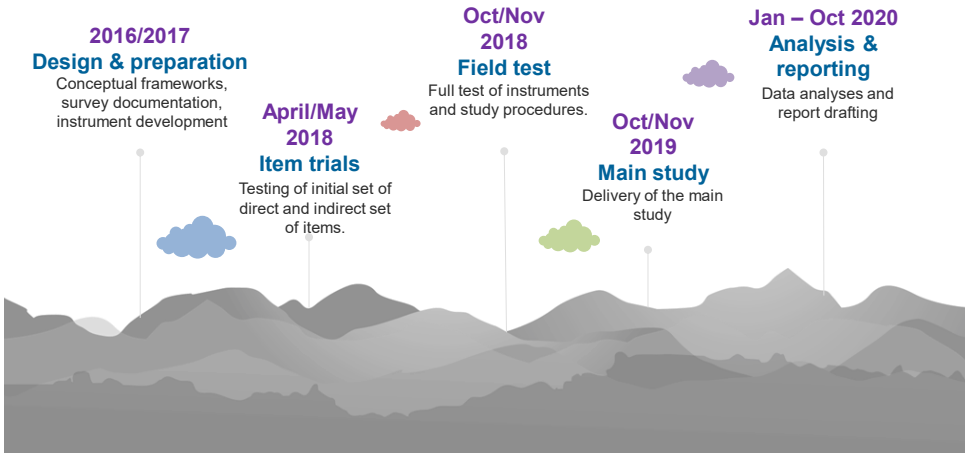
Figure 1.3. Structure of the Study' four questionnaires

STRUCTURE OF INSTRUMENTS		Students	Parents	Teachers	Principals
PART A: Contextual questionnaire		✓	✓	✓	✓
PART B: Assessment of students' skills	Self-reports	✓	✗	✗	✗
	Reports by others	✗	✓	✓	✗

- *Mode of instrument administration.* Student, teacher and principal questionnaires are administered online. Parent questionnaires are also primarily administered online, but parents in some sites are also offered a paper and pencil option.
- *Study administration.* Groups of students belonging to the same cohort sit the assessment. Administration is conducted in schools of the selected students.

1.3.6. Study timeline

Initial preparations for the study started in the end of 2016, with instrument development and survey preparation work being conducted throughout 2017 and 2018. The study is administered in October and November 2019. International report with the main findings will be published in September 2020.



2. Social and emotional skills

2.1. Overview

The Study's conceptual framework, published as a separate document, presents an overview of the literature encompassing social and emotional skills (Chernyshenko, Kankaraš and Drasgow, 2018[3]). It also describes the nature and structure of social and emotional skills, their development, malleability and factors that influence them, their cross-cultural comparability, and their relevance for a wide range of educational, economic and life outcomes. Moreover, it outlines the Big Five model, which is the basis for the Study**.** This section presents some of the most important aspects of the conceptual framework of the SSES.

The Study on Social and Emotional Skills (SSES) has developed its conceptual framework based on extensive literature reviews and incorporation of wide range of research streams in the area of social and emotional skills. In doing so, various existing social and emotional skill frameworks were examined and their differences and overlaps evaluated by leading experts in the field. Importantly, on top of the conceptual analyses, authors of the SSES framework have used empirical findings of the studies that were investigating empirical overlaps between measures belonging to various social and emotional skills frameworks (John and Mauskopf, 2015[24]; Primi, John and de Fruyt, 2016[25]). All of these conceptual and empirical analyses were taken into account during the development of the SSES framework (Chernyshenko, Kankaraš and Drasgow, 2018[3]). The final SSES framework is structurally and conceptually aligned with the Big Five model. At the same time, it is purposefully compiled in such a way that it incorporates concepts and empirical findings not only in the field of personality psychology but also in the fields of education, developmental psychology, 21st century skills, etc.

The Big Five model is a product of multiple research tracks. The large body of research, culminated over time, has led to a general consensus that the Big Five model sufficiently describe the basic dimensions of human personality. Research has also shown that adult personality traits can be organised in a hierarchical fashion where broad, higher-order characteristics can be split into narrower, lower order ones (Markon, 2009[26]), and the Big Five model represents a valuable foundation for this hierarchy (John, Naumann and Soto, 2008[27]). Moreover, the Big Five model can be replicated and compared across countries and cultures (McCrae and Terracciano, 2005[28]).

In this section, we begin by describing the Big Five model and its relevance to children's and adolescents' development. We then identify sets of key sub-domains for each of the Big Five domains. This is followed by an account of the main criteria when choosing sub-domains of the Big Five to include in the Study. Next, we briefly discuss each of the five broad higher order domains, along with the sub-domains selected by the OECD for the Item Trials and Field Test.

The Study focuses on a narrower set of social and emotional skills rather than on the broader Big Five dimensions since in many recent studies these narrower skills:

- have been found to have higher predictive validities than the broad domains (Ashton, 1998[29]; Mershon and Gorsuch, 1988[30]; Paunonen, 1998[31]; Paunonen and Ashton, 2001[32]; Roberts et al., 2005[33])

- provide descriptions that are more aligned with concrete behaviours and educational policy actions than the broad domains – especially when individuals score in the intermediate range on measures of broad dimensions as these scores do not reflect how well individuals do on sub-dimensions, and the number of variations can be quite large
- are more effective than the broad domains for teachers and parents trying to establish effective interventions to help develop social or emotional skills in a child.

Section 2 ends with a description of the three additional compound skills that are also essential to include in the Study.

2.2. The Big Five model

The conceptual framework put forward by the OECD draws on several existing theoretical structures. The most influential is the Big Five taxonomy that distinguishes five basic dimensions of personality (agreeableness, conscientiousness, extraversion, emotional stability, and openness to experience) and provides a parsimonious and efficient summary of social and emotional skills (John and De Fruyt, 2015[34]; Abrahams et al., 2019[35]; Lipnevich, Preckel and Roberts, 2017[36]).

The Big Five model developed from several strands of research (Digman, 1990[37]; Goldberg, 1982[38]; McCrae and Costa Jr, 1987[39]; Norman, 1963[40]; Tupes and Christal, 1958[41]). The model is developed with an objective of identifying general structure of main dimensions of human personality. Each dimension represents a cluster of related thoughts, feelings, and behaviours and can, therefore, be divided into narrower sub-domains. Agreeableness, for example, indicates a pro-social and communal orientation to others and includes sub-domains such as altruism and tender-mindedness. The Big Five model is applied in many countries around the world (McCrae and Terracciano, 2005[28]; McCrae and Costa Jr., 1997[42]). However, such a model with only a few concepts can only broadly cover the universe of social and emotional skills (Hampson, John and Goldberg, 1986[43]) and therefore, sub-dimensions of the Big Five may be more useful in some situations than the broader domains (Paunonen and Ashton, 2001[32]; Roberts et al., 2005[33]). Moreover, recent studies have found that measures of sub-domains have higher predictive validities than broad domains (Ashton, 1998[29]; Paunonen, 1998[31]; Roberts et al., 2005[33]). As measures of sub-dimensions are more descriptive, specific, and accurate, and more clearly point the way to potentially effective interventions, this Study follows John and De Fruyt's (2015[34]) suggestion to include measures of the Big Five *sub-domains*.

The five broad dimensions that refer to different sets of behaviours, thoughts and feelings are as follows:

Conscientiousness (in this study – task performance): Those who are conscientious, self-disciplined and persistent can stay on task, and tend to be high achievers, especially when it comes to education and work outcomes.

Emotional stability (in this study – emotional regulation): This encompasses skills that enable individuals to deal with negative emotional experiences and stressors. Being able to regulate one's emotions is essential for multiple life outcomes and seems to be an especially important predictor of enhanced mental and physical health.

Extraversion (in this study – engaging with others): People who score highly regarding extraversion are energetic, positive and assertive. Engaging with others is critical for

leadership and tends to lead to better employment outcomes. Extroverts also build social support networks more quickly, which is beneficial for mental health outcomes.

Agreeableness (in this study – collaboration): People who are open to collaboration can be sympathetic to others and express altruism. Agreeableness translates into better quality relationships, more pro-social behaviours and less behaviour issues.

Openness to experience (in this study – open-mindedness): Open-mindedness is also predictive of educational attainment, which has life-long positive benefits and seems to equip individuals better to deal with life changes.

2.2.1. The relevance of the Big Five model to school students

As the Big Five model originally stems from research on adults, the OECD Study on Social and Emotional Skills examines whether it can be applied to school-age children. The cumulative body of childhood research investigating this topic mostly concludes that the answer is yes (Caspi and Shiner, 2006[44]; De Fruyt and De Clercq, 2014[45]; Measelle et al., 2005[46]; Shiner, 1998[47]; Shiner and Caspi, 2003[48]; Tackett et al., 2008[49]; Tackett et al., 2012[50]). Three studies are particularly relevant to the Study. Each one involves inventories developed explicitly for school-aged children using a "bottom-up strategy" in which the full range of social and emotional descriptors observed in target age groups is collected and then reduced to a subset of items applicable across multiple ages and cultures:

- Mervield and DeFruyt (1999[51]) developed the Hierarchical Personality Inventory for Children (HiPIC), by analysing over 3 000 personality descriptors found in the Flemish language and reducing these to 144 items, organised into 18 facets, representing the most common personality descriptions of 6-12-year-olds. The items were then tested on three age groups of Belgian school children. These facets could be grouped into five broad characteristics which were highly replicable when tested across the three age groups and related well to the adult Big Five.
- Halverson et al. (2003[52]) describes how the Inventory of Child Individual Differences (ICID) was developed from over 50 000 country and language-specific parental descriptors of children ages 3-12-years, which were then reduced to 141 items that were common descriptors across cultures and measured 15 narrow personality characteristics. Employing confirmatory factor analysis to 1 035 parent ratings on four samples of children ages 3-13 from the People's Republic of China, Greece and the United States, Halverson et al. again matched characteristics to the Big Five model.
- Tackett et al. (2012[50]) analysed parent ratings of children and early adolescents from five countries on 108 ICID items. Among children ages 9-11 and 12-14 (i.e. close to the ages of children the Study targets), three of the Big Five dimensions were consistently replicated across the two age groups. The other two were less stable, although this may, in part, have been due to difficulties involved in measuring internal aspects of children's personality relying solely on parents' reports.

These studies illustrate that youths' social and emotional skills (as is the case with adults) can be organised hierarchically – meaning that each broad dimension can be progressively divided into narrower sub-domains – and that the Big Five model can be reliably measured in childhood and adolescence (Soto et al., 2008[53]; Tackett et al., 2012[50]). The model's underlying sub-domains can also be reliably rated by parents and teachers (Halverson et al., 2003[52]; Mervielde and De Fruyt, 1999[51]), which is important for the Study as it

incorporates parent and teacher reports of youths' social and emotional development in addition to reports from the youths themselves. For a recent overview of different social and emotional skills' frameworks interested readers can consult Abrahams et al. (2019[54]) paper.

Evidence from three initial empirical phases of the Study (cognitive interviews, Item Trials and Field Test) also clearly show that selected social and emotional skills are both distinct from one another and measurable even in case of students as young as 10. In particular, psychometric properties of scales used for assessment students' social and emotional skills were shown to be robust not only in case of older cohort but also in case of students from the younger cohort. In addition, mutual relationships between the assessment scales were found to largely correspond to those observed among the adult population and the theoretical expectations based on the Big Five model. More details on psychometric properties of our assessment scales in the Field Test are available in our recent publication (Kankaraš, Feron and Renbarger, 2019[55])

2.2.2. The relevance of the Big Five model for educational and social policies

There is a large body of empirical evidence indicating relevance of the social and emotional skills from the Big Five domains for important life outcomes and achievement (Roberts et al., 2007[11]; Gutman and Schoon, 2013[56]; Heckman and Kautz, 2012[7]; Kautz et al., 2014[6]). Likewise, the social and emotional skills that belong to the Big Five model are found to be malleable and suspectable to the formal and informal interventions (Helson et al., 2002[57]; Roberts, Walton and Viechtbauer, 2006[58]; Specht, Schmukle and Egloff, 2011[59]; Srivastava et al., 2003[60]). In this section, we present some research findings that pertain to the relevance of the Big Five model for the educational and social policies. More information about the predictive validity as well as malleability of specific Big Five sub-domains is presented in Sections 2.4.1 and 2.4.2 where we focus on those sub-domains chosen to be included in the the Study.

Research results underscore that social and emotional skills, especially those belonging to the domain of Consientiousness, play an important role in educational attainment (Almlund et al., 2011[61]; Heckman, Stixrud and Urzua, 2006[19]). Other studies have also found that the social and emotional skills belonging to the Big Five domain Conscientiousness, such as self-control, persistence and responsibility, are significantly related with students academic achievement, even after controlling for cognitive skills (Heckman and Kautz, 2012[7]; Noftle and Robins, 2007[62]; Rosander and Backstrom, 2014[63]). In fact, these skills are found in several studies to be even stronger predictor of school grades than measures of cognitive skills. Another Big Five domain, emotional regulation, is also a significant predictor of academic achievement. For example, Lounsbury et al. (2004[64]) found that the broad dimension of being able to regulate emotions consistently predicted school absences among 7th, 10th and 12th graders. Open-mindedness and collaboration have also been found to be related to grades although these domains do not appear to have the same predictive magnitude as task performance.

Similarely as in case of their relations with education achievement, numerous empirical studies have shown that social and emotional skills in various Big Five domains are also related to employment outcomes. Skill belonging to the domain of Conscientiousness are found to predict job performance and income across a broad range of occupational categories (see, for example, (Sackett and Walmsley, 2014[65])). The Big Five dimension of extraversion (engaging with others) has also been found to predict income levels and educational attainment. Leadership is another outcome highly related to the Big Five

domain of extraversion that was shown to be important skill in a number of occupations (Bono and Judge, 2004[66]; Judge et al., 2002[67]). Skill belonging to openness to experience domain are shown to better equip individuals to deal with change, which is a capacity that may well have increasing relevance in the future world of work. Using data from the 1997 United States National Longitudinal Survey of Youth, Judge et al. (2012[68]) showed agreeableness domain to be significantly correlated with continuously working and with income.

Social and emotional skills belonging to different Big Five domains are shown to be related to person's physical and mental health and various health-related behaviours (Strickhouser, Zell and Krizan, 2017[69]). For example, skills in the conscientiousness domain are related to a range of health behaviours including safe driving, healthy eating, and substance use, as well as physical and mental health, (Bogg and Roberts, 2004[70]) Moffitt and colleagues (2011[71]). In particular, responsibility inversely relates to drug use, suicide, and violence; and self-control inversely relates to excessive use of alcohol, drug use, risky driving, tobacco use, and violence. Lower levels of skills from this conscientiousness and emotional regulation domains are also found to be related with anti-social, aggressive, and rule-breaking behaviours (Tackett, 2006[72]). Life satisfaction and personal well-being are also strongly affected by the social and emotional skills, especially by those belonging to the emotional regulation domain of the Big Five (Tauber, Wahl and Schroder, 2016[73]) and mental health (Strickhouser, Zell and Krizan, 2017[69]).

Large body of empirical research also shows that the skills belonging to the Big Five model are not fixed at birth with little or no room for improvement, but are instead suspectible to the influence of person's environment and can be changed throughout the lifetime (Helson et al., 2002[57]; Srivastava et al., 2003[60]). For example, it is found that skills belonging to the Big Five domains of conscientiousness, agreeableness and emotional stability, after a period of instability during the adolescence, generally increase with age (Roberts, Walton and Viechtbauer, 2006[58]). Significant life events such as marriage and getting first job can also have a substantially alter personality characteristics such as responsibility or co-operation (Roberts, Walton and Viechtbauer, 2006[58]; Specht, Schmukle and Egloff, 2011[59]). For example, a longitudinal study that assessed the Big Five skills at two different periods over four years, showed that mean levels of agreeableness, conscientiousness and openness changed significantly for men and women who had lost their jobs between the two periods. On ther other hand, changes were limited for those who were still employed or reemployed (Boyce et al., 2015[74]). Empirical research also shows that fostering conscientiousness early on may be an effective way to reduce unemployment throughout adulthood (Egan et al., 2017[75]).

Although early learning interventions are especially effective for the development of skills in the Big Five domain, these skills are malleable even in the very late stage of life, sometimes even more so than cognitive skills (Cunha and Heckman, 2007[12]; Cunha, Heckman and Schennach, 2010[13]). Recent studies on the effectiveness of training interventions also indicate that substantial changes in the Big Five personality characteristics are possible, even after relatively short treatment periods (Roberts et al., 2017[76]). For example, a mindfulness intervention was associated with changes in conscientiousness, agreeableness, empathy and emotional stability among medical residents (Krasner et al., 2009[77]). Similarly, a social-skill training programme for recovering substance abusers led to increases in agreeableness, conscientiousness and emotional stability (Piedmont, 2001[78]). Likewise, a cognitive training intervention for older adults was also associated with increased levels of respondents openness experience (Jackson et al., 2012[79]).

2.3. Identifying key social and emotional skills to include in the Study

Since social and emotional skills are arranged hierarchically, the five general domains of the Big Five framework, can be split into narrower, lower-order sub-domains. Each of these broad domains encompasses a cluster of mutually related narrower social and emotional skills. Task performance, for example, includes achievement motivation, self-control/self-discipline, responsibility/trustworthiness and persistence. These groupings demonstrate mutual similarity of skills belonging to the same domain but also ensure systematic, comprehensive and balanced consideration of individuals' social and emotional skills (Kankaraš, 2017[5]; OECD, 2015[2]). The individual social and emotional skills that we want to assess in the Study can, therefore, be viewed as more contextualised manifestations of broad domains (Roberts, 2006[80]). However, deciding which social and emotional skills to study is not a trivial matter as *"[t]here could be hundreds, if not thousands, of different ways to group typical patterns of behaviours, thoughts and feelings"* (Chernyshenko, Kankaraš and Drasgow, 2018, p. 62[3]).

Although the Big Five model is widely accepted, there is insufficient research – and therefore some disagreement – on the structure of sub-domains that underlie each domain (John, Naumann and Soto, 2008[27]). Different methods have been used in the past to create a set of sub-domains to cover the various dimensions of the broader domains resulting in a somewhat different lower-order structure proposed for every existing personality measure. Many early temperament models were rationally derived from interviews (Thomas and Chess, 1977[81]). And narrow facet taxonomies were established through a purely empirical method of analysing responses to a diverse array of personality indicators by conducting a series of factor analyses (Saucier and Ostendorf, 1999[82]; Ashton et al., 2004[83]; DeYoung, Quilty and Peterson, 2007[84]).

Taxonomies from both rationally derived and empirically based approaches were combined to identify key sub-domains for the Study on Social and Emotional Skills. Sub-domains that were consistently identified and cross-culturally replicated were considered for possible inclusion. This ensures not only that a sub-domain belongs to a particular domain, but also maximises being able to apply the Study's findings to existing personality frameworks. In order to represent diverse viewpoints of the lower-order structure of the Big Five domains of both adults and children, we drew on seven sub-domain-level taxonomies:

1. The Thomas and Chess (1977[81]) temperament model is comprised of nine basic temperament characteristics. As assessment instruments were developed to target infants, pre-, primary and secondary school-age children, the model is useful for the Study (Thomas and Chess, 1977[81]; Hegvik, McDevitt and Carey, 1982[85]).
2. The 18 sub-domains of Mervield and De Fruyt's (1999[51]) Hierarchical Personality Inventory for children (HiPIC) (see also (Mervielde, De Fruyt and De Clercq, 2009[86])) applies to the Study because all HiPIC items are written either in the first person so that children can respond or in the third person singular for parents or other caregivers. This factor structure has also proven to be highly replicable across both childhood and adolescence with broad domains that closely align with those of the Big Five.

3. The 15 personality sub-domains of the Inventory of Children's Individual Differences (ICID) (Halverson et al., 2003[52]) relates to the Study because 50 000 country and language-specific parental descriptors of children ages 3-12 contributed to the reduced common set of 141 items free of cultural bias.
4. The 15 sub-domain structure of the Next Big Five Inventory (BFI-2) (Soto and John, 2017[87]) is rationally designed and provides *"continuity with the original [Big Five model] and previous research on personality structure"* (Soto and John, 2017, p. 121[87]).
5. An empirically derived 18-facet taxonomy consisting of the Big Five broad factors, each one comprising 3-4 narrow facets is based on the widely cited lexical study by Saucier and Ostendorf (1999[82]) who factor-analysed responses to German and English language personality adjectives.
6. The 24 sub-domain taxonomy from the HEXACO personality inventory (Lee and Ashton, 2004[88]) is based on eight independent investigations involving seven different languages and applies to the Study because the developed items were comparable across cultures.
7. The 21 empirically derived sub-domains implemented in the Tailored Adaptive Personality Assessment System (TAPAS) (Drasgow et al., 2012[89]) were drawn from seven major adult personality inventories: the revised NEO Personality Inventory (NEO-PI-R) (Costa, McCrae and Dye, 1991[90]), the Sixteen Personality Factor Questionnaire (16PF) (Conn and Reike, 1994[91]), California Personality Inventory (CPI) (Gough, 1987[92]), the Multidimensional Personality Questionnaire (MPQ) (Tellegen, 1982[93]), the Jackson Personality Inventory – Revised (JPI-R) (Jackson, 1994[94]), the Hogan Personality Inventory (HPI) (Hogan and Hogan, 1992[95]), and the Abridged Big Five-Dimensional Circumplex (AB5C) scales from the International Personality Item Pool (Goldberg, 1999[96]). They have all been widely researched, translated into multiple languages and shown to be cross-culturally relevant.

These seven taxonomies were examined for conceptual overlaps, similarities and differences and then served as a starting point of the development of a common framework underlining the OECD's Study on Social and Emotional Skills (Chernyshenko, Kankaraš and Drasgow, 2018[3]). Also added were sub-domain scales from several well-known adult personality inventories such as AB5C, NEO-PI, 16PF and the Occupational Personality Questionnaire (OPQ) (Saville et al., 1984[97]).

2.4. Criteria for the selection of social and emotional skills

Based on the outlined set of reference taxonomies, a broad scope of sub-domains has been put forward to account for the Big Five conceptual space comprehensively. The Study, however, could not include measures that tap into all of these skills without burdening those who are asked to provide information on students' skills: students, parents and teachers. In order to provide a broad, balanced set of social and emotional skills with sufficient depth and meaning, all parties involved in the Study design agreed that the final set of selected skills should:

- include 2-3 sub-domains within each of the broad Big Five domains, attempting to be as comprehensive in this selection as possible.

Further, selected social and emotional skills should also:

- have *predictive value* across four categories of essential life outcomes and events: educational attainment; economic outcomes; health; and quality of life
- be *malleable* and susceptible to interventions and policy measures, especially during the early years of an individual's life
- be *appropriate* for children and adolescents at ages 10 and 15
- result in scores that are *comparable* across cultures and nations, taking into account sensitivity to cross-cultural issues and the fact that the manner in which some sub-domains are expressed behaviourally may differ from one culture to another
- be *relevant* for the world in the future rather than only relevant now
- be *well researched* and having large body of accumulated empirical evidence.

2.4.1. Predictive value of social and emotional skills

Numerous studies and meta-analyses have found the Big Five dimensions to be associated with academic achievement, health and well-being, job performance and occupational attainment (Roberts et al., 2007[11]; Gutman and Schoon, 2013[56]; Heckman and Kautz, 2012[7]; Kautz et al., 2014[6]), and in some cases the predictive value of the Big Five dimensions rivals that of long-established measures of cognitive skills. Further, personality characteristics influence life outcomes both directly (for example, being socially intelligent can help a person successfully negotiate a job interview), and indirectly (for example, being curious, open-minded and possessing an active approach towards learning are important skills for developing and improving innate cognitive capacities) (Kankaraš, 2017[5]).

In this section, we present some research findings that pertain to the predictive value of the larger domains of task performance, emotional regulation, engaging with others, collaboration, and open-mindedness. More information about the predictive validity of specific sub-domains is presented in Sections 2.5.1 through 2.5.5 where we focus on those sub-domains chosen to be included in the Item Trials for the Study.

Educational attainment is considered one of the most critical outcomes for youth, and the impact of cognitive skills on a whole range of educational attainment measures is well known. Research results also underscore that social and emotional skills, such as task performance, also play an important role in educational attainment (Almlund et al., 2011[61]; Heckman, Stixrud and Urzua, 2006[19]). And Heckman and Kautz's (2012[7]) findings regarding task performance are especially noteworthy because they also controlled for cognitive skills in their analyses. In a large, multi-sample study of University of California students, Noftle and Robbins (2007[62]) also found that task performance (or conscientiousness) was a consistent predictor of grades, even after controlling for gender and IQ, and in several cases was a stronger predictor of grades than verbal and math SAT scores. Rosander and Backstrom's longitudinal study of 197 Swedish high school students also found conscientiousness scores correlated with academic grades 3 years later, and that this relationship remained even after controlling for cognitive ability scores (Rosander and Backstrom, 2014[63]). Emotional regulation is also a significant predictor of academic achievement. For example, Lounsbury et al. (2004[64]) found that the broad dimension of being able to regulate emotions consistently predicted school absences among 7th, 10th and 12th graders. Open-mindedness and collaboration have also been found to be related

to grades although these domains do not appear to have the same predictive magnitude as task performance.

Higher levels of education and better grades typically translate into lower chances of unemployment and higher levels of income (OECD, 2015[2]). Although cognitive skills, such as IQ, have long been considered the most important determinants of employment success, multiple studies have shown that social and emotional skills are also related to employment outcomes. Task performance appears to predict performance and wages across a broad range of occupational categories [see, for example, (Sackett and Walmsley, 2014[65])]. The overall dimension of extraversion (engaging with others) has also been shown to predict income levels and educational attainment. Leadership is another outcome highly related to extraversion (Bono and Judge, 2004[66]; Judge et al., 2002[67]). Being able to engage with others is critical for successful leadership and therefore tends to lead to better employment outcomes. Being open-minded is a dimension that appears to better equip individuals to deal with change, and this is a skill that may well have increasing relevance in the future world of work. Using data from the 1997 United States National Longitudinal Survey of Youth, Judge et al. (2012[68]) showed agreeableness to be significantly correlated with continuously working and with income.

Some domains, such as task performance, have been found to have predictive validity across a wide range of outcomes: in addition to educational attainment and occupational success, task performance has also been found to predict a range of health behaviours including safe driving, healthy eating, and substance use (Bogg and Roberts, 2004[70]) as well as physical and mental health (Strickhouser, Zell and Krizan, 2017[69]), and overall life satisfaction (OECD, 2015[2]). Children with low task performance scores have also been found to exhibit higher rates of anti-social, aggressive and rule-breaking behaviours (Tackett, 2006[72]). Other domains, such as emotional stability, predict some types of outcomes more than others, such as life satisfaction among adults (Tauber, Wahl and Schroder, 2016[73]) and mental health (Strickhouser, Zell and Krizan, 2017[69]). Although as noted by Schoon and colleagues (2015[98]), when it comes to predicting later economic outcomes, the evidence for the predictive capacity of emotional stability is mixed. However, social and emotional skills can *indirectly* affect life outcomes. For example, emotional stability does not directly influence employment and income but can influence various other factors, such as educational achievement, health-related behaviours or quality of social relationships, that in the long run affect employment status, overall health and personal well-being (Kankaraš, 2017[5]).

2.4.2. The malleability of social and emotional skills

Examining whether social and emotional skills are malleable introduces the possibility of changing or developing them for the better. Extensive research shows that children are not born with a fixed set of skills and little room for improvement, but instead have considerable potential to develop social and emotional skills which are influenced throughout life by their environment (Helson et al., 2002[57]; Srivastava et al., 2003[60]). For example, levels of conscientiousness, agreeableness and emotional stability generally increase with age (Roberts, Walton and Viechtbauer, 2006[58]). Significant life events such as marriage and one's first job can also have a substantial influence on personality characteristics (Roberts, Walton and Viechtbauer, 2006[58]; Specht, Schmukle and Egloff, 2011[59]), and conscientiousness has been found to increase in individuals when they start their first job and decrease when they retire (Specht, Schmukle and Egloff, 2011[59]).

While early learning interventions are especially effective for the development of all skills, social and emotional skills are more malleable at later stages in life than cognitive skills (Cunha and Heckman, 2007[12]; Cunha, Heckman and Schennach, 2010[13]). Recent studies on the effectiveness of training interventions also indicate that substantial changes in personality characteristics are possible, even after relatively short treatment periods. Although at the individual level, personality becomes increasingly stable across adulthood, between the ages of 6 and 18, personality can change substantially (Roberts and DelVecchio, 2000[99]). This is especially the case during adolescence (Soto, 2016[100]; Soto and Tackett, 2015[101]) when young people are increasingly influenced by their peers (Grusec and Davidov, 2010[102]), and need to develop skills such as negotiation, conflict resolution, empathy and understanding (Kerr et al., 2003[103]). A multitude of physical, hormonal and psychosocial changes also take place during this period as adolescents grapple with more autonomy and responsibility. One has only to live with a teenager to know that adolescence is a learning phase characterised by turbulent changes in social and emotional skills, such as sociability and regulating emotions (Soto et al., 2011[104]). So long as parents survive this stage, they are rewarded with a young adult who goes back to being pleasant, more sociable, less irritable and possesses an enhanced ability to take on more responsibility. From childhood throughout adolescence, there are therefore many opportunities for parents, teachers and schools to provide learning environments where skills can be developed, enhanced and reinforced through practice and daily experiences.

Investing in the development of social and emotional skills among students is not only crucial for outcomes such as better employment, or reducing the likelihood of anti-social behaviours but also in developing cognitive skills (Chernyshenko, Kankaraš and Drasgow, 2018[3]). The same does not apply to cognitive skills as they have limited impact on future social and emotional skill development (Cunha and Heckmann, 2008[105]; Cunha, Heckman and Schennach, 2010[13]; OECD, 2013[106]).

There have been fewer observational studies on social and emotional development in children and adolescents than in adults. In part, because age-appropriate assessment instruments for adults were developed before they were for children. Furthermore, much of the research on adult and youth populations only covers the broad domains of the Big Five. For example, conscientiousness is one of the Big Five domains found to change during the early years as it appears to decline from late childhood into early adolescence, and then develops rapidly from late adolescence into early adulthood. Soto's and John's (2014[107]) study is a recent exception to only researching the main domains; their findings suggest that developmental patterns do not always hold true for all the sub-domains of a particular Big Five domain.

As the OECD's Study on Social and Emotional Skills focuses on sub-domains, and develops assessment instruments explicitly for two different age groups of children, it has potential to add to our understanding of social and emotional skill development, and provide insights into the sorts of schools, classrooms and family environments that might promote the development of those skills.

2.4.3. Cross-cultural comparability of social and emotional skills

Another issue that is relevant for the Study on Social and Emotional Skills is its cross-cultural comparability. The Study is administered in a variety of cities and countries around the world, and the same questionnaires are used throughout all the sites. However, if the same set of questions has different meanings for people from different cultural backgrounds, then these questions may measure somewhat different constructs in each

culture (Kankaraš and Moors, 2010[108]). Therefore, the study needs to ensure cross-cultural comparability, i.e. that constructs measured in different geographical locations yield the same attributes (Horn and McArdle, 1992[109]).

There is extensive evidence that the Big Five dimensions and their sub-domains are conceptually comparable across cultures; countries and economies (Paunonen et al., 1996[110]; McCrae and Costa Jr., 1997[42]) for both adults (Schmitt et al., 2007[111]; McCrae and Terracciano, 2005[28]) and children from different cultural backgrounds (Tackett et al., 2012[50]). However, there is also evidence that a simple comparison of scale scores across cultures (for example, computed by adding or averaging responses to Likert rated scale items) may not work due to possible method bias resulting from cultural differences in the interpretation of questions. A typical Likert item uses five response options ranging from "strongly agree" to "strongly disagree". Individuals from Western cultures have been found to exhibit an "extreme response" style where they tend to choose response options at the extreme ends of the scale regardless of the meaning of the question. In contrast, individuals from Eastern cultures tend to exhibit a "central response" style by choosing options from the middle of the scale (Kankaraš and Moors, 2011[112]). Another problem is that there may not be a one-to-one correspondence between words in different languages.

Instrument bias stemming from social desirability, response styles and the meaning of particular words receive particular scrutiny in the Study. In order to minimise the bias of cross-cultural comparability as much as possible, the OECD has worked with leading experts in the field to collectively develop comprehensive methodological, translation and statistical procedures that minimise the possibility of method biases. These procedures are elaborated in more detail in Section 4.

2.4.4. The relevance of social and emotional skills for the future

Social and emotional skills sub-domains are relevant for today's context and tomorrow's as well. The ability to think critically and act independently is becoming increasingly important due to the exponential increase in media platforms, overwhelming us with information that needs to be sifted through and evaluated in order to discern where the kernels of truth lie. Children who turn into life-long learners are better prepared to adjust to change. Decreasing levels of social and institutional trust, and a lessening ability to rely on traditional social networks as a result of increased migration place additional emphasis on people's sense of trust and their ability to collaborate with and have compassion for others. Jobs that are increasingly being automatised across a variety of industries forces society to be innovative and rethink the way we work, requiring individuals to understand others' emotions and social situations.

Fostering social and emotional skills is also connected with educational policy priorities, such as excellence and equity. Excellence in education without equity can lead to substantial economic and social disparities, while equity in education at the expense of quality can lead to stagnation. The most advanced education systems now set ambitious goals for all students, focusing on both excellence and equity. They also equip their teachers with the pedagogical skills that have proven effective, such as flexibility, collaboration and appreciation for diversity and meta-cognition, and with enough autonomy so that teachers can use their own creativity in determining the content and instruction they provide to their students (Guerriero, 2017[113]). Equity in education means that personal or social circumstances, such as gender, ethnic origin or family background, but also social and emotional skills, are not obstacles to achieving educational potential (fairness), and that all individuals reach at least a basic minimum level of skills (inclusion). The promotion of

excellence, equity and inclusion are critical aims for education. Empirical research shows that social and emotional skills have the potential to compensate for the effects of socio-economic disparities on academic performance (Steinmayr, Dinger and Spinath, 2012[114]; Suárez-Álvarez, Fernández-Alonso and Muñiz, 2014[115]; Tucker-Drob and Harden, 2012[116]). Equity policies aimed at mitigating the socio-economic conditions of students act as an empowering factor in interventions on social and emotional skills. By ensuring that disadvantaged students develop both cognitive skills and socio-emotional skills, schools and education systems can be at the forefront of creating more inclusive and fair societies.

2.4.5. Other selection criteria taken into consideration

Appropriate for ages 10 and 15

Each selected skill should be a distinct and measurable social and emotional skill already at age 10. Most of the sub-domains identified in this paper are likely to satisfy this criterion. However, some sub-domains are more appropriate for adults, such as introspection from the openness to experience domain, modesty from the agreeableness domain, and honesty/virtue from the conscientiousness domain.

Not burdensome to respondents as a whole

The Study should not burden students, parents, teachers and principals. This is an important consideration and the main reason why all 31 sub-domains identified across the seven taxonomies cannot be assessed.

Emirical evidence

Selected skills should be well-researched and either a critical sub-domain of one of the Big Five domains or skill of particular relevance outside of the Big Five. Sub-domains that are frequently identified across various taxonomies and inventories are the priority as they ensure the comprehensiveness of the selected set of skills.

Broad and balanced as a set

The Study assesses all six[2] social and emotional domains to ensure breadth and comprehensiveness. The selected skills need to provide information at individual sub-domain levels in order to produce more meaningful and actionable results. Each of the broad Big Five domains are represented with 2-3 sub-domains. Selected skills were considered as a whole set rather than individually when deciding which ones to include in the Study. Such an approach ensures selecting a balanced set of skills that do not overlap with each other, thus minimising redundancies and maximising the analytical value of the obtained information.

2.5. Selected skills to include in the Study on Social and Emotional Skills

This section focuses on the 15 sub-domains chosen to represent each of the Big Five domains and additional compound skills, based on the criteria already outlined. The section begins by outlining each of the broad domains and then follows with a description of each

[2] The Study includes the Big Five domains – task performance, emotional regulation, collaboration, open-mindedness and engaging with others – and compound skill: self-efficacy.

of the sub-domains, including information on their predictive value, malleability and relevance for children and adolescents.

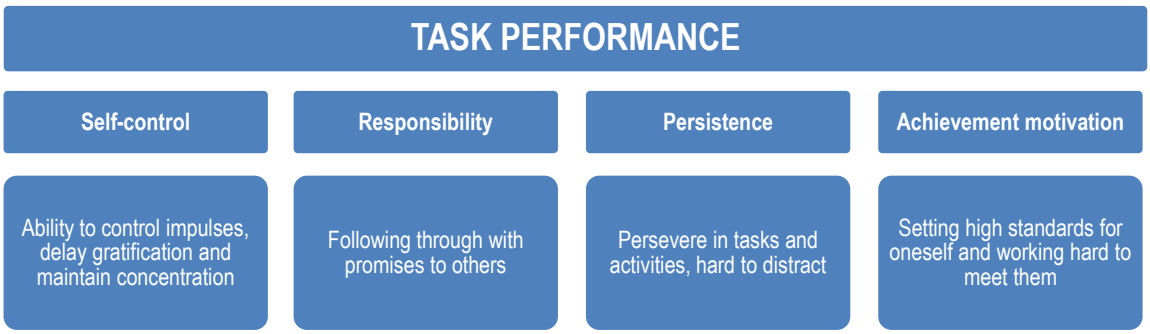
2.5.1. Task performance – getting things done, as required and on time

Known in the Big Five as conscientiousness, task performance includes a range of constructs that describe the propensity to be self-controlled, responsible to others, hardworking, motivated to achieve, honest, orderly, persistent and rule-abiding (Roberts et al., 2009[117]). Since the 1990s, hundreds of research papers have been published on the nature and usefulness of conscientiousness sub-domains, and several papers have exclusively focused on understanding the underlying lower-order structure of this critical domain (Roberts et al., 2005[33]; Roberts, Lejuez and Krueger, 2014[118]).

The Study includes the following sub-domains of task performance:

- self-control/self-discipline
- responsibility/trustworthiness
- persistence
- achievement motivation (index)

Figure 2.1. Task performance sub-domains



Persistence can be defined as an ability to persevere in tasks and activities in spite of challenges and distractions. Whereas in adult inventories, being persistent in the face of challenges is synonymous with industriousness and hence included under achievement motivation, in child-based taxonomies (e.g. HiPIC), industriousness is separated into two sub-domains: achievement motivation and *persistence*. The Study on Social and Emotional Skills also makes this distinction. *Self-control* represents the propensity to control impulses, delay gratification, and maintain concentration. An example item from the HiPIC is *"works with sustained attention"*. *Responsibility* reflects the tendency to follow through with promises to others. Adjectives used to describe this sub-domain are reliable, dependable, prompt, and punctual vs. undependable and unreliable. *Achievement motivation* implies working hard to meet the high standards set for oneself, putting in consistent effort, being highly productive and aspiring to excellence. Positive adjectives identified in the Saucier and Ostendorf (1999[82]) taxonomy belonging to this sub-domain include ambitious, industrious and purposeful, while negative adjectives include aimless, negligent and lazy.

Bogg and Roberts (2004[70]) conducted a meta-analysis investigating relationships between various sub-dimensions of task performance and health-related behaviours. They found that

responsibility inversely relates to drug use, suicide, and violence; and self-control inversely relates to excessive use of alcohol, drug use, risky driving, tobacco use, and violence.

The sub-dimension most often researched in early childhood studies is self-control. In a series of studies examining self-control among a sample of US pre-schoolers from a university community, delay of gratification at age 4 was associated with higher levels of cognitive and self-regulatory competence and coping at age 16, including higher scores on the standardised college entrance exams (SAT) (Shoda, Mischel and Peake, 1990[119]). In another study, 10-year-olds in a US study who exhibited high levels of self-control were shown to have higher academic attainment four years later (Duckworth, Tsukayama and May, 2010[120]).

Using data from the New Zealand Dunedin cohort, Moffitt and colleagues (2011[71]) linked to lack of self-control in childhood to lower-income, low socio-economic status and more self-reported financial difficulties at age 32. While a heightened sense of self-control was associated with greater physical health later on in life (e.g. absence of metabolic abnormality, periodontal disease, airflow limitation, etc.). Children who exhibited greater self-control were also less likely to abuse substances as an adult, including tobacco, alcohol, cannabis, street or prescription drugs. These associations were independent of factors such as intelligence and socio-economic status.

A study using another New Zealand cohort (the Christchurch cohort) found that a self-control score at age 6 was related to a range of adult outcomes including violent offending, welfare dependence, educational attainment and income (controlling for socio-economic status, child conduct disorders, IQ and gender; (Fergusson, Boden and Horwood, 2013[121])). Evidence from the UK National Child Development Study (NCDS) also suggests that childhood self-control mitigates unemployment throughout adulthood (Daly et al., 2015[122]). Furthermore, criminologists often point to a penchant for immediate gratification as an essential factor in leading some youths to engage in anti-social behaviour rather than desisting, knowing negative consequences would follow later.

Prior research has also examined the consistency of various sub-domains of task performance. For example, Soto and colleagues (2011[104]) used a large cross-sectional sample of over a million research participants between 10- and 65-years-old and found self-discipline declined in average levels between ages 10 and the early teens where it bottomed out. Throughout the later teenage years self-discipline sharply increased and then gradually increased from age 20 onwards. More recently, de Haan et al. (2017[123]) utilised longitudinal data from two independent Flemish samples. Focusing on ages 10 and 15, the same ages as in the Study, achievement motivation and self-control decreased from one age group to the other. On the other hand, persistence only decreased for boys and not for girls.

Research has shown that social and emotional skills are malleable between ages 6 and 18 – a range that encompasses the two age groups included in the Study. However, the Study must look at whether systematic interventions can positively influence skills development in children. A meta-analysis of 213 school-based social and emotional learning programmes that involved more than 270 000 primary and secondary school children conducted by Durlak and colleagues (2011[124]) showed that planned and systematic interventions aimed at limiting undesirable behaviours and increasing positive ones can, indeed, be successful. Furthermore, these social and emotional skills also translate into better academic, economic and life outcomes for children who benefited from these social and emotional learning programmes throughout childhood.

Achievement motivation and responsibility/trustworthiness both have theoretical significance and good predictive validity: achievement motivation is especially predictive of quality of life, and responsibility/trustworthiness of economic outcomes. Both are relevant for school settings and have shown to exhibit good cross-cultural comparability. Both are also malleable and therefore good candidates for the Study. A recent study on bullying in Korea, a major concern in many OECD countries, suggests that self-reported engagement in bullying among 14-year-olds is strongly driven by students' lack of responsibility towards others (Sarzosa and Urzua, 2015[125]). Self-control/self-discipline has garnered attention from researchers in numerous fields with empirical literature pointing to its strong relevance for children, theoretical importance and predictive validity, although this sub-domain is slightly less predictive of future socio-economic and health outcomes than are achievement motivation and responsibility/trustworthiness. Persistence appears to predict educational attainment and is considered a highly relevant skill for children.

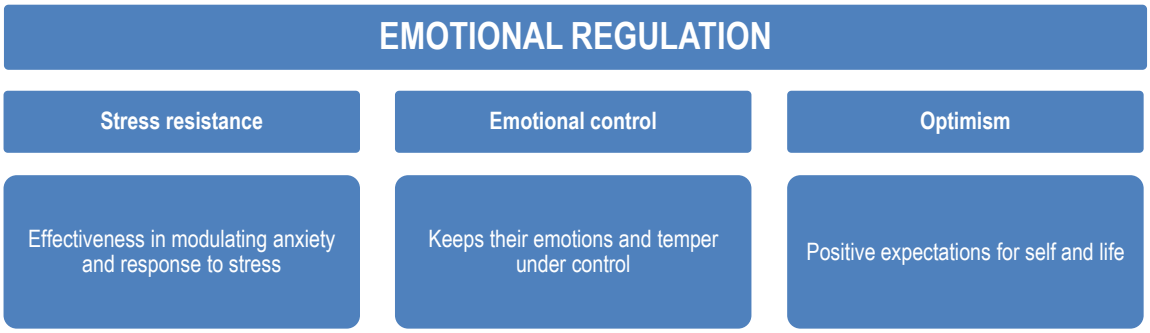
2.5.2. Emotional regulation – having a calm and positive emotionality

Emotional regulation (or emotional stability as it is termed in the Big Five) characterises individual differences in the frequency and intensity of emotional states (Clark and Watson, 2008[126]; Widiger, 2009[127]). It refers to the ability to deal with negative emotional experiences and stressors and is central to managing emotions. Emotional regulation incorporates multiple concepts including anxiety, fear, irritability, depression, self-consciousness, impulsiveness, and vulnerability on the negative side, and notions such as resilience, optimism and self-compassion on the positive side.

The Study includes the following sub-dimensions of emotional regulation:

- stress resistance/resilience vs. anxiety
- emotional control
- optimism/positive emotion

Figure 2.2. Emotional regulation sub-domains



When looking at the predictive ability of separately measured sub-dimensions of emotional stability, optimism consistently predicted school absences for 7th, 10th and 12th graders (Lounsbury et al., 2004[64]). Optimism has also been found to be highly related to life satisfaction (Steel, Schmidt and Shultz, 2008[128]). *Optimism* can be defined as having positive expectations for oneself and optimistic people tend to anticipate success in the

actions they undertake and as having a "can-do" mind-set. In contrast, anxiousness and withdrawal at early ages have also been found to predict anxiety and depression later on in life, in both the Dunedin cohort study (Goodwin, Fergusson and Horwood, 2004[129]) and the Christchurch Health and Development Study (Jakobsen, Horwood and Fergusson, 2012[130]). People who are anxious are unable to solve problems calmly and do not handle stress well.

Soto et al. (2011[104]) focused on developmental changes, especially changes in anxiety and depression between the ages of 10 and 65. They found that anxiety and depression increase sharply in girls between ages 10 and 15 and then gradually decrease into adulthood. However, the pattern for males differs. Anxiety drops quite noticeably between ages 10 and 20 and then gradually decreases into adulthood. Depression, on the other hand, remains relatively constant between ages 10 and 20 and then increases until the early 30s after which time depression gradually declines again until age 65. De Haan et al. (2017[123]) also found that developmental patterns of both anxiety and confidence (i.e. optimism) differ by gender. For girls, confidence slightly decreases while anxiety increases between ages 6 and 17. For boys, a quadratic relation appears to occur for anxiety, with an initial increase followed by a substantial decrease in the later teens. Boys also become less confident until age 14 when their self-confidence increases again.

The sub-dimensions of stress resistance/resilience, emotional control and optimism predict the quality of life, and to a lesser extent, health. They are also all highly relevant for children and be cross-culturally comparable. For the Study, the sub-dimensions should be open to change, and all three of these sub-domains fit this criterion. These sub-domains also can have a positive influence on life outcomes.

2.5.3. Engaging with others – enjoying and excelling in the company of others

Engaging with others relates to extraversion – one of the Big Five dimensions. In early research, there was significant disagreement about which behaviours to include as part of the extraversion domain, but more recent studies suggest focusing on social attention as a valuable sub-domain (Ashton, Lee and Paunonen, 2002[131]). Beginning with the seven facet taxonomies and five adult personality inventories, and especially the three child-based taxonomies (Chernyshenko, Kankaraš and Drasgow, 2018[3]), three sub-domains stood out repeatedly and have been chosen by the OECD to be included in the Study:

- energy/enthusiasm
- assertiveness/dominance
- sociability.

Figure 2.3. Engaging with others sub-domains

Energy/enthusiasm refers to an individual's tendency to approach daily activities with energy, excitement and spontaneity. People with these qualities tend to have a passion and zest for life. Those who are *assertive* can assert their own will in order to accomplish goals in the face of opposition. They can speak out; they will take a stand and are not afraid to confront others. It takes courage to be able to voice one's opinions, needs and feelings, and people who are assertive tend to take on leadership roles. *Sociability* is the ability to approach others and initiate and maintain social connections. Sociable people are skilled at working in teams, are outgoing and comfortable around others.

Research has again shown strong correlations between these sub-dimensions of engaging with others and various outcomes. For example, Judge et al. (2013[132]) found assertiveness/dominance to be related to organisational citizenship, and both dominance and sociability have been found to be strongly correlated with leadership (Judge et al., 2002[67]; Legree et al., 2014[133]). Sociability in childhood has also been shown to be associated with better work competence at age 20 (Masten and Tellegen, 2012[134]), entrepreneurial status at age 34 and earnings among the self-employed at age 34 (Obschonka, Silbereisen and Schmitt-Rodermund, 2012[135]).

Sub-dimensions of extraversion have been found to change over time from childhood through adolescence. For example, de Haan et al. (2017[123]) showed that both energy/enthusiasm declines from age 6 through 17. This substantial decline provides evidence of malleability and hence opportunities for interventions aimed at mitigating these negative trends.

2.5.4. Collaboration – concern for the well-being of others

Individuals who can collaborate successfully with others do so by maintaining positive relations and minimising interpersonal conflict. Showing active emotional concern for others' well-being, treating others well and holding positive generalised beliefs about others are all examples of collaboration (Soto and John, 2017[87]). From the seven facet taxonomies and five adult personality inventories studied by Chernyshenko, Kankaraš, and Drasgow (2018[3]), several sub-domains stand out, three of which have been selected by the OECD for the Study:

- empathy/compassion
- co-operation/relationship harmony
- trust.

Figure 2.4. Collaboration sub-domains

Individuals who score high on *empathy/compassion* are described as warm and sensitive whereas individuals who score low are seen as cold, unsympathetic and insensitive. The sub-dimension, *co-operation/relationship harmony*, distinguishes individuals who are cordial, uncritical, kind and easy to live with. People who score high on *trust* tend to assume the best about people and to act in a trustful way.

Collaborative individuals value interpersonal relationships (Graziano and Tobin, 2002[136]), are more co-operative and helpful (Graziano and Eisenberg, 1997[137]; LePine and Van Dyne, 1998[138]), and are better liked by their peers (Jensen-Campbell et al., 2002[139]). Although collaboration has many positive social benefits and has been found to be negatively related to school absences for 10th and 12th graders (Lounsbury et al., 2004[64]), it is also negatively related to income and earnings (Judge, Livingston and Hurst, 2012[68]; Spurk and Abele, 2010[140]). The ability to collaborate with others translates into stronger relationships, more pro-social behaviours and, among children, fewer behavioural problems.

Some early childhood studies focusing on empathy have found that a lack of empathy is associated with adverse outcomes in adolescents (for example, Fontaine et al., (2011[141]) study of British children). Daniel et al. (2014[142]), also found that sympathy among Swiss children at ages 6 and 9 is associated with social justice values, such as the belief in treating others fairly and minimising inequalities, at age 12.

Soto and colleagues (2011[104]) found co-operation to decline between age 10 and the early teens before returning to initial levels by age 20. De Haan et al. (2017[123]) also found a slight decline in co-operation between ages 6 and 17. These results again show the potential for malleability.

Although none of these three sub-domains appears to be predictive of educational, economic or health-related outcomes, empathy and trust have some predictive validity when it comes to the quality of life measures. On the other hand, agreeableness and its sub-domains are found to be negatively related to various forms of externalising behaviours (e.g. bullying, violence, etc.) both with children and adults. Having empathy, trusting others and being able to co-operate are critical markers of collaboration, and are highly relevant skills for children today and in the future.

2.5.5. Open-Mindedness – exploring the world of things and ideas

Open-mindedness (or openness to experience in Big Five terminology), is regarded as one of the key skills for explaining and understanding the behaviour of individuals in settings characterised by high levels of uncertainty and change (Hough, 2003[143]). Historically, researchers' views have diverged about the precise structure of this broad construct. Consequently, the use of openness measures in applied research has been limited (Ashton et al., 2000[144]). Even at the broadest level, there is disagreement over whether openness to experience should be viewed solely as an intellectual domain (ability to efficiently process information or create new ideas) or whether it should also include other, less intellectualised behaviours, such as tolerance, fantasy, and interest in artistic experiences (Digman, 1990[37]; Goldberg, 1993[145]; McCrae, 1996[146]).

The Study includes the following sub-dimensions of open-mindedness**:**

- intellectual curiosity
- creativity/imagination
- tolerance/cultural flexibility.

Figure 2.5. Open-mindedness sub-domains

Intellectual curiosity exposes interest in ideas and a love of learning, understanding and intellectual exploration. *Creativity/imagination* refers to the capacity to generate novel ways to do things or to come up with new ideas, to think about things through tinkering, learning from failure, or having insight or vision.

Sub-domains belonging to the domain of culture are less commonly considered. However, one sub-domain that surfaces is *cultural tolerance/cultural flexibility* where those who score high on this sub-domain value diversity, appreciate people from different countries and cultures and are open to different points of view.

Different studies have found significant relationships between various openness to experience sub-domains, such as creativity, intellectual curiosity and tolerance, and grade point average (GPA) in both high school and college (Noftle and Robins, 2007[62]; von Stumm, Hell and Chamorro-Premuzic, 2011[147]). Woo et al. (2014[148]) also show that relationships between openness and its sub-domains with cognitive ability (as measured by academic performance) are not very strong, thus indicating that their relationship with grades is not redundant or overlapping. This means that controlling for cognitive ability is unlikely to substantially decrease the ability of the openness sub-domains to predict academic performance. Other studies have found that the intellectual aspect of open-mindedness predicts college graduation and income [for example, Judge et al. (2012[68])].

Both Woo et al. (2014[148]) and Judge et al. (2013[132]) found that openness sub-domains were unrelated to job performance outcomes. The one notable exception was that sub-domains of curiosity and cultural tolerance exhibited higher predictive value with an aggregated adaptive performance criteria measure. This measure included estimates of interpersonal adaptability, expatriate adjustment, creative performance, and coping with organisational change (Woo et al., 2014[148]).

There is also evidence that these social and emotional skills are malleable. For example, Soto et al. (2011[104]) show that average levels of curiosity seem to dip from age 10 to the early teens and this decrease is especially prominent for girls. De Haan et al. (2017[123]) also found a drop in creativity and curiosity between ages 6 and 17.

Intellectual curiosity, creativity/imagination and tolerance/cultural flexibility all appear to moderately predict educational achievement and, therefore, economic outcomes. Particularly, intellectual curiosity is a critical skill that improves learning outcomes and provides intrinsic incentives for personal development throughout life. Tolerance and cultural flexibility are skills with growing social relevance in increasingly diverse and complex societies. Creativity/imagination is a skill that can bring substantial benefits to both individuals and societies.

2.5.6. Additional compound social and emotional skills

The study also includes another type of skills – compound skills – which combine aspects of two or more distinct skills. For example, self-efficacy combines skills from the conscientiousness, emotional stability and extraversion categories of the Big Five. Compound skills are useful to describe and understand certain aspects of behaviour, and they are shown to affect important life outcomes (Chernyshenko, Kankaraš and Drasgow, 2018[3]).

The sub-dimensions of the Big Five tend to be fairly homogeneous. In contrast, compound skills are not one-dimensional as they combine multiple homogeneous skills. Compound skills are advantageous in that they can predict important outcomes as they combine several useful characteristics into an overall domain. However, they are at a disadvantage because it is often unclear which part of the composite measure is driving validity. Just as the five broad domains lack specificity for interventions, so too do compound skills.

The Study included three compound skills in the Item Trials and the Field Test:

- self-efficacy
- self-reflection/meta-cognition
- independence/critical thinking.

Self-Efficacy is selected for inclusion in the Main Study in the form of an index.

Self-efficacy

Self-efficacy reflects the strength of individuals' beliefs in their ability and their effort and dedication in undertaking challenging tasks and achieve goals (Bandura, 1993[149]). And self-efficacy is often a better predictor of students' performance than the actual level of their capabilities, since these beliefs determine how and to what degree they use their knowledge and skills. People with high self-efficacy believe that they can deal with most problems that arise and therefore, do not try and avoid situations they perceive as being difficult. Differences in beliefs about self-efficacy help explain why people with the same level of skills can differ significantly in their performance. Capable individuals may doubt themselves and therefore underachieve, whereas others with modest skills may accomplish more than expected due to their strong belief in their abilities.

Beliefs about self-efficacy are determined by four broad factors: individuals' successful efforts; learning from successful examples (modelling); social persuasion or others' beliefs in individuals' capacity; and physiological factors. Self-efficacy is primarily influenced by the leading social actors in childhood – parents, peers and influential others at school or in their community – and continues to be shaped by experiences and social influences throughout life.

Extensive empirical evidence indicates that self-efficacy influences all aspects of a person's life. They are critical for intrinsic motivation, personal accomplishment and well-being as they influence people's capacity to deal with challenges, and their motivation to initiate actions and persist in the face of difficulty. Moreover, they have a strong influence on people's life choices and the way they interpret the outcomes of their actions and efforts: people with high self-efficacy tend to attribute failure to external factors whereas individuals with low self-efficacy will relate it to their inadequate capacities.

Self-efficacy has been found to be predictive of multiple outcomes. Students with high self-efficacy are likely to take the initiative to learn on their own and actively participate in classes (Bandura et al., 1996[150]; Andrew, 1998[151]) which positively influences their academic performance. Likewise, parents' beliefs regarding their children's academic self-efficacy affect students' self-efficacy and consequently, their academic achievement. Teachers' beliefs in students' self-efficacy influence the kinds of learning environments they create for students (Bandura, 1993[149]). Self-efficacy is also an essential determinant of career choice (Betz and Hackett, 2006[152]; Betz, 2000[153]), job attitudes (Saks, 1993[154]), training proficiency (Martocchio and Judge, 1997[155]) and job performance (Stajkovic and Luthans, 1998[156]; Lunenberg, 2011[157]). High self-efficacy is related to higher job satisfaction and reduced workforce turnover (Cherian and Jacob, 2013[158]; Bradley and Roberts, 2004[159]). However, a meta-analysis of self-efficacy's effect on work-related performance after controlling for the Big Five dimensions – general mental ability, and job or task experience – found the overall predictive power of self-efficacy to be relatively small (Judge et al., 2007[160]). Self-efficacy was a better predictor of performance in low-complexity jobs or tasks than in those of medium or high complexity.

Self-efficacy beliefs also notably relate to the observed under-representation of women in certain occupations such as in science, technology, engineering and mathematics (STEM) fields. Gender differences in self-efficacy expectations seem to influence the career choices of young women; those who are highly competent in mathematics or science often choose other career tracks due to low perceptions about their competence (Zeldin and Pajares, 2000[161]; Herbert and Stipek, 2005[162]).

Self-efficacy also affects a wide range of health-related behaviours, including smoking, exercise, diet, hygiene and self-examination (Conner, 1996[163]). It contributes to the initiation of health improvement or prevention behaviours, the establishment of more ambitious health goals, and persistence in overcoming obstacles. Overall, self-efficacy is a skill that has been widely researched skill and has a high predictive value. It is relatively malleable and especially crucial in school settings.

2.5.7. Behavioural examples of selected skills

Table 2.1 presents a short description of each of the selected skills, accompanied by some typical skill-related behaviour.

Table 2.1. Description of the skills included in the OECD's Study on Social and Emotional Skills

"BIG FIVE" DOMAINS	SKILLS	DESCRIPTION	BEHAVIOURAL EXAMPLES
OPEN-MINDEDNESS (Openness to experience)	**CURIOSITY**	Interest in ideas and love of learning, understanding and intellectual exploration; an inquisitive mind-set.	Likes to read books, to travel to new destinations. Opposite: dislikes change, is not interested in exploring new products.
	TOLERANCE	Is open to different points of view, values diversity, is appreciative of foreign people and cultures.	Has friends from different backgrounds. Opposite: dislikes foreigners or people from different backgrounds.
	CREATIVITY	Generating novel ways to do or think about things through exploring, learning from failure, insight and vision.	Has original insights, creates valued art works Opposite: acts conventionally, not interested in arts.
TASK PERFORMANCE (Conscientiousness)	**ACHIEVEMENT ORIENTATION**	Setting high standards for oneself and working hard to meet them.	Enjoys reaching a high level of mastery in some activity. Opposite: lack of interest in reaching mastery in any activity, including professional competences.
	RESPONSIBILITY	Able to honour commitments, and be punctual and reliable.	Arrives on time for appointments, gets chores done right away. Opposite: doesn't follow through on agreements/promises.

	SELF-CONTROL	Able to avoid distractions and sudden impulses and focus attention on the current task in order to achieve personal goals.	Postpones fun activities until important tasks are completed, doesn't rush into things. Opposite: is prone to say things before thinking them through. Binge drinking.
	PERSISTENCE	Persevering in tasks and activities until they get done.	Finishes homework projects or work once started. Opposite: Gives up easily when confronted with obstacles/distractions.
ENGAGING WITH OTHERS (Extraversion)	**SOCIABILITY**	Able to approach others, both friends and strangers, initiating and maintaining social connections.	Skilled at teamwork, good at public speaking. Opposite: can struggle in working with a larger team, avoids public speaking.
	ASSERTIVENESS	Able to confidently voice opinions, needs, and feelings, and exert social influence.	Takes charge in a class or team. Opposite: waits for others to lead the way, keeps quiet when disagrees with others.
	ENERGY	Approaching daily life with energy, excitement and spontaneity.	Is always busy; works long hours. Opposite: gets tired easily without physical cause.
COLLABORATION (agreeableness)	**EMPATHY**	Understanding and caring for others, and their well-being that leads to valuing and investing in close relationships.	Consoles a friend who is upset, sympathises with the homeless. Opposite: Tends to misinterpret, ignore or disregard other person's feelings.
	TRUST	Assuming that others generally have good intentions and forgiving those who have done wrong.	Lends things to people, avoids being harsh or judgmental. Opposite: is secretive and suspicious in relations with people.
	CO-OPERATION	Living in harmony with others and valuing interconnectedness among all people.	Finds it easy to get along with people, respects decisions made by a group. Opposite: Is prone to arguments or conflicts with others, do not tend to reach compromises.
EMOTION REGULATION (emotional stability)	**STRESS RESISTANCE**	Effectiveness in modulating anxiety and able to calmly solve problems (is relaxed, handles stress well).	Is relaxed most of the time, performs well in high-pressure situations. Opposite: most of the time worries about things, difficulties sleeping.
	OPTIMISM	Positive and optimistic expectations for self and life in general.	Generally in a good mood. Opposite: often feels sad, tends to feel insecure or unworthy.

	EMOTIONAL CONTROL	Effective strategies for regulating temper, anger and irritation in the face of frustrations.	Controls emotions in situations of conflict. Opposite: gets upset easily; is moody.
COMPOUND SKILLS	**SELF-EFFICACY**	The strength of individuals' beliefs in their ability to execute tasks and achieve goals.	Remains calm when facing unexpected events. Opposite: avoids challenging situations.

3. The Study on Social and Emotional Skills Contextual Questionnaires

3.1. Why context is important: Settings in which children develop

Children live and develop in many social settings, including families, schools, peer networks and wider communities. Each of these contexts plays an essential role in their development throughout childhood and adolescence. Characteristics of different environments, their consistency, and how they interact, shape children's social and emotional development. The Study not only assesses students' social and emotional skills across different cultural and linguistic contexts, but it also aims to provide comprehensive information about the characteristics of the environment in which students live. This information gleaned from students themselves, but also their parents, teachers and school principals, helps us understand how family and school contexts affect and enhance various skills development.

Students spend a lot of time with parents and other family members such as siblings and extended family. Although during their teenage years students spend less time with parents, parents remain a significant influence in their children's lives. Families can shape children's social and emotional development by providing guidance, developing routines and habits, imparting values and sharing expectations. Research suggests that supportive and warm families that provide stimulating activities enhance children's social and emotional skills, just as they help to boost children's cognitive skills (Baxter and Smart, 2011[164]; Cabrera, Shannon and Tamis-LeMonda, 2007[165]; Cunha, Heckman and Schennach, 2010[13]). Parental attitudes and disciplinary practices also play an important role in creating children's social and emotional environments (Kiernan and Huerta, 2008[166]). Supportive relationships that generate healthy attachments also positively affect children's understanding and being able to regulate emotions, their feelings of security, and their desires to explore and learn.

Many different factors besides a child's age influence the extent to which parents engage in and affect their children's social and emotional development. Parental characteristics such as their education or employment can influence their academic and career expectations for their children, as well as the time and energy they have to devote to their children. Some argue that employed parents, primarily employed mothers, can impede parent-child bonding (Belsky, 1988[167]; Belsky and Eggebeen, 1991[168]). Others have found little, if any, influence on child behavioural adjustment from mothers working (Cooksey, Joshi and Verropoulou, 2009[169]). Further, both parents working translates into more family income, and therefore being able to purchase materials, services and experiences that positively promote children's cognitive, and social and emotional skills development.

Schools also play an essential role in developing students' social and emotional skills through curricular and extra-curricular activities, or more informally in and outside classroom setting. Teachers can play a particularly prominent role in affecting children's self-esteem, motivation and emotional well-being when they are effective mentors, and help children to learn in the best way they can. Teachers can also affect the way children interact with their peers in classroom settings by encouraging collaboration on school projects, compromise and negotiation, and sociability among pupils. A recent study by Jackson (2013[170]) of 9th graders in the United States found that teachers' abilities to influence cognitive skills, and social and emotional skills were largely independent. This suggests that some teachers may be particularly good at shaping children's social and

emotional skills, but not necessarily as good at shaping their cognitive skills, and vice versa. This finding also suggests that specific teacher characteristics may be particularly conducive to enhancing social and emotional skills (OECD, 2015[2]).

Extra-curricular activities also offer ample opportunities for children to develop social and emotional skills, whether these are focused on sports, music, arts or academics, or revolve around school governance or volunteer activities. Discipline, the ability to work as part of a team, responsibility, negotiation, perseverance, self-control, self-esteem, self-efficacy and curiosity are all potential benefits of these types of activities (Covay and Carbonaro, 2010[171]; Bailey, 2006[172]; Winner, Goldstein and Vincent-Lancrin, 2013[173]).

Peers represent another essential pillar of social influences on student development, with their role gaining importance and surpassing that of parents in adolescent years (Tarrant, 2002[174]). The quality and quantity of peer relations, whether with classmates or with other children, influence all aspects of students' social and emotional skills. With peer groups, children have the opportunity to form relationships with equals and on their terms, without adult control. Peer groups increasingly appeal to children and adolescents as a way to explore socialising and to develop their sense of identity, serving as a source of information on customs, social norms and ideologies. They act as a setting for teaching gender and other social roles as well as promoting group cohesion and collective behaviours (Maslach, Santee and Wade, 1987[175]). Adolescents increasingly also use peer groups to practice preparing for adulthood, learning how to negotiate relationships and to be in contact with different people in the social system outside of the supervision of their parents and teachers.

3.1.1. Structure of contextual questionnaires

Information on background characteristics of students and their parents, as well as on family, school and community learning contexts, was collected through four contextual questionnaires developed for:

- students
- parents
- teachers
- school principals.

The contextual questionnaires aim to capture the most relevant information that influences students' social and emotional skills development in line with characteristics of this Study that tend to be more responsive to policy interventions and adapting teaching methods. They are also relatively reliable and valid information that can be captured in respondents' reports, and concepts that have a firm foundation in empirical research or theory. Along with a range of demographic variables that must be included, content choices have been guided by three questions:

- What lessons can we learn from previous literature and studies?
- Can a contextual characteristic be influenced if it proves to be a significant factor in student socio-emotional development?
- Can a contextual factor be valid and reliably measured when reported by respondents belonging to those same contexts?

In the remainder of Section 3, we outline the content of the four contextual questionnaires, along with a brief elaboration of the importance and relevance of selected topics for the

Study. All four contextual questionnaires administered in the Main Study are provided in the Annex of this document.

3.2. Contextual questionnaire for students

The Study also administers a contextual questionnaire to students in order to gather information on important aspects of their home, school and peer environment. Some of the questions are replicated in the parent and teacher questionnaires. However, collecting information on students' own views about critical aspects of their social context, such as relations with parents, teachers and friends, is important for substantive and methodological reasons. Parents and students often perceive the quality of parent-child relations differently, and each perspective can correlate differently with students' social and emotional skills. Moreover, each perspective can provide new insights into the ways students' environment shapes their skills. Also, obtaining information from both students and parents ensures that this information is captured from at least one source, given that parents' participation rates in surveys such as the Study are often relatively low.

Research shows that environmental factors that influence children's and adolescents' social and emotional skills development are almost exclusively non-shared (Plomin, Kovas and Haworth, 2007[176]). This means that after taking into account parents' genetic influences, shared environmental factors that are common for all siblings in the same family (e.g. family's socio-economic status, parents' education and home possessions) are not that important, especially in regards to children's social and emotional skills development. What seems to be much more important to develop these skills are those factors that siblings do not have in common (Plomin and Daniels, 2011[177]). This does not necessarily mean that parenting styles or the quality of parent-child or parent-teacher relations are not important. Instead, it means that the same situation or environment can be perceived by students differently and that it is this perception that influences students' development rather than the objective reality of a particular situation, e.g. an objective observation of a particular parenting style. In addition, these findings indicate that children themselves shape the behaviour and the quality of their interactions with their family and peers, thus creating personal relations and experiences that are distinct across siblings or across children in the same class. For example, a more active, extrovert and optimistic sibling will evoke different reactions from a parent compared with his withdrawn and introvert brother or sister, thus creating different environmental influences (non-shared environment) within a seemingly shared family environment.

The Study recognises that how students perceive their social environment is critical in determining their experience and development, and why the Study collects relevant information on the context in which students live from students themselves. Coming back to the example on parenting styles, since children's and adolescents' *perception* of how they are parented tends to be more predictive than parents' self-reports of their parenting, the Study asks students for their views.

Box 3.1. Structure of contextual questionnaire for students

- Section A: Demographics
- Section B: Well-being, attitudes and aspirations
- Section C: Relations with parents and friends
- Section D: School life
- Section E: Short cognitive ability measure

3.2.1. Section A: Demographics

This section collects background information in order to measure key socio-demographic indicators that can influence students' social and emotional skills and their educational and well-being outcomes. Questions include date of birth, grade, gender, immigration background and language spoken at home. Students also provide information on their parents' socio-economic status.

The Study's index of economic, social and cultural status (ESCS) is conceptually similar to the corresponding index used in PISA. The ESCS index includes the following variables: parents' occupation using the International Socio-Economic Index of Occupational Status (ISEI); parents' highest level of education, converted into years of schooling; family's cultural capital, using information on the number of books and other cultural items in the family possessions, and the index of other types of material possessions in the family home[3]. Since 2000, the PISA ESCS index has been considered one of the most important variables in educational policy. It influences education outcomes but has also been key in developing equity indicators and research on students' resilience. In the context of the Study, socio-economic background also positively correlates to developing socio-emotional skills and these skills can compensate for the effect of socio-economic disparities on academic performance (Suárez-Álvarez, Fernández-Alonso and Muñiz, 2014[115]).

3.2.2. Section B: Well-being, attitudes and aspirations

This section assesses students' life satisfaction and personal well-being, students' own educational and career aspirations, their perceived mental health, perceived social support from peers, family and teachers, and perceived external pressure to overachieve. Students' implicit theories of the malleability of cognitive and non-cognitive skills are also examined with a short set of questions. Some questions have been taken from PISA 2015, especially from the validated indicators used to develop *PISA 2015 Results (Volume III): Students' Well-Being* (OECD, 2017[1]).

These questions cover important indications of students' quality of life and general well-being. In the Study, they are useful in determining the potential relationship between students' social and emotional skills and different outcomes, such as quality of life.

[3] The SSES index of economic, social and cultural status (ESCS) differs from the one used in PISA in that it distinguishes additional dimension – cultural capital of family, i.e. it separates home possessions into two categories: items indicating wealth and items indicating cultural capital.

3.2.3. Section C: Relations with parents and friends

Previous sections have already addressed how parent and peer relations influence children's development in general, and in particular social and emotional skills. In this section, students are asked about their own view on the quality of these relations, taking into account, as discussed earlier, that their perspective may be somewhat different and more predictive than those obtained from parents and teachers. Questions include: How likely are children to share their problems with family members? How close do they perceive that their relationships are with their parents and siblings? Do children feel loved by their parents? Can they talk to their parents about things that matter to them? Are they involved in shared activities? How do they perceive their parents' parenting behaviours? Analysing the quantity and quality of social interactions between friends, classmates, and parents shed light on how these relations influence learning outcomes and help guide implementing changes (Grunspan, Wiggins and Goodreau, 2014[178]).

Aspects of family culture most relevant to the Study emphasise involvement and connectedness, or lack thereof. Three sets of questions in the student contextual questionnaire measure perceived social support from parents, teachers and friends. These questions look at how students rate how much others understand them, help them when they have problems and accept them when they are not in a good mood. Perceived social support is an established protective factor for mental disorders (Cohen, 2004[179]) and reduces general stress (Lindorff, 2000[180]). Social support has also been found to have cross-cultural measurement invariance (Bieda et al., 2017[181]).

This section also asks students asked about a dimension of perfectionism: external pressure to overachieve. Perfectionism has been associated with student outcomes such as anxiety, depression and life satisfaction, and influencing factors such as parenting styles. It also links to the Big Five, especially conscientiousness (Stoeber, Otto and Dalbert, 2009[182]). Students as young as 10-years-old can show signs of perfectionism, which is important for the Study (Flett et al., 2016[183]; Lozano, Valor-Segura and Lozano, 2015[184]), and it seems to apply across Eastern and Western cultures (Smith et al., 2016[185]). Empirical research also shows that external pressure to overachieve is associated with higher levels of anxiety and depression in children ages 10-15 (Hewitt et al., 2002[186]). Including external pressure in the student contextual questionnaire adds analytical value as a direct assessment of achievement motivation, but it also connects the contexts of students, parents and teachers.

Research suggests that in recent decades the number of social connections is steadily decreasing, which also reflects rising levels of loneliness and alienation in modern societies (Neal and Collas, 2000[187]). One pivotal study in the United States found that the median number of individuals that people could confide in dropped from around 3 in 1985 to around 2 in 2004 (McPherson, Smith-Lovin and Brashears, 2006[188]). The study also found that 1 in 4 adults does not have anyone they can confide in.

These are worrying trends as social connections are one of the most critical factors for a person's health and overall well-being. Social-connectedness represents the subjective experience of interpersonal closeness within society (Lee, Robbins and Steven, 1995[189]). It is based on the quantity and quality of relationships people have, how they evaluate the relationship and its importance in their lives (van Bel et al., 2009[190]). These relationships enable individuals to exchange information, social and emotional support and material aid; to relate or show empathy, belonging and shared identity; and to foster personal growth and well-being.

Not all relationships are equally beneficial, and their effects depend on some factors: level of reciprocity and mutual trust, and on their diversity and intensity. However, even positive relationships can be stressful and daunting at times, while relationships that can be seen as harmful and damaging can have positive aspects. For example, a relationship with someone who is alcoholic can still provide safety, companionship and support (Smyth, Goodman and Glenn, 2006[191]). The diversity of relationships matters as well, especially for children and youth as they need a variety of influences and role models for optimal development (Spencer, Basualdo-Delmonico and Lewis, 2011[192]). Thus, when evaluating social connectedness, it is necessary to understand the broader context, and the diversity and depth of existing long-term relationships.

It is important to distinguish between actual and perceived or subjective social-connectedness since a person may perceive their social network as small and insufficient even if in reality it is relatively large ("loneliness in a crowd"). Likewise, even relatively few relationships may make some people feel well-connected. This perception of social-connectedness determines an individual's well-being more so than her or his real situation. If a person feels well-connected, she or he will enjoy the benefits of being connected irrespective of the actual number of friends, and vice versa.

People are better off when receiving support, but providing support is also associated with positive effects, such as greater self-control, empathy and trust, and higher self-esteem (Thoits, 2011[193]). These outcomes influence others to trust and co-operate more, thus creating a positive cycle of social, emotional, and physical development and stability.

Family greatly influence children during the early years of their development. And peers influence children's behaviour and attitudes more and more as they transition into adolescence (Blakemore and Mills, 2014[194]; Knoll et al., 2015[195]) Knoll et al., 2015). Perceived closeness to other social groups, such as friends, relatives, neighbours, fellow citizens also reflects the degree of students' social-connectedness.

The Study assesses students' closeness to various social groups using a modified version of the Circles of Closeness scale with a similar set of seven overlapping circles (Uleman et al., 2000[196]). It asks respondents to indicate the degree to which they feel close to seven social groups: immediate family, other relatives, friends, classmates, neighbours, fellow citizens and other people in general. Students are presented with pictures of two circles with different degrees of overlap and are asked to pick which pair of circles best describes their relationship with a particular group.

3.2.4. Section D: School life

A number of scales and questions are taken from PISA or developed specifically for this study to measure students' sense of belonging at school, level of their test anxiety, their perception of school safety, their views on their schools' disciplinary climate, and how they view their relationship with their teachers, how engaged they are at school and what their attitudes are towards their school work. Another critical dimension to measure, especially regarding social and emotional skills development, is the relationship that each student has with their peers.

3.2.5. Section E: Short cognitive ability measure

Finally, the student questionnaire includes a short cognitive ability measure. It consists of five verbal/logical/numerical reasoning items and two numerical series items, presented in order of difficulty. The last two items are presented only to students from the older cohort.

Items were sourced from the The International Cognitive Ability Resource (ICAR)[4] item pool of publicly available cognitive assessment measures. Items were selected to be appropriate for both cohorts based on the empirical data available in the ICAR project.

The reason for including cognitive measure in this study is to be able to better examine value of social and emotional skills for a range of life outcomes. In particular, presence of this measure allows us to control for differences in students' cognitive abilities when estimating strength of association between social and emotional skills and various life outcomes. Adding a measure or general cognitive ability for control purposes allows researchers to answer following question: "What are associations of given social and emotional skill and given life outcome for individuals/children with the same level of cognitive ability" (for examples of such results see (Chernyshenko, Kankaraš and Drasgow, 2018[3]; Kankaraš, 2017[5]; OECD, 2015[2])). Given that general cognitive ability is found to be relevant to a wide range of life outcomes and students contextual variables, possibility to control for it in our analytical models would allow us to build a much stronger analytical models.

Another, related reason for use of cognitive ability measure is that it would allow for analysis of the relationships between social and emotional skills and life outcomes for students with different levels of cognitive skills. This type of analysis could offer important policy findings since previous research indicates that social and emotional skills might be even more important for students with less developed cognitive skills (Chernyshenko, Kankaraš and Drasgow, 2018[3]).

3.3. Contextual questionnaire for parents

As parents are the primary caregivers, they have first-hand knowledge of their children's background and family environment (Lippman et al., 2014[197]). In the Study, one of two parents provide information on their child's social and emotional skills (assessment scales) as well as on a wider set of characteristics of family context (contextual questionnaire).

At a broader level, parenting objectives are to improve the knowledge, skills, character and health of the next generation. One measure of successful parenting is that children develop social and emotional skills that enable them to participate fully in the workplace, in their families, and with their friends. The parent questionnaire aims to identify family characteristics that strengthen their children's social and emotional skills development.

The parent questionnaire gathers information on the family's culture, background, parenting behaviours, child's activities and parents' social and emotional skills as well as parents' perceptions of these skills. Previous OECD studies, such as PISA and IELS provide a rich source of questions on family background and home environment. Given the history, technical depth and extensive use of the contextual variables in other OECD surveys, questions are taken from these sources where appropriate, in order to increase consistency and comparability across studies.

[4] The ICAR project is jointly funded by the National Science Foundation (NSF) in the US, the Deutsche Forschungsgemeinschaft (DFG) in Germany, and the Economic and Social Research Council (ESRC) in the UK as part of the Open Research Area Plus for the Social Sciences.

Box 3.2. Structure of contextuel questionnaire for parents

- Section A: Demographics
- Section B: About the child
- Section C: Well-being and skill profile of parents
- Section D: Relations with child
- Section E: Perceptions of social and emotional skills

3.3.1. Section A: Demographics

Socio-economic background and demographic variables are critical factors of students' environment and are particularly important for cross-national research (Adler et al., 2000[198]; Prag, Mills and Wittek, 2016[199]). These kinds of variables enable us to describe and define contexts in which students' skills and life outcomes are socio-economically embedded. The socio-economic status of parents can influence the quality and quantity of parents' involvement in their children's development. Although it is beyond the scope of this Study to gather a full household or relationship history, it does collect information on living situation and family structure. There is a wealth of literature on the benefits of a stable two-parent household, especially on emotional well-being (Amato, 2005[200]). Knowing parents' immigration status and the year they immigrated are useful proxies to determine families' assimilation in society (Jasso, 2011[201]). Other information on families' socio-economic background that is also relevant to students' social and emotional skills development was also collected, such as parents' occupation and employment status, household possessions, and cultural capital.

3.3.2. Section B: About the child

This section gathers relevant information about the students from the parents' perspective. These questions collect information on students' educational trajectory [e.g. if they attended an Early Childhood and Care (ECEC) programme], their general health and habits, peer networks (e.g. the diversity and size of their children's friendship groups), etc. Obtaining reliable information on these indicators not only enables linking results to other OECD surveys and age groups, such as the International Early Learning and Child Well-being Study, but to better understand how students' early experiences, health and health-related behaviours, and peers have a profound and long-lasting impact on students' social and emotional skills.

3.3.3. Section C: Well-being and skill profile of parents

Assessing parents' own social and emotional skill levels provides insight into possible similarities between parents' and children's social and emotional skills. In addition, obtaining information on parents' skill profiles makes it possible to examine the relationship between children's social and emotional skills, and various contextual measures, while also being able to control for possible confounding influences of parental social and emotional skills. Parents' subjective well-being is also assessed in order to examine its possible influences on children's social and emotional skills. For this purpose,

the Study uses the World Health Organisation's WHO-5 scale of subjective well-being (WHO, 1998[202]).

3.3.4. Section D: Relationships with their child

Studies about parenting go back almost a century, but within the past three decades research on parenting has increased dramatically (Holden, 2010[203]). A critical area within parenting research concerns the association between parenting and child outcomes (Fernández-Alonso et al., 2017[185]). How parenting is conceptualised has varied, and a distinction is often made between parenting styles, parenting dimensions and parenting behaviours. Parenting styles relate to Baumrind's (1971[204]) typology of authoritative, authoritarian and permissive parenting. Parenting dimensions incorporate two main constructs: parental responsiveness which refers to *"the extent to which parents intentionally foster individuality, self-regulation, and self-assertion by being attuned, supportive, and acquiescent to children's special needs and demands"* (Baumrind, 1971, p. 62[204]); and parental control which is a multi-dimensional concept made up of psychological control and behavioural control (Barber, 2002[205]). Parenting behaviours refer to different kinds of actions such as discipline or supervision (Socolar, 1997[206]).

Parenting behaviours have been analysed in a variety of ways including interviewing parents and children (Stormshak et al., 1997[207]), directly observing parental behaviours (Collins et al., 2000[208]), and using self-rating questionnaires. Multiple instruments have also been developed to measure parenting behaviour over the past thirty years with the result uncovering many sub-scales. However, not all have been developed for parent self-reports as well as for children to report their perceptions of their parents' behaviours, been translated into other languages, or had their psychometric properties replicated in different cultures.

One scale that does check all these boxes is the Parental Behaviour Scale or PBS (Van Leeuwen and Vermulst, 2004[209]). This scale – widely used and validated in numerous countries – is based on the social interaction theory which initially focused on children's maladaptive and anti-social behaviours and suggested these were related to parents' use of aversive and ineffective management techniques (Meunier and Roskam, 2007[210]; Manrique Millones, Ghesquière and Van Leeuwen, 2014[211]). The evidence showed that (a) positive parenting is inversely associated with problem behaviour in children and levels of stress in parents, and (b) inadequate parenting is positively related to problem behaviour and stress. Finally, results from recent studies by Jannssens et al. (2015[212]) and Van Heel et al., (2017[213]) shows that a measure of psychological control by parents should be added to any comprehensive measurement of parenting behaviours.

Measures of parenting behaviours are essential to include in the Study because they can help to foster or hinder their children's cognitive and socio-emotional development. However, research evidence suggests that children's and adolescents' perceptions of how they are parented may be more relevant than parents' reports for predicting children's outcomes (Gaylord, Kitzmann and Coleman, 2003[214]). This is why the Study also asks students about their views on how their parents raise them.

In additional to parenting behaviours, there are other ways that parents directly or indirectly shape their children's development. Being involved in their children's lives is one such example, and one of the indicators is parents' participation in their children's schools. Empirical research shows that students whose parents use a more indirect parenting style tend to attain better results than those from homes with more control. Parental involvement styles affect achievement at an individual and school level, even after accounting for the

effect of context or background variables (Fernández-Alonso et al., 2017[215]). Numerous studies have shown that parents' willingness to interact with their children is often associated with fewer behaviour problems in children. El Nokali, Bachman and Votruba-Drzal (2010[216]), for example, found social skills improve in children whose parents were more involved, and that parental involvement is amenable to intervention.

Studies on parent-child relationships also show that adolescents with strong positive relationships with their parents are more likely to experience many positive outcomes (De Kemp et al., 2007[217]). Furthermore, it is not just parent-child relationships that are important but family relationships as they are complex, dynamic and intertwined, where each member influences and is influenced by the other members (Minuchin, 1988[218]; O'Brien, 2005[219]). Parents' own relationships are important because relationship problems between parents often negatively affect their children. At the same time, positive couple interactions are beneficial for children residing with both of their parents (Goldberg and Carlson, 2014[220]). One way to examine these issues, is to look at how a family approaches problems.

3.3.5. Section E: Perceptions of social and emotional skills

The Study also asks parents about their views on the malleability of social and emotional, and cognitive skills (i.e. intelligence). Carol Dweck's concept of growth mind-set has gained considerable attention in the research and policy areas in recent decades (Dweck, 2006[221]). People with a growth mind-set think that they can develop or improve any given ability depending on their effort and hard work, while those with a fixed mind-sets think that their abilities are mostly innate and cannot be developed even through effort. Notably, these two sets of people behave and interpret the same situations very differently. In particular, individuals with growth mind-sets are more likely to continue their effort despite encountered setbacks and link their actions' outcomes to their motivation and effort rather than to their talents and abilities. Education settings especially benefit from students who have a growth mind-set, as they are more motivated to invest time and effort in achieving educational goals than students with a fixed mind-set (Yeager and Dweck, 2012[222]).

The Study assesses students' beliefs in the malleability of their cognitive, social and emotional skills. In addition, the Study also assesses parents', teachers' and school principals' growth mind-sets in regards to these abilities in order to examine if their underlying beliefs of skill malleability are related with children's growth mind-sets, and in the long run children's social and emotional skills.

3.4. Contextual questionnaire for teachers

Schools play a large role in preparing students to fully participate and contribute to society – an objective that is universally shared. Compulsory schooling until age 15 is the norm in most countries, and most children over the age of 5 spend 6 to 8 hours a day in school, separated from family and influenced by teachers, school administrators and peers. Numerous studies have focused on school climate and four aspects stand out that the Study takes into account in order to measure the potential influence of school environment on students (Thapa et al., 2013[223]).

These are:

- safety
- teaching and learning
- interpersonal relationships
- institutional environment.

The Study gathers information about these school environment aspects from teachers as they interact directly with students, spend the most time with them, and have the most influence on the classroom environment. Evidence exists from a large body of research on school-based efforts to promote students' social and emotional learning which Elias et al. (1997[224]) defines *"as the process of acquiring core competencies to recognize and manage emotions, set and achieve positive goals, appreciate the perspectives of others, establish and maintain positive relationships, make responsible decisions, and handle interpersonal situations constructively"*. One effective teaching strategy emphasises the importance of active learning where students read, write, discuss and actively participate, instead of passively listen (OECD, 2018[225]; Salas and Cannon-Bowers, 2001[226]).

The Study also asks teachers to provide information on how schools help to develop students' social and emotional skills with questions focusing on factors that enhance students' social and emotional skills. In that context, the Study also looks at teachers' perceptions of and growth mind-sets on the malleability of social and emotional skills.

Box 3.3. Structure of contextual questionnaire for teachers

- Section A: Demographics
- Section B: Education and professional development
- Section C: Teaching practices
- Section D: Teacher's school
- Section E: Perceptions of social and emotional skills

3.4.1. Section A: Demographics

Items include basic demographic information, such as teachers' gender, age, employment status and years of experience. This information provides context when social and emotional skills are connected with educational outcomes and teaching practices.

3.4.2. Section B: Education and professional development

This section includes important details on teachers' education background and the extent to which their training included social and emotional development. Training teachers significantly influence students' social skills in the classroom (Alvarez, 2007[227]). Questions on teachers' continuing professional development also focus on learning pedagogies that enhance students' social and emotional skills as teachers need specific training in order to implement active learning pedagogical practices (described in following section) that promote developing these skills.

3.4.3. Section C: Teaching practices

Teaching practices focus on if and how teachers implement pedagogies that encourage social and emotional skills development. Active learning teaching practices are some of the most suited pedagogies to develop these skills (OECD, 2018[225]). Active learning is generally defined as any teaching method that *"involves students in doing things and thinking about the things they are doing"* (Bonwell and Elson, 1991[228]). In active learning, students are not just passively listening to the teachers' instructions. Instead, they are actively involved in the learning process, with different levels of active learning depending on the degree of their involvement. Active learning encourages students to reflect on the meaning and relevance of the learning material, interact with teachers, and learn from and engage with real-life examples related to content (OECD, 2018[225]).

A growing body of empirical evidence shows that active learning practices can influence a wide range of learning outcomes (Svinicki, 2001[229]; Hoellwarth and Moelter, 2011[230]):

- It improves students' retention of new information, reinforces important insights and skills, and improves the level of understanding and applying learned material to new situations.
- Students feel more connected to topics, increasing their *intrinsic motivation* and *engagement*.
- Students can practice critical *interpersonal skills*, such as *collaboration*, *argumentation*, *communication*, *teamwork*.
- It builds *self-esteem* and a *sense of agency* through actions and interactions during the learning process.
- It improves *higher-order thinking abilities*, such as analysis, evaluation, reflection and synthesis.
- It accommodates different *learning styles* and *personal preferences* of students.
- It creates a *sense of community* in the classroom through *increased interaction* between all actors.

The Study asks teachers how often they implement different active learning practices, such as group learning activities that can foster collaboration, self-paced learning that gives the student more responsibility for their performance, and long-term projects that require students to plan, take responsibility and show initiative in creative activities. *"Socially and emotionally competent teachers set the tone of the classroom by developing supportive and encouraging relationships with their students, designing lessons that build on student strengths and abilities, establishing and implementing behavioral guidelines in ways that promote intrinsic motivation, coaching students through conflict situations, encouraging cooperation among students, and acting as a role model for respectful and appropriate communication and exhibitions of prosocial behaviour."* (Jennings and Greenberg, 2009, p. 492[231]). Apart from teaching practices, in this section teachers are also asked about the criteria and procedures they use to evaluate students' academic performance. These questions highlight to what extent teachers' evaluations of students' performance hinges on their involvement in class, active learning practices and initiatives, and even students' self-evaluation.

3.4.4. Section D: Teacher's school

This section focuses on identifying what schools do in order to promote students' social and emotional skills development. Questions cover whether the school includes the development of these skills in the formal curriculum or whether they are evaluated internally or externally. Gathering information on these relevant topics helps to better connect specific school policies to developing social and emotional skills.

This section also assesses school climate; teachers report on how often students do not comply with school rules and on the quality of relationships between various school groups. Teachers' job satisfaction is assessed as well, as it can relate to teaching outcomes.

3.4.5. Section E: Perceptions of social and emotional skills

The Study evaluates teachers' growth mind-set on the malleability of cognitive, and social and emotional skills. The skills teachers believe cannot be changed might be less likely to be promoted within the classroom setting, even if they are explicitly affirmed in the curriculum and school objectives. Schools could be more effective in promoting social and emotional skills when the teachers and principal share the same growth mind-set about the possibility of changing and improving students' skills.

3.5. Contextual questionnaire for school principals

The schools that children attend – especially their school's climate – greatly influence students' social and emotional development as they spend most of their waking hours on school campuses. School climate comprises many factors: the learning environment, disciplinary climate, rules and regulations, student relations with staff, and staff morale (OECD, 2013[232]). Pedagogical practices mentioned earlier, such as active learning, can help develop students' social and emotional skills in positive ways. School administration and overall school policies, for example, the school's institutional environment, or school safety can also affect teaching and learning. And the overall school climate can have a significant role in influencing the interpersonal relationships that play out within the school setting. For example, Caravita and colleagues (2009[233]) found a positive association between showing empathy and defending victims of bullying.

Education systems from around the world have recognised the importance of fostering a positive school climate, and the necessity of measuring it accurately in order to assess improvements, or lack of them, over time (Cohen, 2012[234]). Researchers have not only documented the positive outcomes associated with social and emotional interventions but suggest that these skills and school climate are interdependent: social and emotional development not only thrives in a positive school environment but also facilitates a supportive climate (Zins and Elias, 2006[235])

Therefore, the Study asks school principals (or their administrative assistants) to provide general information about the school, its curriculum, extra-curricular activities, student body composition, the general level of parental involvement, and the level of conflict or delinquency in the school.

The school principal questionnaire provides relevant contextual information regarding the school's student and teacher demographics. Although not easy to change, school demographics can contribute to explaining student outcomes. School demographics include information such as location, enrolment, percentage of students with immigrant or special needs background, type of school (public or private), funding sources, etc.

Levers, such as school policies and management, influence school climate. While these may differ considerably by jurisdiction, they still share commonalities which are essential for students' social and emotional skills development. These include student tracking based on academic abilities; school pedagogical practices, such as long-term projects and active learning experiences; providing extra-curricular activities; or reasons to use student assessments. Furthermore, the school questionnaire can capture how schools handle conflicts, and which social and emotional skills principals prioritise.

The Study on Social and Emotional Skills aligns many items with the PISA school questionnaire. But because the Study focuses on social and emotional skills and PISA more on cognitive skills, the Study covers more questions on school policies and practices that directly relate to students' social and emotional skills development.

Box 3.4. Structure of contextual questionnaire for school principals

- Section A: Principals' demographics and structure of school
- Section B: Student and teacher population
- Section C: School resources
- Section D: Pedagogical practices, curriculum and assessment
- Section E: School climate
- Section F: Policies and practices

3.5.1. Section A: Principals' demographics and structure of school

Information on principals' demographics and work experience are collected in this section. It also includes questions on the type of community in which the school is located, enrolment by gender, average class size as well as the type of funding source.

3.5.2. Section B: Student and teacher population

This section includes information on the overall characteristics of the student and teacher school population, focusing on those aspects that are relevant to students' social and emotional skills. Questions include the teachers' education profile, and the types of in-house professional development available to teachers.

3.5.3. Section C: School resources

This section provides information on the number and type of school resources dedicated to social and emotional skills development. Questions include whether the school offers formal training on social and emotional skills development, the lack of teaching staff, or the percentage of teachers with more than 20 years of experience.

3.5.4. Section D: Pedagogical practices, curriculum and assessment

This section gathers information on practices that may relate to socio-emotional skills, such as use of active learning teaching practices, extra-curricular activities, the school's and families' community engagement, inclusion of social and emotional skill in school's

curriculum, the school's code of conduct and how it is implemented. This section also looks at how the school incorporates social and emotional learning in assessments, how student assessments are used, and how teacher performance and student learning are assessed. These questions could pinpoint some of the pedagogical elements or broader curriculum and school factors that hinder or foster social and emotional skills development, providing direction for possible policy interventions to improve these practices and overcome identified barriers.

3.5.5. Section E: School climate

This section leans heavily on PISA questions about student and teacher behaviours that harm student learning and general well-being. Questions also focus on students' tendencies to mix socially with different groups, and behaviours that may interfere with student learning and parental involvement in various school-related activities.

3.5.6. Section F: Policies and practices

This section includes information on school procedures on promoting diversity and multicultural tolerance and its' criteria for student selection/admittance. Finally, this section assesses school principals' perception of how malleable cognitive, social and emotional skills are, with the goal to compare them with teacher, parent and student implicit theories of the malleability of these skills.

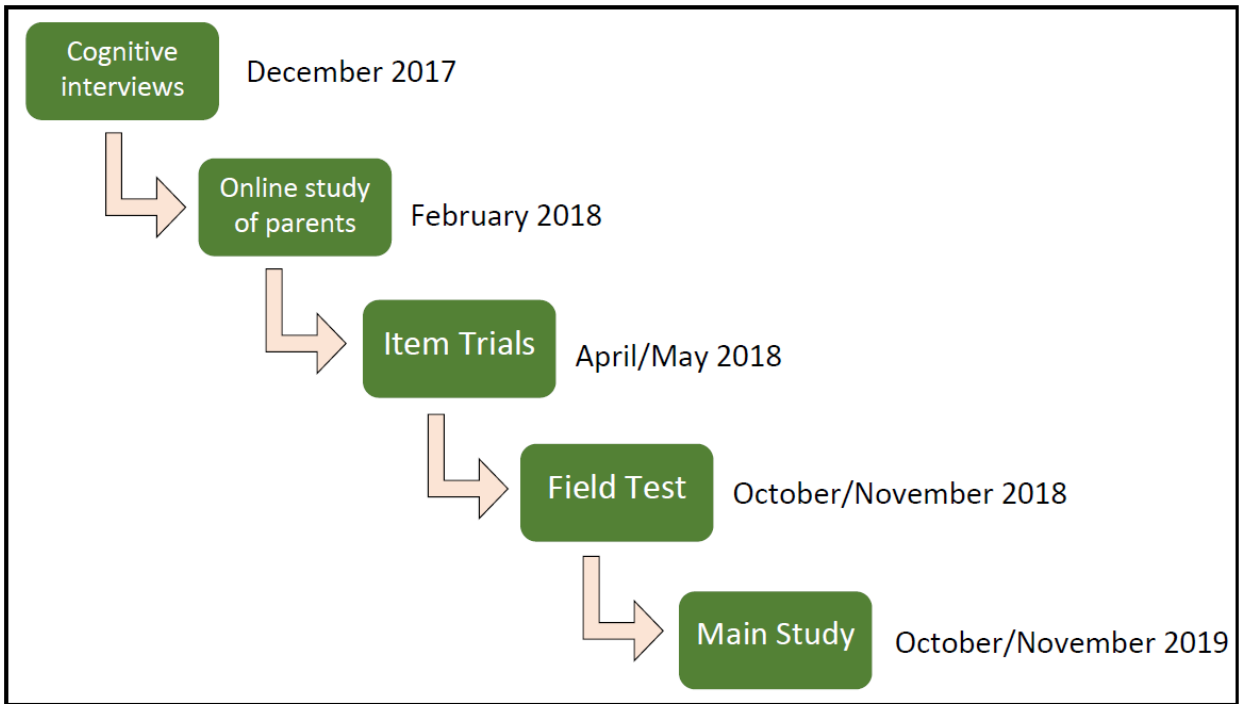
4. Study design

This section outlines the Study of Social and Emotional Skills design – how selected social and emotional skills and contextual factors are assessed and how the Study is implemented. This section also discusses other possible research methods and assessment approaches and the consequences and benefits of different approaches. The process of the instrument development, data collection methods, sampling procedures and the Study's approach to the issue of cross-cultural comparability are also discussed in this section.

4.1. Instrument development

Skill assessment scales and contextual questionnaires needed to be developed for the Study. Wherever possible, existing scales or questions have been used either in their original form or somewhat modified, such as items from the PISA parent, teacher and school contextual questionnaires that were aligned to the Study.

The Study's instrument development process was comprehensive and elaborate, including multitude rounds of empirical testing in various formats (both qualitative and quantitative) and scopes, in order to produce reliable, valid and comparable assessment instruments. Assessment scales were developed specifically for students in the target age groups, with particular focus on being simple, clear, and at an appropriate reading level.



4.1.1. Development of instruments for assessment of students' social and emotional skills

The assessment instruments' development process started with creating an initial item bank of 20 items per skill (or in total 20 items x 19 skills = 380 items). A substantial part of these items were selected from the International Personality Item Pool (IPIP) database of more

than 3 000 items belonging to more than 250 scales designed and used for the assessment of different personality characteristics. Items from a number of other existing scales of social and emotional skills were used as well. Finally, new items were drafted, when necessary, in the same format of the other items obtained from existing scales.

Students' self-reports

Assessment instruments in the Study are divided into two broad groups: self-reports and others-reports. The self-report scales ask students to report on their own behaviour, thoughts and feelings. In others-reports scales, parents and teachers are asked to report on students' behaviours, thoughts and feelings. For comparison purposes, the same items are used in both types of the instruments, although the number of items per scale varies depending on the respondents. Instrument development process for self-report scales, including the number of items in each of the phases is outlined in Table 4.1.

Table 4.1. Development of self-report scales

Study stage	Timeline	Items per skill		Number of skills	
		Older cohort	Younger cohort	Older cohort	Younger cohort
Initial item pool compiled	Nov 2017	20	20	19	19
Item pool reduced after feedback from TAG	Dec 2017	15	12	19	19
Cognitive interviews	Dec 2017		12	19	19
Item Trials	Apr 2018	15	10	19	19
Field Test	Oct/Nov 2018	10	8	19	19
Main Study	**Oct/Nov 2019**	**8**	**8**	**15**	**15**
		Main Study (total items)		**120**	**120**

Note: TAG - Technical Advisory Group.

The International Contractor and OECD teams first compiled an initial item pool of 20 items for each of the 19 skill scales. These items were selected from a wide scope of existing scales with large majority of them being sourced from the International Personality Item Pool (IPIP)[5]. This initial item pool was then evaluated by a group of experts from the Technical Advisory Group (TAG). Based on their feedback, the item bank was reduced to 15 items per skill for the older cohort (15-year-old students). And out of these 15 items per skill, 12 items were selected for the younger cohort (students aged 10).

These 12 items per skill were then examined in the cognitive interviews with 10-year-old students in the United States. Based on results from these cognitive interviews, and TAG members' feedback, some items were reformulated, and two items per scale were excluded from further testing with the younger cohort to reduce their response burden. Thus, in the

[5] https://ipip.ori.org

Item Trials, 15 items per skill were tested for the older cohort, and 10 items per skill for the younger cohort.

Item Trials took place in April and May 2018 in 6 participating sites with 300 students in each of the two cohorts. Based on data from those trials, the best 10 items for the older cohort and 8 items for the younger cohort were selected for the Field Test administration.

The Field Test was implemented in October and November 2018 in all 10 cities. Full two-stage random sampling design was used with 500 students selected in each of the two cohorts, as well as their parents, teachers and school principals (Kankaraš, Feron and Renbarger, 2019[55]). Results from the Field Test were then used to further reduce the number of assessment items for the Main Study. The number of items per skill for the older cohort was reduced from 10 to 8, while it remained the same (8 items per skill) for the younger cohort. In addition, the number of social and emotional skills was reduced from 19 to 15 for both cohorts. Therefore, the entire self-report scales for both cohorts had 120 items.

Parents' and teachers' assessment reports

The same items used in the students' self-report scales across both younger and older cohorts are used in the parent and teacher reports. This facilitates comparing instruments and information about students' social and emotional skills obtained from multiple respondents.

Many assessment items require self-reflection, such as those related to emotional regulation or meta-cognition. In order for items to be comparable across respondents, items must focus on observable behaviours as indicators of social and emotional skills because self-ratings and reports provided by others are more similar for items reflecting observable behaviours, and thus can be more easily compared (Varni et al., 2015[236]). Question wording, syntax and semantics have been kept as simple as possible to minimise the cognitive burden across the three responding groups; maximise the comparability of the assessments; and ease the burden of scale translation. While students and parents provide a report on one student, teachers were often reporting on multiple students. This is why assessment scales for teachers have been kept as brief as possible in order to lower their response burden.

Parents' and teachers' assessment scales were first tested in February 2018 in an online study of parents, with 1 000 parents from the United States participating. This study enabled initially examining measurement properties of 19 assessment scales based on parent reports, including their predictive validity with a range of life outcomes. In this study, assessment instruments consisted of 10 items per skill. After that, in the Item Trials, parents and teachers evaluated their assessment scales through cognitive interviews. They provided information on how they understood the questions if the questions were well-formulated and appropriate, and on the time needed to complete these scales.

Parent and teacher assessment scales were administered during the Field Test, with 8 items per scale for parents' reports and 3 items per scale for teachers' reports (for each of students). Based on the findings from the Field Test, including results from students' self-report scales, parents' and teachers' scales were then further revised for the Main Study. This revision was done with the goal to select best 8 items for the parent scale and best 3 items for teachers scale based on the item properties in all of the scales these items were used. The instrument development process of **parents' and teachers'** assessment scales, including the number of items in each of the phases, is presented in Table 4.2.

Table 4.2. Development of students' self-report assessment instruments

Study stage	Timeline	Items per skill		Number of skills	
		Parent	Teacher	Parent	Teacher
Online study of parents	Feb 2018	10		19	
Item Trials	Apr 2018	10	10	19	19
Field Test	Oct/Nov 2018	8	3	19	19
Main Study	**Oct/Nov 2019**	**8**	**3**	**15**	**15**
		Main Study (total items)		**120**	**45 per student**

4.1.2. Development of contextual questionnaires

The aim behind the development of the contextual questionnaires was first to review the literature in order to map out the factors in students' environment that potentially influence their social and emotional skills development, such as their family, school, peer and broader community environment. The second phase involved examining empirical evidence about the importance of each of the previously selected contextual factors, their characteristics and relations with other relevant variables. In the third phase, existing measures for selected critical contextual factors were evaluated, such as contextual questionnaires from other OECD studies, primarily PISA, in order to see if questions on topics of interest had already been developed. Questions from other OECD studies have not only been well-researched and tested, but they provide opportunities for cross-study comparisons to be made. They have been translated into multiple languages, thus reducing time and costs for the Study. The Study also drew on studies that had carefully developed questions.

New or modified versions of questions used in the Study were initially tested during the Item Trials with the initial group of six cities and countries through parent and teacher cognitive interviews. Results were used to improve the questions and how they were formulated in preparation for administering the contextual questionnaires in the Field Test.

After the Field Test, results from the contextual questionnaire items and scales were thoroughly analysed and evaluated. Based on this examination, decisions were made about which questions and scales should remain in the questionnaires, which need to be modified and finally which questions or scales need to be removed. It was necessary to reduce the number of questions as much as possible while still capturing the most important factors of students' environment in order to limit the response burden of each of the four contextual questionnaires. Such a situation inevitably led to some hard choices, but it also ensures that respondents' time is used more efficiently.

4.2. Methods of assessment

4.2.1. Measurement approaches

There is a long tradition of measuring social and emotional skills both in academic and applied settings, and a wide range of instrument and assessment techniques have been

developed for these purposes. However, the measurement of social and emotional skills has not reached the same level of quality as for cognitive abilities (Heckman and Kautz, 2012[7]). Moreover, the lack of quality in measuring social and emotional skills may be one of the main reasons that they are absent from policy discussions (Brunello and Schlotter, 2011[237]). This is not surprising as social and emotional skills have proven to be more challenging conceptually than cognitive skills at every step of the measurement process. They are more difficult to define and differentiate from other similar concepts; the constructs are harder to transform into reliable and valid scales, or to observe in behaviour; and it is more challenging to establish the hypothesised relations with similar or related constructs and with various life outcomes.

Many different measurement approaches are used to assess social and emotional skills, including self-reports; reports from others, such as peers, teachers and parents; behavioural observations; performance tasks; biographical data; lab experiments, situational judgment tests and think-aloud protocols (Kankaraš, 2017[5]).

Self-reports

By far, most social and emotional skills measures are based on self-report questionnaires, i.e. individuals, whose skills are being assessed, report on their typical behaviours, thoughts and feelings. Various pros and cons exist with self-reporting. On the plus side, self-report questionnaires provide a simple and efficient way to collect information from large numbers of people. They are cost efficient and simple to administer, produce consistent results, and in many cases provide a remarkably good approximation of objective measures (Duckworth, Tsukayama and May, 2010[120]; Connelly and Ones, 2010[238]).

For example, one assessment scales of the SSES study, each of which consists of eight assessment items, takes on average just around one minute of students' response time. Thus, all 15 assessment scales can be administered in about 15 minutes per student of older cohort and around 20 minutes per students of younger cohort. This compares extremely favourably with response times needed in performance-based tests of both cognitive and non-cognitive skills. For example, performance-based assessment of emotional intelligence (Mayer-Salovey-Caruso Emotional Intelligence Test) takes between 30-45 minutes. Likewise, PISA's assessment of Math or Reading proficiencies take 50 minutes per domain.

Similar limitation in terms of efficiency is present in case of situational judgment type of assessments, where children are presented a situation description and have to select the most appropriate response from a predefined set of reactions. Major issues with this type of assessment is that they usually span a very small range of constructs relative to the number of items and time necessary to measure a construct reliably. Besides, these methods identify if people are able to select that particular answer for which subject matter experts agree that this is the most appropriate response, which is very different from explaining how people will react in daily life.

Moreover, a considerable body of literature in the social sciences indicates that people generally react reasonably well to questionnaires and are in general able to describe their typical behaviour in the intended way (Krosnick, 1999[239]; Heine, Buchtel and Norenzayan, 2008[240]). From a practical point of view, they are effectively the only feasible measurement form for use in large-scale international surveys.

One of the most important constraints of the self-report scales is that respondents may misinterpret the questions in a number of way. This may be due to use of difficult or

ambiguous words or formulation, or respondents' lack of motivation, fatique or specific characteristics (i.e. immigrant background). Students are also not equally reliable reporters of their inner states, such as feelings or self-perceptions. They might also have difficulty in retrieving required information in which case they might resort to provide a socially desirable answer. Even when they are motivated and interpret the question correctly, they may be suspectible to various memory biases and inconsistencies.

Social desirability is another threat to the validity of self-reports. It represents the tendency of respondents to answer questions in a way that they believe will be viewed favourably by others. This tendency can either lead to "desirable" behaviour being overstated or "undesirable" behaviour being understated. Social desirability is difficult to disentangle from the substantive interpretations of the scale. It is an especially important issue in high-stakes individual diagnostic situations, such as job selection or clinical screening, where respondents may try to present themselves in the best possible light. However, presence of social desireable answers is possible also in low-stake studies like the SSES. In order to avoid or reduce the effects of social desirability, questions need to be worded in a neutral manner, avoiding overly favourable words and balancing the desirability of opposing response options.

Whenever the so-called Likert scales are used – i.e. where respondents are asked to determine the level of agreement with a particular statement, mostly using five answer options from "strongly agree" to "strongly disagree" – responses are subject to various response-style biases. The most ubiquitous among these is an "acquiescence" response style – the tendency to agree with statements irrespective of their content. The acquiescence response style can be controlled by using both positively and negatively-worded statements in a scale (see our approach to controlling for this response style in section 4.2.3).

Another validity threat to both self-reports is so-called "reference bias", i.e. a situation in which people from different countries answer the same question using different reference standards. In particular, a question such as: "I see myself as someone who tends to be lazy" (a question from the conscientiousness scale of a Big Five questionnaire) may be answered differently depending on a person's standards or reference points regarding what it means to be lazy. Our approach towards controlling for occurrence of reference bias is described in section 4.2.3.

One of the critical issue for the SSES study is the ability of younger students to provide valid and reliable information on their typical behavours, thoughts and feelings. This is why the age of our younger cohort was determined in accordance with previous research that has established that students' self-reports on personality and social and emotional skills can be used from the age of 10 onwards (Soto et al., 2011[104]). This capacity is dependent on a series of critical factors, including language proficiency, but also cognitive and social development (John and De Fruyt, 2015[34]). First, children need to have acquired a certain vocabulary and a basic reading level to be in a position to administer the assessment. Simplicity and clarity in language is anyway an important requirement for skill descriptive items, because assessments not only have to be completed by children and adolescents themselves, but often also by parents of different socio-economic classes, and teachers who will have to rate multiple children in their class. These constraints require grammatically streamlined and short items, an easily understandable response scale format, and clear instructions.

Probably more important is that children also need to have developed some first self-reflective and social-comparison skills. According to Barenboim (1981[241]), children first make behavioural comparisons (e.g. "Ricardo runs faster than Patrick"), and start to

actively use trait terms thereafter (e.g. "Eva is shy"). Furthermore, children's person-perception skills need to develop into a multi-dimensional scheme, to a point where they have a notion that multiple independent trait attributes may apply to themselves or another person. During development, children first associate a single individual with one characteristic [see for example the figures portrayed in children's books and comics that are even named after a single trait, e.g. the different smurfs, each with their typifying characteristics, Asterix (small but smart) and Obelix (raw power), gnome "Lui" (lazy), …], and this perspective needs to progress into a multi-dimensional space of person-perception that can be used to describe differences between, but also within persons. The evidence available right now suggests that this is achieved by age 10-11, in line with the emergence of formal-operational thinking (John and De Fruyt, 2015[34]; Soto et al., 2011[104]).

Others' ratings

Many of the constructs can be measured both through self-rating and others' ratings. The advantage of this approach compared to self-ratings is that other people may be more objective and reliable evaluators of some characteristics of an individual than the individuals themselves. Some research suggests that for certain behavioural characteristics, such as academic achievement or job performance, others' ratings may be more accurate, unbiased and predictive than self-ratings (Connelly and Ones, 2010[238]). The important factor is the degree to which raters know the person they are rating, but even in situations where trained raters have only known the subject for a short time, the predictive value of these ratings may be higher than those obtained from self-rating (Lindqvist and Vestman, 2011[242]). Teachers' ratings have a strong predictive validity for some students' behavioural indicators (Segal, 2011[243]), and are especially valuable for younger students, whose self-ratings tend to be less reliable.

Others' ratings are also useful when combined with self-ratings, providing complementary information in order to create a more comprehensive assessment, and identify and correct certain types of measurement issues, such as social desirability or memory bias. Moreover, others' ratings from different sources can add predictive validity to each other (Connelly and Ones, 2010[238]). For example, parent- and teacher-ratings of students can complement the student's self-rating, adding predictive value for various student outcomes (MacCann et al., 2010[244]).

Others' ratings are also subject to some of the same measurement issues as self-ratings, such as lack of knowledge, memory bias, social desirability (especially when raters are close to the individual they are rating), reference bias and response-style bias (Connelly and Ones, 2010[238]). Moreover, when personal experience, inner feelings and thoughts are the focus of research, others' ratings are less appropriate as external observers have only indirect access to these mental states.

Other approaches to assessment of social and emotional skills

Although alternative assessment approaches can avoid some of the common risks of the self- and others-report scales, each of these alternative methods have their own limitations that largely prevent their widespread use for assessment of social and emotional skills (Duckworth and Yeager, 2015[20]; Connelly and Ones, 2010[238]). Performance tasks and observations have an advantage that they do not rely upon the subjective judgments of students, parents or teachers. In this way they circumvent reference bias, social desirability bias, agreement bias bias, and faking. In addition, by measuring performance in a single moment of time they might be more sensitive than self- and others-reports to subtle changes

in behavior. On the other hand, one of their key drawbacks is that overt behaviour may be caused by different situational factors and personal characteristics that are not related to measured construct. This is especially true in the domain of social and emotional skills where performance in one behavioural task might be influenced by a inter-related set of skills. Thus, although the performance itself is objective phenomenon, its interpretation is based on a strong and often questionable assumptions that given performance is consequence of just one, measured psychological characteristic of a respondent.

Importantly, given that performance tests are usually focused on assessing maximal performance of a respondent, they might not adequately represent his or her typical behaviour in everyday's situation (Duckworth and Yeager, 2015[20]). However, for most social and emotional skills (e.g. empathy, co-operation, emotional control, curiosity, etc.), what matters more is how person usually behaves, rather than how can one behave if he or she really tries his or her best. Performance tests are also suspectible to practice effects, i.e. biasing influence of test familiarity on results of successive administration of a test. Observation and performance task are also significantly lengthier than reports on typical behaviour and are usually more costly and difficult to implement.

The situational judgment tests to avoid some of the limitations of self- and others-reports such as having a different reference point/standard or misunderstanding the question. This technique also overcomes the biggest drawback of the observational approach since it invokes the relevant test situations (and related response) directly, rather than waiting for it to happen in reality, which can often be impractical and/or costly. However, apart from its lengthy administration, the situational judgment test technique also suffers from some of the same limitations as self- and others-reports. For example, they are also suspectible to respondents' tendency to provide socially desirable answers. In addition, the concreteness of their assessment material, although a strong point from the perspective of their ecological validity, can constrain their applicability across varous populations in terms of age, socio-economic background, culture, etc.

Biodata provide another potential route of assessing social and emotional skills. Their main advantage comes from the potential validity of using objective and concrete behavioural outputs as measures of given skills. In this way, these data would be able to avoid many of the pitfals of questionnaires. But use of biodata for estimation of relevant skills is severely limited due to the fact that valid and reliable biographical markers of specific skills are difficult to determine. In addition, privacy concerns and other ethical considerations restrict the scope of potential research topics and applications using these information sources.

Triangulation of data on students' social and emotional skills

Primary and secondary school students' social and emotional skills are the primary focus of this Study and are assessed using both students' self-reports and others' reports provided by their parents and teachers. Therefore, some of the drawbacks of students' self-reports can be offset by using information from parent and teacher reports and vice versa, enabling mutual validation of skill estimates from different sources (Kankaraš, Feron and Renbarger, 2019[55]).

The triangulation of students' socio-emotional skills assessment is a critical component of the Study. First of all, the assessment through parents and teachers increases the content validity of the estimates of students' socio-emotional skills as it provides information on students' behaviours across different contexts. Research shows that all perspectives have unique and valuable viewpoints on individual differences, with reports correlating with one another between 0.30 – 0.60 (John and De Fruyt, 2015[34]). The magnitude of these correlations suggests that all perspectives share some variance, but also have their own

specific and informative viewpoint. This is a critical aspect as students may behave differently in different settings and choosing information from any one of those settings may provide a somewhat biased representation of students' social and emotional skills. The opportunity to combine information on students' skills from personal, school and family perspectives yields a better representation and understanding of students' behaviours in the most important contexts that affect school-age students. Likewise, the ability to obtain information from other sources that know the student well, permits controlling for various sources of measurement error presented in self-reports, such as social desirability or unrealistic self-perceptions.

Parents are valuable sources of information on typical behaviours, thoughts and feelings of their children, especially at a younger age. They have long-term and close relationships with their children, have seen them grow and develop, and know first-hand their life situation, personal preferences and practices. They have observed their children in the family context but also across a range of situations, including children's relations with other influential people, such as their friends and peers.

Given the Study's focus on students' social and emotional skills, teachers' reports are highly valuable because they have much experience dealing with many children and can evaluate students' social and emotional skills as a reasonably objective non-family member. Furthermore, teachers have experience with children in the more structured context of the classroom, and are in a good position to observe more interpersonal and task-oriented skills, whereas parents provide ratings relying on the home-context (John and De Fruyt, 2015[34]). In addition, teachers rely on a much broader frame-of-reference to describe pupils' characteristics, because they accumulate professional experience with tens of new pupils in their classroom each year, whereas the scope of parents is usually much smaller and more idiosyncratic.

The parents' and teachers' reports, just like the students' self-reports, follow a standard Likert-type format, with statements describing typical behavioural patterns of students and response categories representing various degrees of agreement with the statement. Additionally, the assessment items are a selection of those used in the students' self-reports. This improves comparability of estimates of students' social and emotional skills and the instruments' properties across the three groups of respondents. For fuller examination of the triangulation approach implemented in the Study and its outcomes based on the Field Test data please consult recent publication on the topic (Kankaraš, Feron and Renbarger, 2019[55]).

Short self- and others- rating scales

In large-scale surveys, existing instruments designed for assessing social and emotional skills are often too long to use. Over the past few decades, studies have tried to shorten the length of instruments, leading to positive and negative outcomes. They reduce respondent burden and boredom which can increase the quality of answers, and multiple studies have found that short scales have the same comparable levels of criterion validity as longer scales measuring the same construct (Burisch, 1984[245]; Robins, Hendin and Trzesniewski, 2001[246]; Thalmayer, Saucier and Eigenhuis, 2011[247]). Other studies have shown that short scales have satisfactory test-retest reliability and convergence validity (Gosling, Rentfrow and Swann, 2003[248]; Robins, Hendin and Trzesniewski, 2001[246]). Moreover, shorter and quicker assessments make them more cost effective and leave more room to assess other relevant concepts and contextual factors, thus enriching the overall analytical potential of a study. On the negative side, the use of short scales tends to lead to increased measurement error and consequently lower estimate reliability (Kankaraš, 2017[5]). In addition, fewer

items means that fewer aspects of a measured construct can be assessed, reducing content validity. Finally, when measures only have one or two items per scale, they cannot be used to identify underlying, or statistically latent, scale structures. This has both methodological and substantive consequences and can affect the Study's aim to assess latent constructs that cannot be measured directly.

Therefore, the Study attempts to use the smallest amount of questions for assessing each of the selected social and emotional skills, while still ensuring satisfactory levels of the reliability, validity and comparability of obtained skill estimates. An extensive process of instrument development was used to select the best items that provide the most information per unit of response time. In this way, the Study has developed a set of assessment instruments of students' social and emotional skills based on students' self-reports as well as parent and teacher reports. These instruments can be applied in the restrictive context of large-scale international studies, in different cultures, at different times and for children of different ages.

Future development of assessment instruments

Self- and others-reports that are developed for use in this study are by far the most common assessment method in the area of assessment of social and emotional skills. This has to do to their unique set of advantages such as efficiency and solid psychometric quality that sets them apart from other assessment approaches. Important limitations of alternative approaches, such as performance tests, have reduced their applicability and amount of use.

However, in recent years, a lot of important findings and developments has happened in the area of psychological assessment in general and assessment of social and emotional skills in particular. Availability of information technologies has vastly improved the possibility for interactive, adaptive and formative assessment methods, as well as for powerful and fast statistical analyses. Coupled with increased attention to the area by both researchers and policy-makers, this has led to proliferation of new types of instruments and assessment approaches (e.g. game-based tests, computer adaptive testing, situational judgment tests, use of log-data, etc.). Most of these new methods are in relatively early stages of development and, at this moment, none of them seem to have psychometric superiority over the more traditional measures, especially once development costs and response times are taken into account. Nevertheless, given the pace of progression and the amount of research in this area, it seems that it is just a matter of time before new types of instruments successfully overcome some of the pitfalls of the older measures.

This is why in the longer-term strategy of the Study, there is a possibility for testing and development of new types of assessment approaches. Such methods could be used in conjuction with existing self- and others-report scales in order to further improve the quality of our assessment data.

4.2.2. Response formats

The type of response format that assessment items use also has pros and cons. Consider a hypothetical sub-domain – sociability. As discussed in the previous section, survey respondents are often asked to rate themselves or someone they know on a series of statements (item stems). Frequently, they are presented with a standardised response format, and researchers infer scores on the variable of interest from these ratings. For example, an indicator of sociability might be the statement, "I like to attend parties". A variety of standardised response formats are then possible for such a statement.

The most often used response format is the Likert responses agree/disagree scale with either four or five options. This scale is sometimes also presented as a dichotomous agree/disagree scale. There are various positives and negatives associated with each of these two options. For example, a virtue of the dichotomous choice format is its simplicity as the respondent does not need to calibrate *the extent* to which the item applies. It may also be less subject to response scale biases that can vary substantially across countries. He, Bartram, Inceoglu, and van de Vijver (2014[249]), for example, showed that large cultural differences exist in the extent of extreme response style (ERS; i.e. the use of 1 and 5 on a five-point rating scale), as well as midpoint response style (MRS; i.e. the use of 3 on a five-point rating scale). Reducing cross-cultural response bias is crucial in the Study in order to facilitate comparisons across international study sites. A question that remains, however, is whether or not eliminating extreme and middle options would eliminate the problem of response bias. It also does not help with acquiescence bias.

On the other hand, a dichotomous format provides less psychometric information than a multiple-choice format. In other words, dichotomous response formats require more questions to obtain the same precision of the skill estimate compared to a 4- or 5-point, multiple-choice format. This decreased efficiency represents especially important concern in large scale international studies, such as the Study. Additionally, a dichotomous format may not provide a meaningful option in all situations and for all participants.

Another important decision regarding the response scale format is whether or not to include a neutral point category. Some people might honestly neither like nor dislike attending parties. So for them, neutral response option (e.g. “neither agree nor disagree”) is the most valid answer, i.e. it best represents their situation in regards of the measured construct. An argument against including a neutral category is that some people who lean more in one direction or the other might use it as an “escape” category, i.e. might choose a non-committal answer.

On the other hand, judgments needed to meaningfully respond to a standard Likert scale with anchors from *strongly disagree* to *strongly agree* may be difficult for young children to make (Chambers and Johnston, 2002[250]; Mellor and Moore, 2013[251]) – a problem that is especially salient for this Study’s assessment of 10-year-olds.

Whether dichotomous or polytomous, scales with *agree/disagree* response formats tend to be influenced by acquiescence response style (Krosnick, 1999[239]; Saris et al., 2010[252]). One strategy for mitigating this problem is to provide a balanced set of items, in which half represent high levels of the latent construct (positively-worded items) and half represent low levels of the construct (negatively-worded items). The latter is then reverse-scored before analysis. Again, this option is not without problems as inclusion of negatively-worded items tends to complicate the dimensionality of a measure (Edwards et al., 2010[253]; Woods, 2006[254]); item reversals can create complications in translating measures across populations with different socio-linguistic backgrounds (Wong, Rindfleisch and Burroughs, 2003[255]); and negatively written items may be confusing to 10-year-old children (Suárez-Álvarez et al., 2018[256]). Thus, the use of negatively-worded items is a topic of continuous discussions among methodologists.

A five-point agree/disagree Likert scale was selected for the Study due to the amount of empirical evidence regarding its measurement properties and to the familiarity of this scale across survey respondents of different ages, backgrounds and nationalities. Also, given the time limit, students can only respond to a limited number of items. Using a five-point response scale increases the amount of information obtained per question, thus reducing response burden and increasing the assessment instruments’ efficiency.

4.2.3. Innovative assessment design elements

Since 2000, OECD surveys like PISA, PIAAC and TALIS have been using a Likert-type scale for self-report items in order to measure contextual factors in their questionnaires. Self-reports have provided the most validity evidence in scientific literature for social and emotional skills assessments, and are the most widely-used survey design in policy research. However, to increase the validity of students' self-report assessment scales and cross-country comparability, some considerations must be taken into account. For instance, PISA has consistently found that the directionality in the relationships between some background constructs measured with Likert scales and achievement outcomes at the individual student level is not consistent with those at the aggregated country level. While such inconsistencies might stem from real differences in how relationships play out at individual and country levels or from omitted variable bias, they might, in fact, result from systematic differences among countries in how students interpret the agreement response scale or in response styles.

In order to address these potential issues, several new survey methods were introduced in the Study on Social and Emotional Skills to enhance the validity of questionnaire indices, especially for cross-country comparisons. The following innovative assessment designs complement and do not replace the direct self-report assessment.

Anchoring vignettes to improve cross-cultural comparability

Reference bias represents one source of cross-cultural incomparability for self-report measures (Kankaraš, 2017[5]). It refers to a situation in which people from different countries answer the same question using different reference standards. In particular, a question such as: "I see myself as someone who tends to be lazy" (a question from the conscientiousness scale of a Big Five questionnaire) may be answered differently depending on a person's standards or reference points regarding what it means to be lazy. Possibly as a consequence of this, national rankings on the Big Five scale of conscientiousness do not correlate with factual measures such as average working hours (Schmitt et al., 2007[111]). Reference bias is a problem when comparing aggregate data between cultures, but not when comparing individual scores within the same culture. Analysis of data from the Programme for International Student Assessment (PISA) shows the expected positive association between self-reported academic performance and conscientiousness within countries, but between-country results indicate a negative association, with countries with higher scores on conscientiousness, performing worse in math and reading (Kyllonen and Bertling, 2013[257]).

Anchoring vignettes are sets of questions specially designed to account for reference bias. They are designed to identify the reference system used by respondents for evaluating behaviours presented in a given scale. Based on the answers obtained from anchoring vignettes, respondents' answers to assessment scales are adjusted to account for differences in their reference systems. This adjustment could reduce possible bias introduced by respondents from different cultures using different reference systems for evaluating the same behaviours. Scientific literature seems to agree that vignettes should be used for correcting content-related item scores, rather than correcting across unrelated scales. This means that the Study should have specific vignettes for each Big Five domain, which would then be used to correct items within that domain only.

Previous studies using PISA 2012 data show that responses to vignette questions represent valid individual and country differences and that the adjusted responses tend to show higher levels of comparability (He, Buchholz and Klieme, 2017[258]). However, the effectiveness

of this strategy depends, to a high degree, on whether two assumptions hold true: anchoring vignettes are supposed to be invariant across respondents, and the responses to vignette prompts are supposed to be without error and strictly ordered (von Davier et al., 2017[259]).

Kyllonen and Bertling (2013[257]) have reported some success in minimising cross-national differences by using anchoring vignettes in PISA. With this methodology, respondents are presented with three vignettes to which they respond using a Likert scale rating. The same scenarios are used in all nations/cultures, and the mean rating for individuals within each culture are computed. These within nation means are then used to adjust ratings on items assessing the constructs of interest. Put another way, when items are scored based on vignettes, numerical values for responses are not assigned based on the concrete response option chosen (e.g. the value 5 for *"strongly agree"* and 4 for *"agree"*) but instead on the self-reported answer relative to the personal standard captured by the rating of three vignettes. The anchoring vignettes approach has been used for cross-country comparisons in various fields of research (Kapteyn, Smith and Van Soest, 2007[260]; Saloman, Tandon and Murray, 2004[261]; Kristensen and Johansson, 2008[262]), but PISA 2012 is the first large-scale education assessment to use this approach. Anchoring vignettes do require additional administration and increased assessment time, which make it important that their benefits are explored and confirmed.

Aware of these considerations and their potential limitations, and learning from PISA's 2012 cycle, new anchoring vignettes were developed for initial testing in the Study. For each Big Five model domain, a set of three anchoring vignettes were developed, depicting behavioural characteristic for the high, medium and low end of that dimension. Thus, a total set of 15 anchoring vignettes were drafted and tested in the Study. Table 4.3 presents an example of the three anchoring vignettes for the domain, Emotional regulation.

The same 15 anchoring vignettes are included in all three respondents' questionnaires: student, parent and teacher. They were first tested on students in both cohorts during the Item Trials. Based on the Item Trials findings, anchoring vignettes were modified and prepared for the Field Test.

In the Field Test, anchoring vignettes were tested in two ways. Firstly, results from students, parents and teachers assessment scales were compared before and after correcting for their responses to anchoring vignette, with particular emphasis on measuring cross-cultural comparability in order to see if it improves by using anchoring vignettes. Secondly, anchoring vignettes were used to see if they improved the quality of assessment report answers, by looking at whether answering anchoring vignettes before assessment scales helps respondents learn to better use given response scales of assessment items. This, in turn, could lead to respondents' answers that are more faithful representation of their assessed skills, thus improving assessment scales' measurement properties. In order to test this hypothesis, a split-design method was used in the Field Test, with half of the randomly assigned respondents having completed anchoring vignettes (before the assessment scales) and other half of respondents not having anchoring vignettes at all.

Findings from the Field Test indicate that anchoring vignettes could be used to improve cross-cultural comparability and overall measurement properties of the assessment scales. Therefore, it was decided that anchoring vignettes would be used in the Main Study as well. However, in the Main Study they are implemented in student and teacher questionnaires using a split design in order to reduce respondent burden. In particular, each respondent provided answers to one of five sets of three related vignettes that were randomly assigned across respondents. On the other hand, all five sets of 15 anchoring vignettes were administered in all parent questionnaires.

Table 4.3. Example of anchoring vignettes

		How much do you agree that each of the following students handles his/her emotions well?	Strongly disagree	Disagree	Neither agree nor disagree	Agree	Strongly agree
Emotional regulation	+	[Mary] never looks stressed. She is calm and positive even before an exam. How much do you agree that [Mary] handles her emotions well?	1	2	3	4	5
	0	[John] is usually calm during classes, but there are a few moments when she can get upset and change moods. How much do you agree that [John] handles his emotions well?	1	2	3	4	5
	-	[Anna] is frequently in a bad mood and gets upset every time someone does something she doesn't like. How much do you agree that [Anna] handles her emotions well?	1	2	3	4	5

Behavioural indicators

Gathering information on how students behave – or manifest their social and emotional skills – is an important addition to the student, parent and teacher reports on these skills. A set of questions are included in the student, parent and teacher questionnaires regarding students' behaviours at home or in class/school that correspond to key social skills, such as class behaviours, school absenteeism, behaviours with friends and parents, disruptive behaviours and health-related behaviours. Using national longitudinal data from the United States, Segal (Segal, 2011[243]) created a misbehaviour measure in which the students' teacher rated the students' behaviour together with concrete behavioural indicators. He found that misbehaviour ratings predicted lower educational attainment and lower earnings at age 26-27 (Kyllonen, 2012[263]). Social and emotional skills can be expressed as behavioural indicators, and these concrete indicators can act as an intermediate step between developing these skills and obtaining outcomes.

Educational assessments usually present sub-domain results (e.g. math or achievement motivation) and their connection with key outcomes such as educational attainment, labour market, or well-being. This traditional focus on broad life outcomes is still important for all education surveys, including the Study. In the last few years, a complementary approach, connecting assessment results to specific observable behaviours has gathered validity evidence in personality assessment (Chapman and Goldberg, 2017[264]). More importantly, using behavioural indicators is especially relevant for policy context since it improves the link between survey results, and policy actions and implementation.

The OECD Secretariat, in consultation with the project's stakeholders, decided to include behavioural indicators in the student, parent and teacher questionnaires. After reviewing the literature and consulting with TAG members, 24 behavioural indicators for parents and teachers, and 20 indicators for students were selected for administration in the Field Test. Most of the behavioural indicators were the same for all three respondent groups (students, parents and teachers). However, there were also some indicators asked only to students and teachers (about students' behaviour in school classes) and to students and parents (about their behaviours at home). Finally, there were also some questions that were asked only to students from the older cohort.

After the Field Test, this list was reduced to six behavioural indicators for younger cohort and four additional ones for older cohort. In addition, five indicators were administered to parents and six to teachers. In the Main Study each of the behavoural indicators was used only with one group of respondents, i.e. either with students, parents, or teachers.

Examples of behavioural indicators related to the Study include:

- He/she skips school classes
- He/she has trouble sleeping
- He/she always participates in class activities
- He/she does house chores (e.g. cleaning room, making bed, etc.)
- He/she never gets in fights

Controlling for acquiescence response style

Whenever the so-called Likert scales are used – i.e. where respondents are asked to determine the level of agreement with a particular statement, mostly using five answer options from "strongly agree" to "strongly disagree" – responses are subject to various response-style biases. One of the most common among them is an "acquiescence" response style – the tendency to agree with statements irrespective of their content. This response bias is found to have substantial effects on people's responses to questionnaire items and a consequential detrimental impact on the detectability and clarity of the five-factor (Rammstedt, Goldberg and Borg, 2010[265]; Rammstedt and Kemper, 2011[266]).

One way to control for the acquiescence response style is by using both positively and negatively-worded statements in a scale, so that the effects of the agreeing tendency can be identified and controlled for. In some cases the number of positively and negatively-worded items in a scale is kept the same in order to achieve full balance out over the scale (Rammstedt, Goldberg and Borg, 2010[265]). However, negatively-worded items are often harder to understand (with occasional use of 'double negations'), have higher cognitive load and response times, are more sensitive to differences in educational and socio-economic backgrounds in respondents (better educated respondents from privileged backgrounds are more likely to understand them in correct way), and are in general more likely to be misunderstood and consequently less valid. This is the reason why they tend to have somewhat lower psychometric quality (reliability and validity indices) and to lead to the scales with lower precision and internal consistencies. Also, negatively-worded items are less used and developed so it is harder to find an equal number of negatively-worded items that have at least remotely similar empirical foundation and psychometric properties.

This is why in the initial item pool, as well as in the final set of scales used in the Main Study, an unequal number of negatively-worded items was selected, depending on the

number of available negatively formulated items of a satisfactory quality. In the Main Study, number of negatively formulated items in the 15 assessment scales varies from 1 to 7 (out of total 8 items per scale). Analyses of the Field Test data have shown that negatively formulated items are useful for identification and control of the acquiescence response style. They will be used for the same purposes during the analysis of the Main Study data.

4.2.4. Mode of survey administration

All students instruments (students' self-reports and students' contextual questionnaires) are administered via the Internet (online administration) using desktop computers. There are multiple advantages to using online delivery: significant cost savings due to a more efficient way in preparing and administering instruments; improved data quality due to a decrease in scoring errors as they are captured automatically in real time and do not have to be manually inputted later. The centralised online administration also facilitates seamless administration of the assessment instruments in multiple languages and for multiple respondents at the same time, thus enabling administering assessment measures to groups of students. Finally, online administration facilitates continuous monitoring of the process of gathering data, making it possible to track not only participation rates but also timing information and even some of the aspects of data quality at any moment of study administration. Importantly, research has shown that young learners are adept at answering questions online (Anderson, 2015[267]).

The teachers' assessment reports as well as teacher and principal contextual questionnaires are also administered online. However, parents' reports and parent contextual questionnaires are administered in both online and a paper and pencil format. The paper and pencil format is useful if parents do not have access to the Internet or in case that that they prefer such format.

Given that teachers assessg students' social and emotional skills, it is important to gather this information from one teacher who is likely to know the student well. In the case of 10-year-olds that will probably be their main classroom teacher. For 15-year-olds, this could be any of the student's teachers who knows the student well. Selection of these teachers for each student was conducted by School Co-ordinators, which were usually either school principals or school management staff who had access to school administration data and have known school organisation and teacher class assignments well.

One parent or guardian completes the parent questionnaire although some important information on the socio-economic background of both parents is asked in this questionnaire.

Ordering of different sections of assessment instruments and contextual questionnaires across all four respondent groups is presented Table 4.4.

Table 4.4. Order of instrument administration across four groups of respondents

Order	Students	Parents	Teachers	Principals
1	Anchoring vignettes	Anchoring vignettes	Part A (only once): Contextual questionnaire	Contextual questionnaire
2	Assessment instrument	Assessment instrument	Part A (only once): Anchoring vignettes	
3	Behavioural indicators	Behavioural indicators	Part B (for each assigned student): Assessment instrument	
4	Contextual questionnaire	Contextual questionnaire	Part B (for each assigned student): Behavioural indicators	

4.2.5. Timings of questionnaires

Respondents should not be too burdened by the time it takes them to complete their surveys. The table 4.5 presents expected average lengths of time for administration of assessment instruments and contextual questionnaires for all four respondents in the Study:

Table 4.5. Timings of questionnaires across respondents

Timing estimates	Contextual questionnaire	Assessment of students' skills	Total time (Main Study)
Students	15 – 25 minutes	15 – 25 minutes	~ 40 minutes
Parents	15 – 25 minutes	~ 15 minutes	30 – 40 minutes
Teachers	15 – 20 minutes	4 – 6 minutes	30 – 40 minutes (with average number of assigned students)
Principals	25 – 35 minutes		25 – 35 minutes

Note: These expected timings are based on the response times in the Item Trials and Field Test.

Suggested response times do not include logging in times as well as, in case of students, times for instructions. A minimum of 1 hour and 30 minutes was allotted for the student assessment sessions[6] in order to accommodate for those students who progress more slowly through the questionnaires. The maximum number of student reports per teacher is capped at 15.

Study respondents

The Study includes students from two age cohorts, 10- and 15-year olds. Taking a snapshot of two age groups – 10 and 15 – enables cross-national comparability and also makes it easier to link Study results with relevant child development literature, most of which is age-referenced.

Assessing students at age 10 provides information on how they are progressing in school and what may be needed during the next years of schooling. And age 15 is, in many countries, the last point in time where nearly all young people are in some form of formal education setting. This is also the age when students are assessed in PISA. Finally, including two age levels enables participating cities and countries to (1) identify differences

[6] These were administered in schools from which students were sampled, in form of group administration.

in social and emotional skills between two different age groups, and (2) better understand the likely development of social and emotional skills between two points in students' schooling.

Including data from schools and parents also enables participants to ascertain the extent to which differences in social and emotional skills between the two age groups are associated with individual characteristics, students' family and home backgrounds, and the approaches in their schools and education systems. When matched with information on students' educational attainment, researchers can examine the extent to which social and emotional skills either hinder or foster cognitive skills development and how this might differ between the two age groups.

- *Target population:* The Study assesses students in two age cohorts – 10 and 15 – that attend educational institutions located within the administrative borders of participating cities and countries. Ten-year-old students are considered the youngest that can reliably answer questions about their behaviours, thoughts, and feelings. While 15-year-olds are at a different period in their lives, they are also at a point where "nearly all" members of their cohorts are still in formal schooling. Also, they are the same age as adolescents assessed in PISA, providing an opportunity for cross-study comparability. Defining the target population by age instead of by grade provides an opportunity to compare results across countries and economies. Likewise, age-based samples make it easier to link the Study's results with relevant literature on the development of social and emotional skills, most of which is age-referenced.
- *Sample of students:* Data are collected from 3 000 students in each of the two cohorts (ages 10 and 15). Sampling is a two-stage process: first, schools within a city/country are randomly selected, followed by randomly selecting students within those schools. The Study uses a stratified random sampling of schools with the selection probability being proportional to school size. This is standard practice for rigorous sampling and is the approach used in peer OECD school-based studies such as PISA and TALIS. This sampling design aims to be a reliable representation of the entire target population outlined above.
- *Sample of parents:* For each sampled student, parents or legal guardians are asked to participate in the Study by filling out a contextual questionnaire, and assessment instrument (report on students' skills). Only one of the two parents/guardians fill out the questionnaire, and the parents/guardians themselves decide.
- *Sample of teachers:* For each sampled student, the teacher that knows the student the best or with whom the student has spent the most time is selected. These teachers are asked to fill out the teacher contextual questionnaire as well as a small assessment instrument for each of the assigned students.
- *Sample of school principals:* For each sampled school, school principals are asked to fill out the contextual questionnaire for school principals.

Cross-cultural comparability

One critical goal of the Study is to provide an internationally recognised measurement tool for policy-makers and education practitioners to use in order to assess essential social and emotional skills for students across diverse student populations and settings. However,

distinct methodological challenges, such as problems relating to bias and equivalence can occur, which hinder cross-cultural comparisons.

In order for results to be comparable across cultures and countries and be free of reference bias, they need to measure the same construct or trait in each group. Therefore, accurately translating questions into each local language is inadequate as the people responding must understand the questions in the same way as intended in the source version (Hui and Trandis, 1985[268]; Van de Vijver and Leung, 2001[269]; Kankaraš and Moors, 2010[108]). As noted by Van de Vijver and Tanzer (2004[270]), *"In every cross-cultural study, the question as to whether test scores obtained in different cultural populations can be interpreted in the same way across these populations has to be dealt with"* (p. 119[270]). Implementing the same assessment instruments across participating cities and countries in the Study does not, by itself, ensure that obtained estimates of social and emotional skills are directly comparable. The degree of such comparability would need to be examined, and any identified bias would need to be accounted for, before direct comparison of findings across participating sites in the Study takes place.

Extensive evidence exists on the Big Five domains' and their sub-domains' comparability across cultures and nations (Paunonen et al., 1996[110]; McCrae and Costa Jr., 1997[42]). Although research has shown that some culture-specific constructs exist [e.g. Cheung et al. (2001[271])], the common Big Five dimensions and their facets are clearly present in most cultures and languages, making cross-cultural comparisons feasible (fuller examination of this topic is presented in the conceptual framework of the Study – Chernyshenko, Kankaraš and Drasgow, (2018[3]). For example, in their Personality Profiles of Cultures (PPOC) project, McCrae and Terracciano (2005[28]) examined factor replicability of a widely-used adult personality measure, the NEO-PI-R, in 50 *cultures* using translations into several languages. The factor structure was replicated in most cultures and was recognisable in all (McCrae and Terracciano, 2005[28]). Ashton et al. (2004[83]) also showed that the factor structure of adjectives was remarkably similar across seven studied language groups (Dutch, French, German, Hungarian, Italian, Korean and Polish). The Big Five and their facets have not only been replicated across cultures but also across age groups and genders [see, for example, Tackett et al.'s (2012[50]) study involving 3 751 children in 5 countries and 4 age groups between ages 3 and 14].

Two primary sources of incomparability occur in cross-cultural research: construct and method biases (Van de Vijver and Leung, 1997[272]). Construct bias occurs when an instrument measures constructs that have different meanings or only partially overlap across countries, making cross-cultural comparison impossible. Method bias originates from methodological and procedural aspects of cross-cultural studies – from characteristics of samples from different cultures (e.g. when a sample in one culture includes minority populations and in others it does not), survey instruments to which individuals from different cultural groups react in consistently dissimilar ways (e.g. when different words or item content are more familiar to one culture than to others, or response styles differ across countries), and through administration bias due to various procedural aspects of the data collection effort, for example interviewer characteristics or use of different administration formats across countries.

As the Study's focus must be relevant for all cities involved by including social and emotional skills that are not subject to construct bias, it conducted an extensive literature review investigating the relevance and meaningfulness of social and emotional skills in diverse cultures (Chernyshenko, Kankaraš and Drasgow, 2018[3]). The review confirmed

previous findings that the Big Five and their sub-domains are found to be generally applicable across cultures and nations.

Multiple steps were also taken to mitigate method bias. First in terms of instrument development, items chosen to assess the various sub-domains were 1) reviewed for idiomatic expressions as these rarely translate well; 2) reviewed for item content to identify material that might not be relevant in all cultures; and 3) examined by a multicultural team of experts from a wide-range of backgrounds. Second, an extensive and careful translation process was set up to ensure accurate and culturally-appropriate translations that minimise linguistic bias. This process is described in Section 4.3.1. Third, sampling and survey administration procedures were standardised across all participating countries in accordance with the *Study on Social and Emotional Skills Technical Standards* in order to ensure that differences in results across participating cities and countries are not due to different ways in which the Study is administered in their jurisdictions.

Fourth, in order for scores to be comparable across cultures and countries, accurately translating questions into each local language doesn't go far enough. The Study needed to ensure that people responding to questions understand them in the same way (Hui and Trandis, 1985[268]; Kankaraš and Moors, 2010[108]; Van de Vijver and Leung, 2001[269]). Extensive sets of psychometric analyses, both at the scale and the item levels, were conducted at different stages of this process to evaluate cross-cultural comparability of all assessment instruments rigorously. After each of these phases, the analyses results were used to exclude or modify those items that were found to be culturally incomparable. This process ensured that only items that could work best cross-culturally were used in the Main Study instruments.

Finally, 15 anchoring vignettes were developed and implemented in order to improve cross-cultural comparability by controlling for possible reference bias across countries. Possible improvements in being able to compare skill estimates across participating sites after accounting for respondents' answers in anchoring vignettes were examined following the Field Test and the Main Study.

4.3. Adaptation and translation

The Study was administered in a large number of countries with different languages, cultures and school systems. Students' social and emotional skills were assessed through both self-reports and reports from teachers and parents. In addition, the study instruments also included four contextual questionnaires used to collect information on the students' family background, peer and school environment.

In order to collect internationally comparable data in the study, translations of these instruments were required, and each of the national versions used by participating cities and countries must meet stringent quality standards. It is crucial to ensure that the translation process does not introduce biases likely to distort international comparisons by:

- unintentionally modifying the content of the questions by changing their meaning or by expressing the questions in a manner that might re-frame the stimulus and hence the response
- introducing ambiguities that could impair some of the variables collected through the background questionnaires

- adapting instruments to the national context in ways that extensively change the data collected (e.g. when adaptations result in undesired changes in the test administration or coding procedures).

Adaptation

Specific terms within the source versions of each questionnaire need to be adapted to ensure their applicability in local settings. Examples include names of people or places which may need to maintain their specific significance but should be easy to read, and sound familiar to students, or terms relating to countries' education systems, such as the equivalent term for ISCED 2.

The International Contractor vetted and recorded all the adaptations proposed by national teams. In a number of cases, adaptations of original source wording was required. For example, local teams needed to add their own language or educational institution titles to some questions. In other cases, adaptations were optional and depended on the local teams' and the International Contractor's evaluation of the local context. Although adaptations of individual terms were allowed or required where necessary, no structural changes to the existing questions were permitted: inserting or adding extra rows to accommodate additional items, or deleting existing rows with existing items within the question. However, participating cities and countries could added up to ten national items at the end of each of the four questionnaires.

Translation

All source instruments for the study were developed in English. Participating cities and countries in which the local language or languages were not English were tasked with translating these instruments. The Study assesses students in the language in which they are taught. Therefore, survey questions were translated into languages to which students, teachers and parents were comfortable responding. In addition, even participating sites where the main language of administration is English (Ottawa and Houston) needed to adapt the instruments in order to make sure that all the items corresponded to their local contexts.

The translation process was twofold. Each local team hired three translators, fluent in English, for each language their site used in order to translate student assessments, contextual questionnaires and any additional documentation, such as instructions that the Test Administrators read to students. Then, cApStAn was responsible for verifying the translations.

The team translation approach is summarised below and described fully in the Translation Manual and corresponds to the ITC Guidelines for Translating and Adapting Tests (ITC, 2017[273]).

- Each site hires three translators.
- ACER[7] trains translators on how to implement the team translation approach within ACER's web-based translation software system.
- Items that need to be translated are divided among the site's three translators who work on their own sections.

[7] Australian Council for Educational Research.

- Each translator is assigned to translate 2/3 of the material, thus ensuring that each item is translated by two independent translators.
- After translating 2/3 of material assigned to them, each translator reviews and provides feedback on the work of the other two translators, and then they discuss in order to reach a consensus on each of the translated items.
- Then, a local psychometric expert reviews these initial translation, analysing whether the translation and phrasing for each of the items aligns to the concepts that they intend to measure and other psychometric considerations.
- The translated instruments are then sent to the verification subcontractor (cApStAn) who reviews the translations and suggests changes, when necessary. The two parties discuss any suggested changes and agree on the joint version.
- This version is then sent to an independent translation referee, who evaluates the translations agreed by local translation teams and cApStAn. Translation referees also intervene in cases where an agreement between local teams and cApStAn cannot be reached.
- After the local teams, cApStAn and a translation referee confirm the final translations, they are sent to the International Contractor and uploaded on the online platform.

This extensive process of instrument translations was conducted three times among the six sites participating in the Item Trials, and then in all 10 sites with 11 different languages for the Field Test and the Main Study. In those cases where items remained the same between the Field Test and the Main Study, local teams used existing translation of the item from the Field Test. However, in all other cases where a new or modified item was introduced for the first time in the Main Study, the translation process needed to be organised in accordance with the steps outlined.

4.4. The Study on Social and Emotional Skills Technical Standards

The Study on Social and Emotional Skills is based on the highest scientific standards in the area of large-scale survey research that have been developed in order to ensure that the data collected is of the highest quality. The standards and guidelines used in the Study are presented in the Technical Standards document (OECD, Forthcoming[274]). The standards for data collection and submission were developed with three major, and inter-related, goals in mind: (1) consistency, (2) precision and (3) generalisability of the data. Furthermore, the standards serve to ensure a timely progression of the project in general.

The Technical Standards detail the procedures and required standards on the following topics:

- target population and sampling
- adaptation and translation processes
- field test administration procedures
- main study administration procedures
- confidentiality and security protocols and requirements
- quality monitoring process

- assessment mode
- communication protocols
- data management.

Interested readers are invited to refer to this document for a full description of all technical aspects of the Study, i.e. how the study is implemented. The Technical Standards, together with this Assessment Framework and the previously published Conceptual Framework (Chernyshenko, Kankaraš and Drasgow, 2018[3]) are key references of the Study on Social and Emotional Skills.

5. References

Abrahams L., P. (2019), "Social-Emotional Skill Assessment in Children and Adolescents: Advances and Challenges in Personality, Clinical, and Educational Contexts", *Psychological Assessment*, Vol. 31/4, pp. 460-473. [54]

Abrahams, L. et al. (2019), "Social-emotional skill assessment in children and adolescents : advances and challenges in personality, clinical and educational contexts", *Psychological Assessment*, Vol. 31/4, pp. 460-473. [35]

Adler, N. et al. (2000), "Relationship of subjective and objective social status with psychological and physiological functioning: Preliminary data in healthy, white women", *Health Psychology*, Vol. 19/6, pp. 586-592. [198]

Almlund, M. et al. (2011), "Personality psychology and economics", *National Bureau of Economic Research Working Paper 16822*, NBER, Cambridge. [61]

Alvarez, H. (2007), "The impact of teacher preparation on responses to student aggression in the classroom", *Teaching and Teacher Education*, Vol. 23/7, pp. 1113-1126. [227]

Amato, P. (2005), "The impact of family formation change on the cognitive, social, and emotional well-being of the next generation", *Future of Children*, Vol. 2005/15, pp. 75-96. [200]

Anderson, P. (2015), *Teaching Reading in the Early Years: Research Developments*, ACER, Melbourne. [267]

Andrew, S. (1998), "Self-efficacy as a predictor of academic performance in science", *Journal of Advanced Nursing*, Vol. 27, pp. 596-603. [151]

Ashton, M. (1998), "Personality and job performance: The importance of narrow traits", *Journal of Organizational Behaviour*, Vol. 19, pp. 289-303. [29]

Ashton, M., K. Lee and S. Paunonen (2002), "What is the central feature of extraversion? Social attention versus reward sensitivity", *Journal of Personality and Social Psychology*, Vol. 83, pp. 245-252. [131]

Ashton, M. et al. (2004), "A six-factor structure of personality-descriptive adjectives: Solutions from psycholexical studies in seven languages", *Journal of Personality and Social Psychology*, Vol. 86, pp. 356-366. [83]

Ashton, M. et al. (2000), "Fluid intelligence, crystallized intelligence, and the Openness/Intellect factor", *Journal of Research in Personality*, Vol. 34, pp. 198-207. [144]

Bailey, R. (2006), "Physical education and sport in schools: A review of benefits and outcomes", *Journal of School Health*, Vol. 28/2, pp. 117-148. [172]

Bandura, A. (1993), "Perceived self-efficacy in cognitive development and functioning", *Educational Psychologist*, Vol. 28/2, pp. 117-148. [149]

Bandura, A. et al. (1996), "Multifaceted impact of self-efficacy beliefs on academic functioning", *Child Development*, Vol. 67/3, pp. 1206-1222. [150]

Barber, B. (ed.) (2002), *Intrusive Parenting: How Psychological Control Affects Children and Adolescents*, American Psychological Association, Washington DC. [205]

Barenboim, C. (1981), "The development of person perception in childhood and adolescence - from behavioral-comparisons to psychological constructs to psychological comparisons", *Child Development*, Vol. 52/1, pp. 129-144. [241]

Baumrind, D. (1971), "Current patterns of parental authority", *Developmental Psychology Monographs*, Vol. 4, pp. 1-103. [204]

Baxter, J. and D. Smart (2011), "Fathering in Australia among couple families with young children", *Occasional Paper No. 37*, Department of Families, Housing, Community Services and Indigenous Affairs, Australian Government, Canberra. [164]

Belsky, J. (1988), "The effects of infant daycare reconsidered", *Early Childhood Research Quarterly*, Vol. 3/2, pp. 235-272. [167]

Belsky, J. and D. Eggebeen (1991), "Early and extensive maternal employment and young children's socioemotional development: Children of the National Longitudinal Survey of Youth", *Journal of Marriage and Family*, Vol. 53/4, pp. 1083-1098. [168]

Benítez, I. et al. (2016), "Linking extreme response style to response processes: A cross-cultural mixed methods approach", *International Journal of Psychology*, Vol. 51, pp. 464-473. [296]

Benson, P., P. Scales and A. Syvertsen (2010), "The contribution of the developmental assets framework to positive youth development theory and practice", *Advances in Child Development and Behavior*, Vol. 41, pp. 197-230. [18]

Betz, N. (2000), "Self-efficacy theory as a basis for career assessment", *Journal of Career Assessment*, Vol. 8/3, pp. 205-222. [153]

Betz, N. and G. Hackett (2006), "Career self-efficacy theory: Back to the future", *Journal of Career Assessment*, Vol. 14/1, pp. 3-11. [152]

Bieda, A. et al. (2017), "Universal happiness? Cross-cultural measurement invariance of scales assessing positive mental health", *Psychological Assessment*, Vol. 29/4, pp. 408-421, http://dx.doi.org/10.1037/pas0000353. [181]

Blakemore, S. and K. Mills (2014), "Is adolescence a sensitive period for sociocultural processing?", *Annual Review of Psychology*, Vol. 65, pp. 187-207. [194]

Bogg, T. and B. Roberts (2004), "Conscientiousness and health-related behaviors: A meta-analysis of the leading behavioral contributors to mortality", *Psychological Bulletin*, Vol. 130, pp. 887-919, http://dx.doi.org/10.1037/0033-2909.130.6.887. [70]

Bono, J. and T. Judge (2004), "Personality and transformational and transactional leadership: A meta-analysis", *Journal of Applied Psychology*, Vol. 89, pp. 901-910. [66]

Bonwell, C. and J. Eison (1991), "Active learning: Creating excitement in the classroom", *ASHE-ERIC Higher Education Report No. 1*, The George Washington University, School of Education and Human Development, Washington DC. [302]

Bonwell, C. and J. Elson (1991), "Active learning: Creating excitement in the classroom", *ERIC*, https://eric.ed.gov/?id=ED336049. [228]

Boyce, C. et al. (2015), "Personality Change Following Unemployment", *Journal of Applied Psychology*, Vol. 100/4, pp. 991-1011. [74]

Bradley, D. and J. Roberts (2004), "Self-employment and job satisfaction: Investigating the role of self-efficacy, depression and seniority", *Journal of Small Business Management*, Vol. 42/1, pp. 37-58. [159]

Brown, P., M. Corrigian and A. Higgins-D'Alessandro (eds.) (2012), *Measuring and improving school climate: A pro-social strategy that recognizes, educates and supports the whole child and the whole school community*, Rowman and Littlefield. [234]

Brunello, G. and M. Schlotter (2011), "Non-cognitive skills and personality traits: Labor market relevance and their development in education and training systems", *IZA Discussion Paper no. 5743*. [237]

Burisch, M. (1984), "Approaches to personality inventory construction: A comparison of merits", *American Psychologist*, Vol. 39, pp. 214-227. [245]

Cabrera, N., J. Shannon and C. Tamis-LeMonda (2007), "Fathers' influence on their children's cognitive and emotional development: From toddlers to pre-K", *Applied Developmental Science*, Vol. 11/4, pp. 208-213. [165]

Caravita, S., P. Di Blasio and C. Salmivalli (2009), "Unique and interactive effects of empathy and social status on involvement in bullying", *Social Development*, Vol. 18/1, pp. 140-163. [233]

Caspi, A. and R. Shiner (2006), "Personality development", in Damon, W., R. Lerner and N. Eisenberg (eds.), *Handbook of Child Psychology: Social, Emotional, and Personality Development*, 6th ed., Wiley, New York. [44]

Chambers, C. and C. Johnston (2002), "Developmental differences in children's use of rating scales", *Journal of Pediatric Psychology*, Vol. 27, pp. 27-36. [250]

Chapman, B. and L. Goldberg (2017), "Act-frequency signatures of the Big Five", *Personality and Individual Differences*, Vol. 116, pp. 201-205, http://dx.doi.org/10.1016/j.paid.2017.04.049. [264]

Cherian, J. and J. Jacob (2013), "Impact of self-efficacy on motivation and performance of employees", *International Journal of Business and Management*, Vol. 8/14, pp. 80-88. [158]

Cherney, I. (2008), "The effects of active learning on students' memories for course content", *Active Learning in Higher Education*, Vol. 9/2, pp. 152-171, http://dx.doi.org/10.1177/1469787408090841. [301]

Chernyshenko, O., M. Kankaraš and F. Drasgow (2018), "Social and emotional skills for student success and well-being: Conceptual framework for the OECD Study on Social and Emotional Skills", *OECD Education Working Papers*, No. 173, OECD Publishing, Paris, http://dx.doi.org/10.1787/db1d8e59-en. [3]

Cheung, F. et al. (2001), "Indigenous Chinese personality constructs - Is the five-factor model complete?", *Journal of Cross-Cultural Psychology*, Vol. 32, pp. 407-433, http://dx.doi.org/10.1177/0022022101032004003. [271]

Clark, L. and D. Watson (2008), "Temperament: An organizing paradigm for trait psychology", in John, O. and R. Robins (eds.), *Handbook of Personality: Theory and Research*, 3rd ed., Guilford Press, New York, NY. [126]

Cohen, S. (2004), "Social relationships and health", *American Psychologist*, Vol. 59, pp. 676-684. [179]

Collins, W. et al. (2000), "Contemporary research on parenting: The care for nature and nurture", *American Psychologist*, Vol. 55, pp. 218-232. [208]

Conn, S. and M. Reike (eds.) (1994), *The 16PF fifth edition technical manual*, Institute for Personality and Ability Testing, Champaign. [91]

Connelly, B. and D. Ones (2010), "An other perspective on personality: Meta-analytic integration of observers' accuracy and predictive validity", *Psychological Bulletin*, Vol. 136/6, pp. 1092-1122, http://dx.doi.org/10.1037/a0021212. [238]

Conner, M. (1996), *Predicting Health Behaviour: Research and Practice with Social Cognition Models*, Open University Press, Maidenhead. [163]

Cooksey, E., H. Joshi and G. Verropoulou (2009), "Does mothers' employment affect children's development? Evidence from the children of the British 1970 Birth Cohort and the American NLSY79", *Longitudinal and Life Course Studies*, Vol. 1/1, pp. 95-115. [169]

Costa, P., R. McCrae and D. Dye (1991), "Facet scales for agreeableness and conscientiousness: A revision of the NEO personality inventory", *Personality and Individual Differences*, Vol. 12/9, pp. 887-898. [90]

Covay, E. and W. Carbonaro (2010), "After the bell: Participation in extracurricular activities, classroom behavior, and academic achievement", *Sociology of Education*, Vol. 83/1, pp. 20-45. [171]

Cunha, F. and J. Heckman (2007), "The technology of skill formation", *American Economic Review*, Vol. 97/2, pp. 31-47. [12]

Cunha, F., J. Heckman and S. Schennach (2010), "Estimating the technology of cognitive and noncognitive skill formation", *Econometrica*, Vol. 78/3, pp. 883-931. [13]

Cunha, F. and J. Heckmann (2008), "Formulating, identifying and estimating the technology of cognition and noncognitive skill formation", *Journal of Human Resources*, Vol. 43/4, pp. 738-782. [105]

Daly, M. et al. (2015), "Childhood self-control and unemployment throughout the life span: Evidence from two British cohort studies", *Psychological Science*, Vol. 26, pp. 709-723. [122]

Daniel, E. et al. (2014), "Developmental relations between sympathy, moral emotion attributions, moral reasoning, and social justice values from childhood to early adolescence", *Journal of Adolescence*, Vol. 37, pp. 1201-1214. [142]

De Fruyt, F. and B. De Clercq (2014), "Antecedents of personality disorder in childhood and adolescence: Toward an integrative developmental model", *Annual Review of Clinical Psychology*, Vol. 10, pp. 449-476. [45]

De Fruyt, F., B. Wille and O. John (2015), "Employability in the 21st Century: Complex (Interactive) Problem Solving and Other Essential Skills", *Industrial and Organizational Psychology*, Vol. 8/2, pp. 276-281. [14]

de Haan, A. et al. (2017), "Long-term developmental changes in children's lower-order Big Five personality facets", *Journal of Personality*. [123]

De Kemp, R. et al. (2007), "Early adolescent empathy, parental support, and antisocial behavior", *The Journal of Genetic Psychology*, Vol. 165, pp. 5-18. [217]

Department of Education and Skills (2012), *A Framework for Junior Cycle*, Department of Education and Skills, Dublin. [278]

DeYoung, C., L. Quilty and J. Peterson (2007), "Between facets and domains: 10 aspects of the Big Five", *Journal of Personality and Social Psychology*, Vol. 93, pp. 880-896. [84]

Digman, J. (1990), "Personality structure: Emergence of the five-factor model", *Annual Review of Psychology*, Vol. 41, pp. 417-440. [37]

Drasgow, F. (1984), "Scrutinizing psychological tests: Measurement equivalence and equivalent relations with external variables are the central issues", *Psychological Bulletin*, Vol. 95, pp. 134-135. [299]

Drasgow, F. and C. Hulin (1987), "Cross-cultural measurement", *Interamerican Journal of Psychology*, Vol. 21, pp. 1-24. [298]

Drasgow, F. et al. (2012), *Development of the Tailored Adaptive Personality Assessment System (TAPAS) to support Army selection and classification decisions*, (Tech. Rep. No. 1311), US Army Research Institute for the Behavioral and Social Sciences, Arlington. [89]

Duckworth, A., E. Tsukayama and H. May (2010), "Establishing causality using longitudinal hierarchical linear modeling: An illustration predicting achievement from self-control", *Social Psychological and Personality Science*, Vol. 1, pp. 311-317, http://dx.doi.org/10.1177/1948550609359707. [120]

Duckworth, A. and D. Yeager (2015), "Measurement matters: Assessing personal qualities other than cognitive ability for educational purposes", *Educational Researcher*, Vol. 44/4, pp. 237-251. [20]

Durlak, J. et al. (2011), "The impact of enhancing students' social and emotional learning: A meta-analysis of school-based universal interventions", *Child Development*, Vol. 82, pp. 405-432, http://dx.doi.org/10.1111/j.1467-8624.2010.01564.x. [124]

Dweck, C. (2006), *Mindset: The new psychology of success*, Random House, New York. [221]

Edwards, B. and D. Woehr (2007), "An examination and evaluation of frequency-based personality measurement", *Personality and Individual Differences*, Vol. 43, pp. 803-814. [292]

Edwards, M. et al. (2010), "A reexamination of the factor structure for the Center for Epidemiologic Studies Depression Scale: Is a one-factor model plausible?", *Psychological Assessment*, Vol. 22, pp. 711-715. [253]

Egan, M. et al. (2017), "Adolescent conscientiousness predicts lower lifetime unemployment", *Journal of Applied Psychology*, Vol. 102/4, pp. 700-709. [75]

El Nokali, N., H. Bachman and E. Votruba-Drzal (2010), "Parent involvement and children's academic and social development in elementary school", *Child Development*, Vol. 81/3, pp. 988-1005, http://dx.doi.org/10.1111/j.1467-8624.2010.01447.x. [216]

Elias, M. et al. (1997), *Promoting Social and Emotional Learning: Guidelines for Educators*, Association for Supervision and Curriculum Development, Alexandria. [224]

Fergusson, D., J. Boden and L. Horwood (2013), "Childhood self-control and adult outcomes: Results from a 30-year longitudinal study", *Journal of the American Academy of Child and Adolescent Psychiatry*, Vol. 52, pp. 709-717. [121]

Fernández-Alonso, R. et al. (2017), "Parental involvement and academic performance: Less control and more communication", *Psicothema*, Vol. 29/4, pp. 453-461. [215]

Flavell, J. (1979), "Metacognition and cognitive monitoring: A new area of cognitive-developmental inquiry", *American Psychologist*, Vol. 34/10, pp. 906-911. [280]

Flett, G. et al. (2016), "The child-adolescent perfectionism scale: Development, psychometric properties, and associations with stress, distress, and psychiatric symptoms", *Journal of Psychoeducational Assessment*, Vol. 34/7, pp. 634-652. [183]

Fontaine, N. et al. (2011), "Predictors and outcomes of joint trajectories of callous-unemotional traits and conduct problems in childhood", *Journal of Abnormal Psychology*, Vol. 120, pp. 730-742. [141]

Gaylord, N., K. Kitzmann and J. Coleman (2003), "Parents' and children's perceptions of parental behavior: Associations with children's psychosocial adjustment in the classroom", *Parenting*, Vol. 3/1, pp. 23-47, http://dx.doi.org/10.1207/S15327922PAR0301_02. [214]

Goldberg, L. (1999), "A broad-bandwidth, public domain, personality inventory measuring the lower-level facets of several five-factor models", in Mervielde, I. et al. (eds.), *Personality Psychology in Europe*, Tilburg University Press, Tilburg. [96]

Goldberg, L. (1993), "The structure of phenotypic personality traits", *American Psychologist*, Vol. 48, pp. 26-34. [145]

Goldberg, L. (1982), "From ace to zombie: Some explorations in the language of personality", in Spielberger, C. and J. Butcher (eds.), *Advances in Personality Assessment*, Erlbaum, Hillsdale. [38]

Goldberg, L. and M. Carlson (2014), "Parents' relationship quality and children's behavior in stable married and cohabiting families", *Journal of Marriage and Family*, Vol. 76/4, pp. 762-777. [220]

Goodwin, R., D. Fergusson and L. Horwood (2004), "Early anxious/withdrawn behaviours predict later internalising disorders", *Journal of Child Psychology and Psychiatry*, Vol. 45, pp. 874-883. [129]

Gosling, S., P. Rentfrow and W. Swann (2003), "A very brief measure of the Big Five personality domains", *Journal of Research in Personality*, Vol. 37/6, pp. 504-528. [248]

Gough, H. (1987), *The California Psychological Inventory Administrator's Guide*, Consulting Psychologists Press, Palo Alto. [92]

Graziano, W. and N. Eisenberg (1997), "Agreebleness: A dimension of personality", in Hogan, R., J. Johnson and S. Briggs (eds.), *Handbook of Personality Psychology*, Academic Press, San Diego. [137]

Graziano, W. and R. Tobin (2002), "Agreebleness: Dimension of personality or social desirability artifact?", *Journal of Personality*, Vol. 2002, pp. 695-727. [136]

Grunspan, D., B. Wiggins and S. Goodreau (2014), "Understanding classrooms through social network analysis: A primer for social network analysis in education research", *Research Methods*, Vol. 13, pp. 167-178, http://dx.doi.org/10.1187/cbe.13-08-0162. [178]

Grusec, J. and M. Davidov (2010), "Integrating different perspectives on socialization theory and research: A domain-specific approach", *Child Development*, Vol. 81, pp. 687-709. [102]

Guerriero, S. (ed.) (2017), *Pedagogical Knowledge and the Changing Nature of the Teaching*, OECD Publishing, http://dx.doi.org/10.1787/9789264270695-en. [113]

Gutman, L. and I. Schoon (2013), "The impact of non-cognitive skills on outcomes for young people", *Education Endowment Foundation*, London, UK, https://educationendowmentfoundation.org.uk/public/files/Publications/EEF_Lit_Review_Non-CognitiveSkills.pdf. [56]

Halverson, C. et al. (2003), "Personality structure as derived from parental ratings of free descriptions of children: The inventory of child individual differences", *Journal of Personality*, Vol. 71, pp. 995-1026, http://dx.doi.org/10.1111/1467-6494.7106005. [52]

Hampson, S., O. John and L. Goldberg (1986), "Category breadth and hierarchical structure in personality: Studies of asymmetrics in judgments of trait implications", *Journal of Personality and Social Psychology*, Vol. 51, pp. 37-54. [43]

Harkness, S. and C. Super (2002), "Culture and parenting", in Bornstein, M. (ed.), *Handbook of Parenting*, Erlbaum, Mahwah. [288]

Heckman, J. and T. Kautz (2014), "Achievement tests and the role of character in American life", in Heckman, J., J. Humphries and T. Kautz (eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, University of Chicago Press, Chicago. [9]

Heckman, J. and T. Kautz (2014), "Fostering and measuring skills: Interventions that improve character and cognition", in Heckman, J., J. Humphries and T. Kautz (eds.), *The Myth of Achievement Tests: The GED and the Role of Character in American Life*, University of Chicago Press, Chicago. [8]

Heckman, J. and T. Kautz (2012), "Hard evidence on soft skills", *Labour Economics*, Vol. 19, pp. 451-464. [7]

Heckman, J., J. Stixrud and S. Urzua (2006), "The effects of cognitive and non-cognitive abilities on labor market outcomes and social behavior", *Journal of Labor Economics*, Vol. 24, pp. 411-482. [19]

Hegvik, R., S. McDevitt and W. Carey (1982), "The middle childhood temperament questionnaire", *Journal of Developmental Behavior Pediatrics*, Vol. 3, pp. 197-200. [85]

Heine, S., E. Buchtel and A. Norenzayan (2008), "What do cross-national comparisons of personality traits tell us? The case of conscientiousness", *Psychological Science*, Vol. 19/4, pp. 309-313, http://dx.doi.org/10.1111/j.1467-9280.2008.02085.x. [240]

He, J. et al. (2014), "Response styles and personality traits: A multilevel analysis", *Journal of Cross-Cultural Psychology*, Vol. 45/7, pp. 1028-1045. [249]

He, J., J. Buchholz and E. Klieme (2017), "Effects of anchoring vignettes on comparability and predictive validity of student self-reports in 64 cultures", *Journal of Cross-Cultural Psychology*, Vol. 48/3, pp. 319-334, http://dx.doi.org/10.1177/0022022116687395. [258]

Helson, R. et al. (2002), "The growing evidence of personality change in adulthood: Findings from research with personality inventories", *Journal of Research in Personality*, Vol. 36/4, pp. 287-306. [57]

Herbert, J. and D. Stipek (2005), "The emergency of gender differences in children's perceptions of their academic competence", *Journal of Applied Developmental Psychology*, Vol. 26/3, pp. 276-295. [162]

Hewitt, P. et al. (2002), "Perfectionism in children: Associations with depression, anxiety, and anger", *Personality and Individuality Differences*, Vol. 32/6, pp. 1049-1061, http://dx.doi.org/10.1016/S0191-8869(01)00109-X. [186]

Hoellwarth, C. and M. Moelter (2011), "The implications of a robust curriculum in introductory mechanics", *American Journal of Physics*, Vol. 79/5, p. 540, http://dx.doi.org/10.1119/1.3557069. [230]

Hogan, R. and J. Hogan (1992), *Hogan Personality Inventory Manual*, Hogan Assessment Systems, Tulsa. [95]

Holden, G. (2010), *Parenting: A Dynamic Perspective*, Sage, California. [203]

Horn, J. and J. McArdle (1992), "A practical and theoretical guide to measurement invariance", *Experimental Aging Research*, Vol. 18, pp. 117-144. [109]

Hough, L. (2003), "Emerging trends and needs in personality research and practice: Beyond main effects", in Barrick, M. and A. Ryan (eds.), *Personality and Work: Reconsidering the Role of Personality in Organizations*, Jossey-Bass, San Francisco. [143]

Hui, C. and H. Trandis (1985), "Measurement in cross-cultural psychology: A review and comparison of strategies", *Journal of Cross-Cultural Psychology*, Vol. 16, pp. 131-152. [268]

Irwin, D. et al. (2009), "Cognitive interviewing methodology in the development of a pediatric item bank: A patient reported outcomes measurement information system (PROMIS) study", *Health and Quality of Life Outcomes*, Vol. 7/3, http://dx.doi.org/10.1186/1477-7525-7-3. [294]

ITC (2017), "ITC Guidelines for translating and adapting tests", *International Journal of Testing*, Vol. 18/2, pp. 101-134, http://dx.doi.org/10.1080/15305058.2017.1398166. [273]

Jackson, C. (2013), "Non-cognitive ability, test scores, and teacher quality: Evidence from 9th grade teachers in North Carolina", *NBER Working Paper No. 18624*. [170]

Jackson, D. (1994), *Jackson Personality Inventory-Revised Manual*, Sigma Assessment Systems, Port Huron. [94]

Jackson, J. et al. (2012), "Can an old dog learn (and want to experience) new tricks? Cognitive training increases openness to experience in older adults", *Psychology and Aging*, Vol. 27, pp. 286-292. [79]

Jakobsen, L., L. Horwood and D. Fergusson (2012), "Childhood anxiety/withdrawal, adolescent parent-child attachment and later risk of depression and anxiety disorder", *Journal of Child and Family Studies*, Vol. 21, pp. 303-310. [130]

Janssens, A. et al. (2015), "Parents' and adolescents perspectives on parenting: Evaluating conceptual structure, measurement invariance, and criterion validity", *Assessment*, Vol. 22/4, pp. 473-489. [212]

Jasso, G. (2011), "Migration and Stratification", *Social Science Research*, Vol. 40, pp. 1292-1336. [201]

Jennings, P. and M. Greenberg (2009), "The prosocial classroom: Teacher social and emotional competence in relation to student and classroom outcomes", *Review of Educational Research*, Vol. 79/1, pp. 491-525, http://dx.doi.org/10.3102/0034654308325693. [231]

Jensen-Campbell, L. et al. (2002), "Agreebleness, extraversion, and peer relations in early adolescence: Winning friends and deflecting aggression", *Journal of Research in Personality*, Vol. 36, pp. 224-251. [139]

John, O. and F. De Fruyt (2015), *Framework for the Longitudinal Study of Social and Emotional Skills in Cities*, OECD Publishing, Paris. [34]

John, O. and S. Mauskopf (2015), *Self-reported socio-emotional qualities: Five factors for 21st century skills?*. [24]

John, O., L. Naumann and C. Soto (2008), "Paradigm shift to the integrative Big Five trait taxonomy: History, measurement, and conceptual issues", in John, O., R. Robins and L. Pervin (eds.), *Handbook of Personality: Theory and Research*, Guildford Press, New York. [27]

John, O. and S. Srivastava (1999), "The Big Five trait taxonomy: History, measurement, and theoretical perspectives", in Pervin, L. and O. John (eds.), *Handbook of Personality: Theory and Research*, Guilford, New York. [286]

Johnson, T. (1998), "Approaches to equivalence in cross-cultural and cross-national survey", *ZUMA-Nachrichten Spezial*, Vol. 3, pp. 2-40. [279]

Judge, T. et al. (2002), "Personality and leadership: A qualitative and quantitative review", *Journal of Applied Psychology*, Vol. 87, pp. 765-780. [67]

Judge, T. et al. (2007), "Self-efficacy and work-related performance: The integral role of individual differences", *Journal of Applied Psychology*, Vol. 92/1, pp. 107-127. [160]

Judge, T., B. Livingston and C. Hurst (2012), "Do nice guys - and gals - really finish last? The joint effects of sex and agreebleness on income", *Journal of Personality and Social Psychology*, Vol. 102, pp. 390-407. [68]

Judge, T. et al. (2013), "Hierarchical representations of the five-factor model of personality in predicting job performance: Integrating three organizing frameworks with two theoretical perspectives", *Journal of Applied Psychology*, Vol. 98, pp. 875-925, http://dx.doi.org/10.1037/a0033901. [132]

Kankaraš, M. (2017), "Personality matters: Relevance and assessment of personality characteristics", *OECD Education Working Papers*, OECD Publishing, Paris, http://dx.doi.org/10.1787/8a294376-en. [5]

Kankaraš, M. (2009), "Implicit Theories of Creativity: A Cross-Cultural Study", *Psihologija*, Vol. 42/2, pp. 187-202, http://dx.doi.org/10.2298/PSI0902187K. [303]

Kankaraš, M. (2004), "Metakognicija - nova kognitivna paradigm (Metacognition - a new cognitive paradigm)", *Psihilogija*, Vol. 37/2, pp. 149-161. [281]

Kankaraš, M., E. Feron and R. Renbarger (2019), *Assessing students' social and emotional skills through triangulation of assessment methods*, OECD Publishing. [55]

Kankaraš, M. and G. Moors (2011), "Measurement equivalence and extreme response bias in the comparison of attitudes across Europe: A multigroup latent-class factor approach", *Methodology*, Vol. 7, pp. 68-80. [112]

Kankaraš, M. and G. Moors (2010), "Researching measurement equivalence in cross-cultural studies", *Psihologija*, Vol. 43/2, pp. 121-136. [108]

Kapteyn, A., J. Smith and A. Van Soest (2007), "Vignettes and self-reports of work disability in the US and the Netherlands", *American Economic Review*, Vol. 97, pp. 461–73. [260]

Kautz, T. et al. (2014), "Fostering and measuring skills: Improving cognitive and non-cognitive skills to promote lifetime success", *OECD Education Working Papers*, OECD Publishing, Paris, http://dx.doi.org/10.1787/5jxsr7vr78f7-en. [6]

Kautz, T. and W. Zanoni (2014), "Measuring and fostering non-cognitive skills in adolescents: Evidence from Chicago public schools and the OneGoal program", Unpublished manuscript, University of Chicago, Department of Economics. [10]

Kerr, M. et al. (2003), "Relationships with parents and peers in adolescence", in Lerner, R., M. Easterbrooks and J. Mistry (eds.), *Handbook of Pscyhology: Developmental Psychology*, Wiley, Hoboken. [103]

Kiernan, K. and M. Huerta (2008), "Economic deprivation, maternal depression, parenting and children's cognitive and emotional development in early childhood", *British Journal of Sociology*, Vol. 59/4, pp. 783-806. [166]

Knoll, L. et al. (2015), "Social influence on risk perception during adolescence", *Psychological Science*, Vol. 26/5, pp. 583-592, http://dx.doi.org/10.1177/0956797615569578. [195]

Kotchick, B. and R. Forehand (2002), "Putting parenting in perspective: A discussion of the contextual factors that shape parenting practices", *Journal of Child and Family Studies*, Vol. 11, pp. 255-269. [289]

Krasner, M. et al. (2009), "Association of an educational program in mindful communication with burnout, empathy, and attitudes among primary care physicians", *Journal of the American Medical Association*, Vol. 302, pp. 1284-1293. [77]

Kristensen, N. and E. Johansson (2008), "New Evidence on cross-country differences in job satisfaction using anchoring vignettes", *Labour Economics*, Vol. 15, pp. 96-117. [262]

Krosnick, J. (1999), "Survey research", *Annual Review of Psychology*, Vol. 50, pp. 537-567. [239]

Kuhn, D. (1999), "A Developmental Model of Critical Thinking", *Educational Researcher*, Vol. 28/2, pp. 16-46, http://dx.doi.org/10.3102/0013189X028002016. [284]

Kyllonen, P. (2012), "The importance of higher education and the role of noncognitive attributes in college success", *Pensamiento Educativo. Revista de Investigación Educacional Latinoamericana*, Vol. 49/2, pp. 84-100, http://dx.doi.org/10.7764/PEL.49.2.2012.7. [263]

Kyllonen, P. and J. Bertling (2013), "Innovating questionnaire assessment methods to increase cross-country comparability", in Rutkowski, L., M. von Davier and D. Rutkowski (eds.), *Handbook of International Large-Scale Assessment: Background, Technical Issues, and Methods of Data Analysis*, Chapman and Hall, New York. [257]

Lapsley, D. and D. Yeager (2012), "Moral character education", in Reynolds, W., G. Miller and L. Weiner (eds.), *Handbook of Psychology: Volume 7 Educational Psychology*, John Wiley and Sons Inc, New Jersey. [23]

Lee, K. and M. Ashton (2004), "Psychometric properties of the HEXACO personality inventory", *Multivariate Behavioral Research*, Vol. 39, pp. 329-358. [88]

Lee, R., M. Robbins and B. Steven (1995), "Measuring belongingness: The social connectedness and the social assurance scales", *Journal of Counseling Psychology*, Vol. 42/2, pp. 232-241, http://dx.doi.org/10.1037/0022-0167.42.2.232. [189]

Legree, P. et al. (2014), "Identifying the leaders of tomorrow: Validating predictors of leader performance", *Military Psychology*, Vol. 26, pp. 292-309. [133]

LePine, J. and L. Van Dyne (1998), "Predicting voice behavior in work groups", *Journal of Applied Psychology*, Vol. 83, pp. 853-868. [138]

Lila, M. and E. Grecia (2005), "Determinants of parental acceptance-rejection", *Psicothema*, Vol. 17, pp. 107-111. [291]

Lindorff, M. (2000), "Is it better to perceive than receive? Social support, stress and strain for managers", *Psychology, Health and Medicine*, Vol. 5, pp. 271-286. [180]

Lindqvist, E. and R. Vestman (2011), "The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment", *American Economic Journal: Applied Economics*, Vol. 3/1, pp. 101-128, http://dx.doi.org/10.1257/app.3.1.101. [242]

Lipnevich, A., F. Preckel and R. Roberts (2017), *Psychosocial skills and school systems in the 21st century: Theory, research, and practice*, Springer International Publishing, Switzerland. [36]

Lippman, L. et al. (2014), *Flourishing Children: Defining and Testing Indicators of Positive Development*, Springer. [197]

Lippman, L. et al. (2015), "Workforce connections: Key "soft skills" that foster youth workforce success: Toward a consensus across fields", *Child Trends Publication*, Vol. 24. [17]

Lounsbury, J. et al. (2004), "An investigation of personality traits in relation to adolescent school absenteeism", *Journal of Youth and Adolescence*, Vol. 33, pp. 457-466, http://dx.doi.org/10.1023/B:JOYO.0000037637.20329.97. [64]

Lozano, L., I. Valor-Segura and L. Lozano (2015), "Could a perfectionism context produce unhappy children?", *Personality and Individual Differences*, Vol. 80, pp. 12-17. [184]

Lunenberg, F. (2011), "Expectancy theory of motivation: Motivating by altering expectations", *International Journal of Business Administration*, Vol. 15/1, pp. 1-6. [157]

MacCann, C. et al. (2010), "Emotional intelligence and the eye of the beholder: Comparing self- and parent-rated situational judgments in adolescents", *Journal of Research in Personality*, Vol. 44/5, pp. 673-676, http://dx.doi.org/10.1016/j.jrp.2010.08.009. [244]

Manrique Millones, D., P. Ghesquière and K. Van Leeuwen (2014), "Evaluation of a parental behavior scale in a Peruvian context", *Journal of Child and Family Studies*, Vol. 23/5, pp. 885-894, http://dx.doi.org/10.1007/s10826-013-9744-z. [211]

Markon, K. (2009), "Hierarchies in the structure of personality traits", *Social and Personality Psychology Compass*, Vol. 3, pp. 812-826. [26]

Martocchio, J. and T. Judge (1997), "Relationship between conscientiousness and learning in employee training: Mediating influences of self-deception and self-efficacy", *Journal of Applied Psychology*, Vol. 82/5, pp. 764-773. [155]

Maslach, C., R. Santee and C. Wade (1987), "Individuation, gender role, and dissent: Personality mediators of situational forces", *Journal of Personality and Social Psychology*, Vol. 53/6, pp. 1088-1093, http://dx.doi.org/10.1037/0022-3514.53.6.1088. [175]

Masten, A. and A. Tellegen (2012), "Resilience in developmental psychopathology: Contributions of the Project Competence Longitudinal Study", *Development and Psychopathology*, Vol. 24, pp. 345-361. [134]

McCrae, R. (1996), "Social consequences of experiential openness", *Psychological Bulletin*, Vol. 120, pp. 323-337. [146]

McCrae, R. and P. Costa Jr. (1997), "Personality trait structure as a human universal", *American Psychologist*, Vol. 52, pp. 509-516. [42]

McCrae, R. and P. Costa Jr (1987), "Validation of the five-factor model of personality across instruments and observers", *Journal of Personality and Social Psychology*, Vol. 52, pp. 81-90. [39]

McCrae, R. and A. Terracciano (2005), "Universal features of personality traits from the observer's perspective: Data from 50 cultures", *Journal of Personality and Social Psychology*, Vol. 88, pp. 547-561. [28]

McPherson, M., L. Smith-Lovin and M. Brashears (2006), "Social isolation in America: Changes in core discussion networks over two decades", *American Sociological Review*, Vol. 71/3, pp. 353-375, http://dx.doi.org/10.1177/000312240607100301. [188]

Measelle, J. et al. (2005), "Can children provide coherent, stable, and valid self-reports on the Big Five dimensions? A longitudinal study from ages 5 to 7", *Journal of Personality and Social Psychology*, Vol. 89, pp. 90-106. [46]

Mellor, D. and K. Moore (2013), "The use of Likert scales with children", *Journal of Pediatric Psychology*, Vol. 39, pp. 369-379. [251]

Mershon, B. and R. Gorsuch (1988), "Number of factors in the personality sphere: Does increase in factors increase predictability of real-life criteria?", *Journal of Personality and Social Psychology*, Vol. 55, pp. 675-680. [30]

Mervielde, I. and F. De Fruyt (1999), "Construction of the hierarchical personality inventory for children (HiPIC)", in Mervielde, I. et al. (eds.), *Personality Psychology in Europe, Proceedings of the Eighth European Conference on Personality Psychology*, Tilburg University Press, Tilburg. [51]

Mervielde, I., F. De Fruyt and B. De Clercq (2009), *Hiërarchische Persoonlijkheidsvragenlijst voor Kinderen [Hierarchical Personal Inventory for Children]: Handleiding*, Hogrefe Publishers, Amsterdam. [86]

Meunier, J. and I. Roskam (2007), "Psychometric properties of a parental childrearing behavior scale for French-speaking parents, children, and adolescents", *European Journal of Psychological Assessments*, Vol. 23, pp. 113-124. [210]

Minuchin, P. (1988), "Relationships within the family: A systems perspective on development", in Hinde, R. and J. Stevenson-Hinde (eds.), *Relationships Within Families: Mutual Influences*, Oxford University Press, New York. [218]

Moffitt, T. et al. (2011), "A gradient of childhood self-control predicts health, wealth, and public safety", *Proceedings of the National Academy of Sciences*, Vol. 108, pp. 2693-2698. [71]

Murphy, N. and J. Hall (2011), "Intelligence and interpersonal sensitivity: A meta-analysis", *Intelligence*, Vol. 39/1, pp. 54-63. [21]

National Academy of Sciences (2012), *Education for Life and Work: Developing Transferable Knowledge and Skills in the 21st Century*, National Academic Press, Washington. [16]

National Youth Policy Institution (2009), *Introduction to Creative Activities*, Ministry of Education, Science and Technology Notice 41, National Youth Policy Institution, Seoul, Korea. [275]

Neal, A. and S. Collas (2000), *Intimacy and Alienation: Forms of Estrangement in Female/Male Relationships*, Psychology Press. [187]

Noftle, E. and R. Robins (2007), "Personality predictors of academic outcomes: Big Five correlates of GPA and SAT scores", *Journal of Personality and Social Psychology*, Vol. 93, pp. 116-130, http://dx.doi.org/10.1037/0022-3514.93.1.116. [62]

Norman, W. (1963), "Toward an adequate taxonomy of personality attributes: Replicated factor structure in peer nomination personality ratings", *Journal of Abnormal and Social Psychology*, Vol. 66, pp. 574-583. [40]

O'Brien, M. (2005), "Studying individual and family development: Linking theory and research", *Journal of Marriage and Family*, Vol. 67/4, pp. 880-890. [219]

Obschonka, M., R. Silbereisen and E. Schmitt-Rodermund (2012), "Explaining entrepreneurial behavior: Dispositional personality traits, growth of personal entrepreneurial resources, and business idea generation", *Career Development Quarterly*, Vol. 60, pp. 178-190, http://dx.doi.org/10.1002/j.2161-0045.2012.00015.x. [135]

OECD (2018), "PISA: Preparing our youth for an inclusive and sustainable world: The OECD PISA global competence framework", OECD Publishing, Paris, http://www.oecd.org/pisa/Handbook-PISA-2018-Global-Competence.pdf. [22]

OECD (2018), *Teaching for the Future: Effective Classroom Practices To Transform Education*, OECD Publishing, https://doi.org/10.1787/9789264293243-en. [225]

OECD (2017), *PISA 2015 Results (Volume III): Students' Well-Being*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264273856-en. [1]

OECD (2015), *Skills for Social Progress: The Power of Social and Emotional Skills*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264226159-en. [2]

OECD (2013), *How the quality of the learning environment is shaped*, OECD Publishing, http://dx.doi.org/10.1787/9789264201156-en. [232]

OECD (2013), *PISA 2012 Results: Ready to Learn (Volume III): Students' Engagement, Drive and Self-Beliefs*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264201170-en. [106]

OECD (Forthcoming), "Study on Social and Emotional Skills Technical Standards", *OECD Working Papers*, This will be updated with authors before publishing. [274]

Patcher, L. and T. Dumont-Mathieu (2004), "Parenting in culturally divergent settings", in Hoghughi, M. and N. Long (eds.), *Handbook of Parenthood, Theory and Research for Practice*, Sage, London. [290]

Paunonen, S. (1998), "Hierarchical organization of personality and prediction of behavior", *Journal of Personality and Social Psychology*, Vol. 74, pp. 538-556. [31]

Paunonen, S. and M. Ashton (2001), "Big Five factors and facets and the prediction of behavior", *Journal of Personality and Psychology*, Vol. 81, pp. 524-539. [32]

Paunonen, S. et al. (1996), "The structure of personality in six cultures", *Journal of Cross-Cultural Psychology*, Vol. 27, pp. 339-353. [110]

Piedmont, R. (2001), "Cracking the plaster cast: Big Five personality change during intensive outpatient counseling", *Journal of Research in Personality*, Vol. 35, pp. 500-520. [78]

Pintrich, P. and E. De Groot (1990), "Motivational and Self-Regulated Learning Components of Classroom Academic Performance", *Journal of Educational Psychology*, Vol. 82/1, pp. 33-40, http://dx.doi.org/10.1037/0022-0663.82.1.33. [283]

Plomin, R. and D. Daniels (2011), "Why are children in the same family so different from one another?", *International Journal of Epidemiology*, Vol. 40/1, pp. 563-582, http://dx.doi.org/10.1093/ije/dyq148. [177]

Plomin, R., Y. Kovas and C. Haworth (2007), "Generalist genes: Genetic links between brain, mind, and education", *International Mind, Brain, and Education*, Vol. 1/1, pp. 11-19, http://dx.doi.org/10.1111/j.1751-228X.2007.00002.x. [176]

Prag, P., M. Mills and R. Wittek (2016), "Subjective socioeconomic status and health in cross-national comparison", *Social Science and Medicine*, Vol. 149, pp. 84-92. [199]

Primi, R., O. John and F. de Fruyt (2016), "Development of an inventory assessing social and emotional skills in Brazilian youth", *European Journal of Psychological Assessment*, Vol. 32, pp. 5-16. [25]

Rammstedt, B., L. Goldberg and I. Borg (2010), "The measurement equivalence of Big-Five factor markers for persons with different levels of education", *Journal of Research in Personality*, Vol. 44/1, pp. 53-61, http://dx.doi.org/10.1016/j.jrp.2009.10.005. [265]

Rammstedt, B. and C. Kemper (2011), "Measurement equivalence of the Big Five: Shedding further light on potential causes of the educational bias", *Journal of Research in Personality*, Vol. 45/1, pp. 121-125, http://dx.doi.org/10.1016/j.jrp.2010.11.006. [266]

Rauthmann, J. and J. Denissen (2011), "I often do it vs. I like doing it: Comparing a frequency- and valency-approach to extraversion", *Personality and Individual Differences*, Vol. 50, pp. 1283-1288. [293]

Ravens-Sieberer, U. et al. (2014), "The European KIDSCREEN approach to measure quality of life and well-being in children: Development, current applications, and future advances", *Quality of Life Research*, Vol. 23, pp. 791-803. [295]

Roberts, B. (2006), "Personality development and organizational behavior", in Staw, B. (ed.), *Research on Organizational Behavior*, Elsevier Science/JAI Press. [80]

Roberts, B. et al. (2005), "The structure of conscientiousness: An empirical investigation based on seven major personality questionnaires", *Personnel Psychology*, Vol. 58, pp. 103-139. [33]

Roberts, B. and W. DelVecchio (2000), "The rank-order consistency of personality traits from childhood to old age: A quantitative review of longitudinal studies", *Psychological Bulletin*, Vol. 126, pp. 3-25, http://dx.doi.org/10.1037/0033-2909.126.1.3. [99]

Roberts, B. et al. (2009), "Conscientiousness", in Leary, M. and R. Hoyle (eds.), *Handbook of Individual Differences in Social Behavior*, Guilford, New York. [117]

Roberts, B. et al. (2007), "The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting life outcomes", *Perspectives on Psychological Science*, Vol. 2/4, pp. 313-345, http://dx.doi.org/10.1111/j.1745-6916.2007.00047.x. [11]

Roberts, B., C. Lejuez and R. Krueger (2014), "What is conscientiousness and how can it be assessed?", *Developmental Psychology*, Vol. 50, pp. 1407-1425. [118]

Roberts, B. et al. (2017), "A systematic review of personality trait change through intervention", *Psychological Bulletin*, Vol. 143/2, pp. 117-141. [76]

Roberts, B., K. Walton and W. Viechtbauer (2006), "Patterns of mean-level change in personality traits across the life course: A meta-analysis of longitudinal studies", *Psychological Bulletin*, Vol. 132, pp. 1-25, http://dx.doi.org/10.1037/0033-2909.132.1.1. [58]

Robins, R., H. Hendin and K. Trzesniewski (2001), "Measuring global self-esteem: construct validation of a single-item measure and the Rosenberg self-esteem scale", *Personality and Social Psychology Bulletin*, Vol. 27/2, pp. 151-161. [246]

Rosander, P. and M. Backstrom (2014), "Personality traits measured at baseline can predict academic performance in upper secondary school three years later", *Scandinavian Journal of Psychology*, Vol. 55, pp. 611-618. [63]

Sackett, P. and P. Walmsley (2014), "Which personality attributes are most important in the workplace?", *Perspectives on Psychological Science*, Vol. 9, pp. 538-551, http://dx.doi.org/10.1177/1745691614543972. [65]

Saks, A. (1993), "Moderating and mediating effects of self-efficacy for the relationship between training and newcomer adjustment", *Academy of Management Proceedings*, Vol. 1, pp. 126-130. [154]

Salas, E. and A. Cannon-Bowers (2001), "The science of training: A decade of progress", *Annual Review of Psychology*, Vol. 52, pp. 471-499. [226]

Saloman, J., A. Tandon and C. Murray (2004), "Comparability of self-rated health: Cross-sectional multi-country survey using anchoring vignettes", *British Medical Journal*, Vol. 328, pp. 258–261. [261]

Saris, W. et al. (2010), "Comparing questions with agree/disagree response options to questions with item-specific response options", *Survey Research Methods*, Vol. 4, pp. 61-79. [252]

Sarzosa, M. and S. Urzua (2015), "Bullying among adolescents: The role of cognitive and non-cognitive skills", *NBER Working Paper No. w21631*, https://ssrn.com/abstract=2675909. [125]

Saucier, G. and F. Ostendorf (1999), "Hierarchical subcomponents of the Big Five personality factors: A cross-language replication", *Journal of Personality and Social Psychology*, Vol. 76/4, pp. 613-627. [82]

Saville, P. et al. (1984), *The Occupational Personality Questionnaire (OPQ)*, SHL, London. [97]

Schmitt, D. et al. (2007), "The geographic distribution of big five personality traits: Patterns and profiles and human self-description across 56 nations", *Journal of Cross-Cultural Psychology*, Vol. 38, pp. 173-212, http://dx.doi.org/10.1177/0022022106297299. [111]

Schoon, I. et al. (2015), "The impact of early life skills on later outcomes", *OECD Education Working Papers*, OECD Publishing, Paris. [98]

Segal, C. (2011), "Misbehavior, education, and labor market outcomes", *Journal of the European Economic Association*, Vol. 1, pp. 1-40. [243]

Shiner, R. (1998), "How shall we speak of children's personalities in middle childhood: A preliminary taxonomy", *Psychological Bulletin*, Vol. 124, pp. 308-332. [47]

Shiner, R. and A. Caspi (2003), "Personality differences in childhood and adolescence: Measurement, development, and consequences", *Journal of Child Psychology and Psychiatry*, Vol. 44, pp. 2-32. [48]

Shoda, Y., W. Mischel and P. Peake (1990), "Predicting adolescent cognitive and self-regulatory competencies from preschool delay of gratification: Identifying diagnostic conditions", *Developmental Psychology*, Vol. 26, pp. 978-986. [119]

Shuey, E. and M. Kankaraš (2018), "The Power and Promise of Early Learning", *OECD Education Working Paper* 186, https://doi.org/10.1787/f9b2e53f-en. [4]

Smith, M. et al. (2016), "Cultural similarities in perfectionism: Perfectionistic strivings and concerns generalize across Chinese and Canadian groups", *Measurement and Evaluation in Counseling and Development*, Vol. 49/1, pp. 63-76, http://dx.doi.org/10.1177/0748175615596785. [185]

Smyth, K., F. Goodman and C. Glenn (2006), "The full-frame approach: A new response to marginalized women left behind by specialized services", *The American Journal of Orthopsychiatry*, Vol. 76/4, pp. 489-502, http://dx.doi.org/10.1037/0002-9432.76.4.489. [191]

Socolar, R. (1997), "A classification scheme for discipline: Type, mode of administration, context", *Aggression and Violent Behavior*, Vol. 2, pp. 355-364. [206]

Solomon, M. (1993), "Transmission of cultural goals: Social network influences on infant socialization", in Demick, J., K. Bursik and R. DiBlaise (eds.), *Parental Development*, Erlbaum, Hillsdale. [287]

Soto, C. (2016), "The little six personality dimensions from early childhood to early adulthood: Mean-level age and gender differences in parents' reports", *Journal of Personality*, Vol. 84, pp. 409-422. [100]

Soto, C. and O. John (2017), "The next Big Five Inventory (BFI-2): Developing and assessing a hierarchical model with 15 facets to enhance bandwidth, fidelity, and predictive power", *Journal of Personality and Social Psychology*, Vol. 113/1, pp. 117-143. [87]

Soto, C. and O. John (2014), "Traits in transition: The structure of parent-reported personality traits from early childhood to early adulthood", *Journal of Personality*, Vol. 82, pp. 182-199. [107]

Soto, C. et al. (2011), "Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample", *Journal of Personality and Social Psychology*, Vol. 100, pp. 330-348, http://dx.doi.org/10.1037/a0021717. [104]

Soto, C. et al. (2008), "The developmental psychometrics of Big Five self-reports: Acquiescence, factor structure, coherence and differentiation from ages 10 to 20", *Journal of Personality and Social Psychology*, Vol. 94, pp. 718-737. [53]

Soto, C. and J. Tackett (2015), "Personality traits in childhood and adolescence: Structure, development, and outcomes", *Current Directions in Psychological Science*, Vol. 24, pp. 358-362. [101]

Specht, J., S. Schmukle and B. Egloff (2011), "Stability and change of personality across the life course: The impact of age and major life events on mean-level and rank-order stability of the Big Five", *Journal of Personality and Social Psychology*, Vol. 101, pp. 862-882. [59]

Spencer, R., A. Basualdo-Delmonico and T. Lewis (2011), "Working to make it work: The role of parents in the youth mentoring process", *Journal of Community Psychology*, Vol. 39/1, pp. 51-59. [192]

Spurk, D. and A. Abele (2010), "Who earns more and why? A multiple mediation model from personality to salary", *Journal of Business and Psychology*, Vol. 26, pp. 87-103. [140]

Srivastava, S. et al. (2003), "Development of personality in early and middle adulthood: Set like plaster or persistent change?", *Journal of Personality and Social Psychology*, Vol. 84, pp. 1041-1053. [60]

Stajkovic, A. and F. Luthans (1998), "Self-efficacy and work-related performance: A meta-analysis", *Psychological Bulletin*, Vol. 124/2, pp. 240-261. [156]

Standing Council on School Education and Early Childhood (2013), "National Safe Schools Framework", http://www.safeschoolshub.edu.au/documents/nationalsafeschoolsframework.pdf. [276]

Stark, S., O. Chernyshenko and F. Drasgow (2005), "An IRT approach to constructing and scoring pairwise preference items involving stimuli on different dimensions: The multi-unidimensional pairwise-preference model", *Applied Psychological Measurement*, Vol. 29, pp. 184-203. [297]

Steel, P., J. Schmidt and J. Shultz (2008), "Refining the relationship between personality and subjective well-being", *Psychological Bulletin*, Vol. 134, pp. 138-161, http://dx.doi.org/10.1037/0033-2909.134.1.138. [128]

Steinberg, L. (2005), "Cognitive and affective development in adolescence", *Trends in Cognitive Sciences*, Vol. 9/2, pp. 69-74, http://dx.doi.org/10.1016/j.tics.2004.12.005. [285]

Steinmayr, R., F. Dinger and B. Spinath (2012), "Motivation as a mediator of social disparities in academic achievement", *European Journal of Personality*, Vol. 26/3, pp. 335-349. [114]

Stoeber, J., K. Otto and C. Dalbert (2009), "Perfectionism and the Big Five: Conscientiousness predicts longitudinal increases in self-oriented perfectionism", *Personality and Individual Differences*, Vol. 47, pp. 363-368. [182]

Stormshak, E. et al. (1997), "Observed family interaction during clinical interviews: A comparison of families containing preschool boys with and without disruptive behavior", *Journal of Abnormal Child Psychology*, Vol. 25, pp. 345-357. [207]

Strickhouser, J., E. Zell and Z. Krizan (2017), "Does personality predict health and well-being? A metasynthesis", *Health Psychology*, Vol. 36/8, pp. 797-810, http://dx.doi.org/10.1037/hea0000475. [69]

Suárez-Álvarez, J., R. Fernández-Alonso and J. Muñiz (2014), "Self-concept, motivation, expectations, and socioeconomic level as predictors of academic performance in mathematics", *Learning and Individual Differences*, Vol. 30, pp. 118-123, http://dx.doi.org/10.1016/j.lindif.2013.10.019. [115]

Suárez-Álvarez, J. et al. (2018), "Using reserved items in Likert Scales: A questionable practice", *Psicothema*, Vol. 30/2, pp. 149-158, http://dx.doi.org/10.7334/psicothema2018.33. [256]

Svinicki, M. (2001), *EDP 398T College Teaching Methodology*, University of Texas. [229]

Tackett, J. (2006), "Evaluating models of the personality-psychopathology relationship in children and adolescents", *Clinical Psychology Review*, Vol. 26, pp. 584-599. [72]

Tackett, J. et al. (2008), "Personality in middle childhood: A hierarchical structure and longitudinal connections with personality in late adolescence", *Journal of Research in Personality*, Vol. 42, pp. 1456-1462. [49]

Tackett, J. et al. (2012), "The hierarchical structure of childhood personality in five countries: Continuity from early childhood to early adolescence", *Journal of Personality*, Vol. 80, pp. 847-879, http://dx.doi.org/10.1111/j.1467-6494.2011.00748.x. [50]

Tarrant, M. (2002), "Adolescent peer groups and social identity", *Social Development*, Vol. 11/1, pp. 110-123, http://dx.doi.org/10.1111/1467-9507.00189. [174]

Tauber, B., H. Wahl and J. Schroder (2016), "Personality and life satisfaction over 12 years: Contrasting mid- and late life", *The Journal of Gerontopsychology and Geriatric Psychiatry*, Vol. 29, pp. 37-48. [73]

Tellegen, A. (1982), *Manual for the Multidimensional Personality Questionnaire*, Unpublished manuscript. [93]

Thalmayer, A., G. Saucier and A. Eigenhuis (2011), "Comparative validity of brief to medium-length Big Five and Big Six personality questionnaires", *Psychological Assessment*, Vol. 23/4, pp. 995-1009. [247]

Thapa, A. et al. (2013), "A review of school climate research", *Review of Educational Research*, Vol. 83/3, pp. 357-385, http://dx.doi.org/10.3102/0034654313483907. [223]

Thoits, P. (2011), "Mechanisms linking social ties and support to physical and mental health", *Journal of Health and Social Behavior*, Vol. 52/5, pp. 145-161, http://dx.doi.org/10.1177/0022146510395592. [193]

Thomas, A. and S. Chess (1977), *Temperament and Development*, Brunner/Mazel, New York. [81]

Trilling, B. and C. Fadel (2009), *21st Century Skills: Learning for Life in Our Times*, Jossey-Bass, Francisco. [15]

Tucker-Drob, E. and K. Harden (2012), "Learning motivation mediates gene-by-socioeconomic status interaction on mathematics achievement in early childhood", *Learning and Individual Differences*, Vol. 22/1, pp. 37-45. [116]

Tupes, E. and R. Christal (1958), *Stability of Personality Trait Rating Factors Obtained Under Diverse Conditions*, Lackland Air Force Base, TX, US Air Force. [41]

Uleman, J. et al. (2000), "The relational self: Closeness to ingroups depends on who they are, culture, and the type of closeness", *Asian Journal of Social Psychology*, Vol. 3, pp. 1-17. [196]

US Department of Education (2005), "Character education ... Our shared responsibility", US Department of Education, http://www.ed.gov/admins/lead/character/brochure.html. [277]

van Bel, D. et al. (2009), "Social connectedness: Concept and measurement", *Ambient Intelligence and Smart Environments*, Vol. 2, pp. 67-74. [190]

Van de Vijver, F. and K. Leung (2001), "Personality in cultural context: methodological issues.", *Journal of personality*, Vol. 69/6, pp. 1007-31, http://dx.doi.org/10.1111/1467-6494.696173. [269]

Van de Vijver, F. and K. Leung (1997), *Methods and Data Analysis for Cross-Cultural Research*, Sage, Newbury Park. [272]

Van de Vijver, F. and N. Tanzer (2004), "Bias and equivalence in cross-cultural assessment: an overview", *European Review of Applied Psychology*, Vol. 54/2, pp. 119-135. [270]

Van Heel, M. et al. (2017), "Measuring parenting throughout adolescence: Measurement invariance across informants, mean-level and differential continuity", *Assessment*, http://dx.doi.org/10.1177/1073191116686827. [213]

Van Leeuwen, K. and A. Vermulst (2004), "Some psychometric properties of the Ghent Parental Behaviour Scale", *European Journal of Psychological Assessment*, Vol. 20, pp. 283-298. [209]

Varni, J. et al. (2015), "Item-level informant discrepencies between children and their parents on the PROMIS pediatric scales", *Quality of Life Research*, Vol. 24, pp. 1921-1937. [236]

von Davier, M. et al. (2017), "The effects of vignette scoring on reliability and validity of self-reports", *Applied Psychological Measurement*, Vol. 42/4, pp. 291-306, http://dx.doi.org/10.1177/0146621617730389. [259]

von Stumm, S., B. Hell and T. Chamorro-Premuzic (2011), "A hungry mind: Intellectual curiosity is the third pillar of academic performance", *Perspectives on Psychological Science*, Vol. 2011, pp. 574-588, http://dx.doi.org/10.1177/ 1745691611421204. [147]

WHO (1998), *Wellbeing measures in primary health care/the DepCare Project*, http://www.euro.who.int/__data/assets/pdf_file/0016/130750/E60246.pdf. [202]

Widiger, T. (2009), "Neuroticism", in Leary, M. and R. Hoyle (eds.), *Handbook of Individual Differences in Social Behavior*, Guilford, New York. [127]

Winner, E., T. Goldstein and S. Vincent-Lancrin (2013), *Art for Art's Sake? The Impact of Arts Education*, Centre for Educational Research and Innovation, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264180789-en. [173]

Wong, N., A. Rindfleisch and J. Burroughs (2003), "Do reverse-worded items confound measures in cross-cultural consumer research? The case of the Material Values Scale", *Journal of Consumer Research*, Vol. 30, pp. 72-91. [255]

Woods, C. (2006), "Careless responding to reverse-worded items: Implications for confirmatory factor analysis", *Journal of Psychopathology and Behavioral Assessment*, Vol. 28, pp. 189-194. [254]

Woo, S. et al. (2014), "Openness to experience: Its lower level structure measurement, and cross-cultural equivalence", *Journal of Personality Assessment*, Vol. 96/1, pp. 29-45, http://dx.doi.org/10.1080/00223891.2013.806328. [300]

Woo, S. et al. (2014), "Validity of six openness facets in predicting work behaviors: A meta-analysis", *Journal of Personality Assessment*, Vol. 96/1, pp. 76-86, http://dx.doi.org/10.1080/00223891.2013.806329. [148]

Yeager, D. and C. Dweck (2012), "Mindsets that promote resilience: When students believe that personal characteristics can be developed", *Educational Psychologists*, Vol. 47/2, pp. 302-314, http://dx.doi.org/10.1080/00461520.2012.722805. [222]

Zeldin, A. and F. Pajares (2000), "Against the odds: Self-efficacy beliefs of women in mathematical, scientific and technological careers", *American Educational Research Journal*, Vol. 37/1, pp. 215-246. [161]

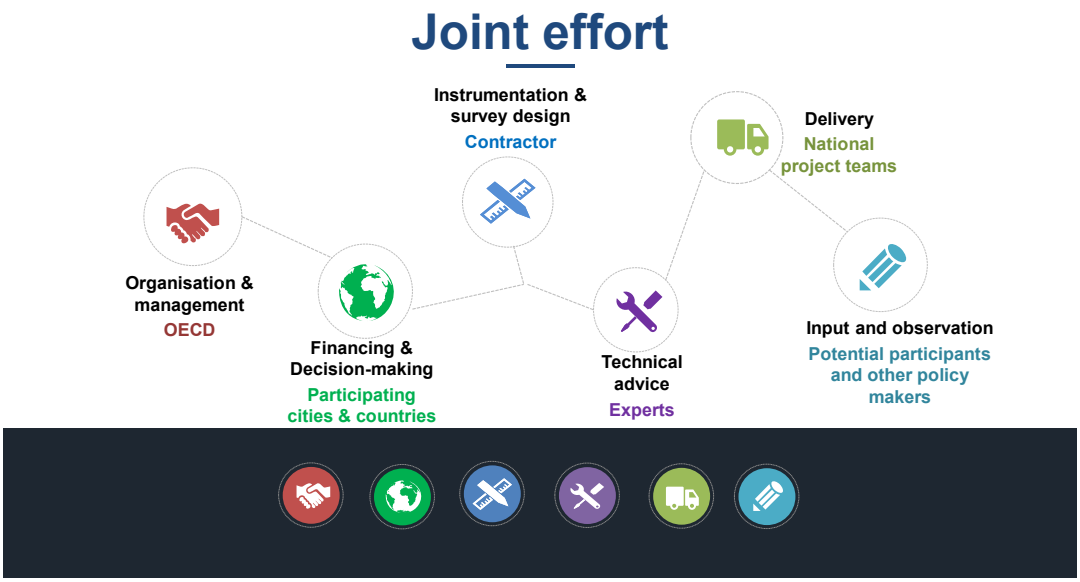
Zimmerman, B. (2002), "Becoming a self-regulated learner: An overview", *Theory into Practice*, Vol. 41/2, pp. 64-70. [282]

Zins, J. and M. Elias (2006), "Social and emotional learning", in Bear, G. and K. Minke (eds.), *Children's Needs*, National Association of School Psychologists. [235]

Annex

Project organisation and the main Study stakeholders

This section sets out the expected roles of the OECD Secretariat and other key players: the Informal Advisory Group (IAG), Study Project Managers (SPMs), the International Contractor (ACER) and the Technical Advisory Group (TAG).



The OECD Secretariat

The OECD Secretariat is responsible for the overall management of the Study. The Secretariat works collaboratively with participating cities and countries to ensure their priorities and interests are reflected in the design and implementation of the Study.

The OECD Secretariat leads or participates actively during the development of all instruments, protocols and procedures, documents and reports and approves all documents before public release.

The OECD Secretariat is also responsible for:

- actively engaging participating cities and countries in the development and implementation of the Study
- keeping the OECD's governance bodies regularly updated on progress and issues that arise
- overseeing the International Contractor, and managing and monitoring potential risks, issues and deviations from timelines
- providing a central point of contact for resolving any debates between the International Contractor and Study Project Managers over responsibilities,

workflow and timelines that have not been resolved through the processes of communication set up by the International Contractor

- monitoring the budgets and milestones of the International Contractor and resolving budgetary or contractual issues
- establishing and maintaining an archive of all project resources, documents, materials and databases
- providing additional support to SPMs, obtaining regular feedback from SPMs, and dealing with queries or problems that cannot be resolved by the International Contractor.

Informal Advisory Group

Representatives from cities and countries participating in the Study and other stakeholders interested in the development of the project participate in the Informal Advisory Group. Among the participants, there are government representatives of non-participating countries/jurisdictions, leading academics, representatives of research and policy institutes, NGOs and foundations in the field of social and emotional learning and education in general. Participants provide advice and other input to the OECD Secretariat on the Study as they develop from an individual city or country perspective. Two face-to-face two-day meetings of the Informal Advisory Group are held each year, in addition to shorter webinars and conference calls.

International Contractor

The OECD Secretariat appointed the Australian Council for Educational Research (ACER) as the International Contractor to implement the Study. As part of its management role, the International Contractor maintains an overall project plan for each stage of the Study, including implementation timelines for participating cities and countries.

ACER works closely with three other organisations on this project:

- cApStAn - Linguistic Quality Control CTM in Brussels, Belgium manages the international translation verification procedures.
- SoNET in Melbourne, Australia is responsible for the test delivery platform.
- Béatrice Halleux from HallStat SPRL in Liège Area, Belgium acts as the translation referee.

The International Contractor is responsible for implementing sampling requirements, manuals and other tools, training Study Project Managers in assessment administration and for analysing the findings.

The International Contractor is responsible for supporting and overseeing the preparations and implementation of the assessment in participating cities/countries – from the first phases of the translation, adaptation and field testing, to implementing the Study. The International Contractor is required to establish tools and procedures for effectively communicating with Study Project Managers, for collecting and collating regular progress updates from Study Project Managers, and for keeping the OECD Secretariat regularly updated on progress and issues that arise.

The International Contractor is the main point of contact for Study Project Managers. The International Contractor is required to specify and implement procedures that promote

excellent communication with Study Project Managers. The International Contractor are expected to maintain a communication portal, where Study Project Managers can communicate about tasks, and where they can find manuals, guidance and regularly updated information on the Study's progress.

Technical Advisory Group

A Technical Advisory Group (TAG) of international experts provides advice to the OECD Secretariat and the International Contractor on both substantive and methodological aspects of the Study. They helped the OECD Secretariat to design the study, setup instrument development process and outline all of the key aspects of Study's assessment framework. The Technical Advisory Group is made up of leading international experts in a range of relevant fields, including psychological assessment, skill development, social psychology, cross-cultural comparability, survey methodology, etc.

Following international experts were members of the Technical Advisory Group of the SSES:

Prof. Dr. **Filip De Fruyt**, Ghent University, Belgium

Dr. **Pat Kyllonen**, Senior Research Director, Educational Testing Service, United States

Prof. Dr. **Bruno Losito**, Università Degli Studi Roma Tre, Rome, Italy

Dr. **Kristin Moore**, Senior Program Area Director and Senior Scholar, Child Trends, United States

Dr. **Ricardo Primi**, Universidade São Francisco, São Paulo, Brazil

Prof. Dr. **Beatrice Rammstedt**, vice president, GESIS – Leibniz - Institute for the Social Sciences, Mannheim, Germany

Prof. Dr. **Brent Roberts**, University of Illinois Urbana-Champaign, United States

Prof. Dr. **Christopher Soto**, Colby College, United States

Em. Prof. Dr. **Fons van de Vijver**, Tilburg University, the Netherlands

City/Country Teams

Study Project Managers

Each participating city or country appoints a Study Project Manager who is responsible for implementing the Study in their jurisdiction. Study Project Managers are managed by the International Contractor, as described below.

Study Project Managers are the primary liaison between participating cities/countries and the International Contractor throughout the implementation of the Study. Study Project Managers play a vital role in ensuring that the Study is of high quality and producing verifiable and reliable results.

Study Project Managers are responsible for the translation of assessment items and other documents, and adapting wording in items to the local context, supported by and following procedures set out by the International Contractor. The Study Project Manager is also responsible for contracting and training local staff, such as Study Administrators, and for liaising with School Co-ordinators.

School Co-ordinators

A staff member from each sampled school is nominated as School Co-ordinator (SC). Their roles include collecting and recording information about the target population within the school, disseminating information to the school community, and co-ordinating and overseeing the administration of the assessments within their school.

Study Administrators

Study Administrators (TAs) administer the Study according to the Study's Technical Standards and Study Procedures. Teachers of students taking part in the assessment cannot be Study Administrators.

Quality Monitors

Quality Monitors (QMs) report on how well local sites followed protocol while conducting the student assessments. Study Project Managers nominate Quality Monitors but they are employed by and report directly to the International Contractor.