OECD Education Working Papers No. 208

Assessing students' social
and emotional skills through
triangulation of assessment
methods

Miloš Kankaraš,
Eva Feron ,
Rachel Renbarger

<https://dx.doi.org/10.1787/717ad7f2-en>

Assessing students' social and emotional skills through triangulation of assessment methods

OECD Education Working Paper No. 208

Miloš Kankaraš, Eva Feron and Rachel Renbarger (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

Eva Feron (eva.feron@oecd.org)

JT03454643

OECD EDUCATION WORKING PAPERS SERIES OECD

Working Papers should not be reported as representing the official views of the OECD or of its member countries. The opinions expressed and arguments employed herein are those of the author(s).

Working Papers describe preliminary results or research in progress by the author(s) and are published to stimulate discussion on a broad range of issues on which the OECD works. Comments on Working Papers are welcome and may be sent to the Directorate for Education and Skills, OECD, 2 rue André-Pascal, 75775 Paris Cedex 16, France.

This document, as well as any data and map included herein, are without prejudice to the status of or sovereignty over any territory, to the delimitation of international frontiers and boundaries and to the name of any territory, city or area.

You can copy, download or print OECD content for your own use and you can include excerpts from OECD publications, databases and multimedia products in your own documents, presentations, blogs, websites and teaching materials, provided that suitable acknowledgement of OECD as source and copyright owner is given. All requests for public or commercial use and translation rights should be submitted to rights@oecd.org.

Comment on the series is welcome and should be sent to edu.contact@oecd.org.

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

www.oecd.org/edu/workingpapers

© OECD 2019

Acknowledgments

The authors would like to thank Zsuzsa Bakk, Rose Bolognini, Marco Paccagnella, Ricardo Primi and William Thorn for valuable feedback on earlier drafts of this paper. 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

Triangulation – a combined use of different assessment methods or sources to evaluate psychological constructs – is still a rarely used assessment approach in spite of its potential in overcoming inherent constraints of individual assessment methods. This paper uses field test data from a new OECD Study on Social and Emotional Skills to examine the triangulated assessment of 19 social and emotional skills of 10- and 15-year-old students across 11 cities and countries. This study assesses students' social and emotional skills combining three sources of information: students' self-reports and reports by parents and teachers. We examine convergent and divergent validities of the assessment scales and the analytical value of combining information from multiple informants. Findings show that students', parents' and teachers' reports on students' skills overlap to a substantial degree. In addition, a strong 'common rater' effect is identified for all three informants and seems to be reduced when we use the triangulation approach. Finally, triangulation provides skill estimates with stronger relations to various life outcomes compared with individual student, parent or teacher reports.

Résumé

Triangulation – l'utilisation combinée de différentes méthodes ou sources d'analyse permettant d'évaluer les concepts psychologiques – est toujours rarement utilisée en tant que méthode d'analyse, malgré son potentiel à surpasser les contraintes inhérentes des méthodes d'analyse individuelle. Ce document utilise les données du test de terrain de la nouvelle enquête de l'OCDE sur les compétences sociales et émotionnelles afin de combiner trois sources d'information : l'auto-évaluation des élèves, et les évaluations des parents et des professeurs. Nous examinons les validités convergentes et divergentes des échelles d'évaluation, et la valeur analytique combinant l'information de multiple informateurs. Les recherches démontrent que les évaluations des élèves, parents, et professeurs sur les compétences des élèves se recoupent à un degré substantiel. De plus, un fort effet d'« évaluateur commun » est identifié pour les trois groupes d'informateurs et semble se réduire lorsque l'approche de triangulation est utilisée. Enfin, la méthode de triangulation fournit des estimations de compétence plus fortement reliées aux différents résultats de la vie, comparé à une évaluation individuelle d'un élève, parent ou professeur.

Table of Contents

Acknowledgments	3
Abstract.....	4
Résumé	4
1. Introduction.....	8
2. Triangulation of assessment methods.....	9
2.1. Triangulation of assessment scales: theory and practice	9
2.2. Triangulation in the Study on Social and Emotional Skills.....	12
3. Data and methodology	15
3.1. Sample	15
3.2. Assessment scales.....	17
3.3. Missing data and selection.....	19
3.4. Methodology.....	20
4. Results	22
4.1. Psychometric properties of assessment scales	22
4.2. Correlations between students', parents' and teachers' assessment estimates	23
4.3. Multi-Trait Multi-Method analysis.....	27
4.4. Creating a 'triangulated' measure of social and emotional skills.....	34
5. Discussion.....	44
6. Conclusion	48
References	50
7. Appendices.....	55

Tables

Table 3.1. Participating students and sites	15
Table 3.2. Descriptive statistics for student respondents	16
Table 3.3. Descriptive statistics for parent respondents.....	17
Table 3.4. Descriptive statistics for teacher respondents	17
Table 3.5. Overview of included social and emotional skills	19
Table 4.1. Psychometric properties of skill scales	23
Table 4.2. Relationship between scale reliabilities and their MTMM skill factor loadings for student, parent and teacher scales	33
Table 4.3. Factor loadings of student, parent and teacher scales on the triangulated factor, by cohort	36
Table 4.4. Relation between co-operation and politeness across four measures of informants	39
Table 4.5. Strength of relationship between skill estimates and behavioural indicators.....	40

Table 7.1. Raw skill scale descriptive statistics by respondent and cohort.....	55
Table 7.2. Comparison of student characteristics across informants	56
Table 7.3. Comparison of student characteristics across informants II	57
Table 7.4. Comparison of student characteristics across informants III	57
Table 7.5. Mean responses and their standard deviations in skill scales between all three informants	58
Table 7.6. Correlations between parent (x-axis) and student (y-axis) reports of student skills for the older cohort.....	59
Table 7.7. Correlations between parent (x-axis) and student (y-axis) reports of student skills for younger cohort.....	60
Table 7.8. Correlations between teacher (x-axis) and student (y-axis) reports of student skills for the older cohort.....	61
Table 7.9. Correlations between the teacher (x-axis) and student (y-axis) reports of student skills for the younger cohort.....	62
Table 7.10. Correlations between parent (x-axis) and teacher (y-axis) reports of student skills for the older cohort.....	63
Table 7.11. Correlations between parent (x-axis) and teacher (y-axis) reports of student skills for the younger cohort.....	64
Table 7.12. Correlations between skill factors in correlated MTMM models, by cohort	65

Figures

Figure 4.1. Correlation between corresponding assessment scales across different informants – older cohort.....	25
Figure 4.2. Correlation between corresponding assessment scales across different informants – younger cohort.....	25
Figure 4.3. Average correlation between student, parent and teacher assessment scales measuring the same or different skills, before and after corrections for attenuation.....	27
Figure 4.4. Example of an MTMM model with correlated skills from the same domain.....	28
Figure 4.5. Factor loadings of MTMM models with correlated skills from the same domain by cohort	29
Figure 4.6. Example of an MTMM model with uncorrelated skills from different domains	30
Figure 4.7. Selected examples of factor loadings of MTMM models with uncorrelated skills from different domains, by cohort.....	31
Figure 4.8. Average factor loadings of MTMM models across the three sets of models, by cohort.....	34
Figure 4.9. Estimation of triangulated measures of social and emotional skills from three respondents.....	36
Figure 4.10. Relationship between optimism and personal well-being, by type of skill estimate and cohort	41
Figure 4.11. Relationship between stress resistance and school test anxiety, by type of skill estimate and cohort	42
Figure 4.12. Relationship between achievement motivation and school grades, by type of skill estimate and cohort	43
Figure 7.1. Relationship between persistence and school grades, by type of skill estimate and cohort	66

Figure 7.2. Relationship between curiosity and school grades, by type of skill estimate and cohort	66
---	----

Boxes

Box 2.1. The Study on Social and Emotional Skills	13
---	----

1. Introduction

The OECD's new Study on Social and Emotional Skills is one of a growing number of large-scale international studies that aims to assess students' social and emotional skills. Studies on this topic assess these skills in different ways, such as through performance-based assessment methods, self-reports, reports from others (e.g. from students, parents and teachers) and observational data; and each method has its own advantages and disadvantages. Nevertheless, in all such studies, the assessment methods' validity – their capacity to evaluate the intended concepts – is critical.

Combining information from different assessment methods – or triangulation – in order to evaluate the same psychological construct is one way to overcome the inherent limitations or validity threats of each of the individual assessment approaches. Although this strategy has clear benefits, few empirical studies use triangulation because its implementation requires more resources, especially for large-scale studies. However, the Study on Social and Emotional Skills uses this assessment strategy by assessing students' social and emotional skills via three different sources of information – students' self-reports combined with their parents' and teachers' reports – offering a rare and valuable opportunity to examine the characteristics and the value of triangulation in a large-scale international assessment.

This paper investigates how the combination of information from three different informants on students' social and emotional skills affects the validity evidence in skill measurement. In particular, we examine the psychometric properties of each of the three types of skill estimates and compare them to each other. We then use Multi-Trait Multi-Method models to determine the amount of overlap between skill estimates measuring the same construct using different methods (convergent validity) as well as between estimates measuring different constructs using the same method (divergent validity). Finally, we create a triangulated estimate of skills by combining the three estimates from students, parents and teachers into a new skill estimate and compare its criterion validity with that of the three individual skill estimates. We use data from the study's field test which was implemented across 11 international sites and included two cohorts of students: 10- and 15-year-olds.

The first section of this paper discusses the triangulation method, its theoretical underpinnings and the empirical results of related literature. In the second section, we present the study's triangulated assessment approach. The third section shows the data and methodology used in this paper. The fourth section presents the results of our analyses. The fifth section discusses the presented results and the sixth section concludes.

2. Triangulation of assessment methods

In this section, we discuss the general approach of triangulation of assessment methods, its theoretical underpinnings as well as results of previous studies that have used this approach. Afterwards, we describe how the Study on Social and Emotional Skills implements this approach.

2.1. Triangulation of assessment scales: theory and practice

Triangulation in social science is a process that uses several research methods to examine the same psychological or social phenomena which is especially useful in the assessment of rich and complex human behaviour by providing data from more than one standpoint (Cohen, Manion and Morrison, 2002^[1]). Although its name implies the existence of three different methods, it can actually involve more independent sources of information about the same topic of interest¹. These methods can include different types of evidence (e.g. observational, written and conversational data) or different informants (e.g. friends, relatives, colleagues, self-reports, etc.) providing the same type of evidence. In this study, we examine triangulation in terms of questionnaires completed by students, their parents and their teachers.

Self-reported questionnaires are a preferred method of assessing human characteristics for a variety of reasons. Self-reported questionnaires ask participants questions about themselves rather than gathering information from another source (Gonyea, 2005^[2]). Self-reports offer a relatively simple and efficient way of assessing psychological characteristics. They are cost effective and quick to administer, tend to produce consistent results and in many cases can be a quite accurate approximation of objective measures (Connelly and Ones, 2010^[3]; Gonyea, 2005^[2]). Furthermore, research in the social sciences indicates that people tend to react to questionnaires in the intended way and are generally able to describe their typical behaviour accurately (Heine, Buchtel and Norenzayan, 2008^[4]; Krosnick, 1999^[5]; Vazire and Carlson, 2010^[6]). People's conceptions can converge with others' perceptions of them but they can also understand how their own perceptions differ from others' beliefs about them (Carlson, Vazire and Furr, 2011^[7]; Vazire, 2010^[8]). Finally, self-reported scales are often the only feasible assessment form in large-scale national or international research studies where a relatively large sample of respondents needs to be assessed within (strict) time constraints (Kankaraš, 2017^[9]).

Nevertheless, self-reported questionnaires have numerous drawbacks that can substantially impair their quality and usability. When assessing personality, these drawbacks include: respondents' lack of knowledge or

¹ The concept of 'triangulation' is borrowed from navigational and land surveying techniques that identify a single point in space using convergence of measurements taken from two other distinct points (Rothbauer, 2008^[50]).

misinterpretation of the item, the tendency to give socially desirable answers, inconsistency in answers and memory biases (Paulhus and Vazire, 2007_[10]). The situation is further complicated when these measures are used across cultures, as additional potential measurement biases are introduced, such as reference bias and measurement non-equivalence (Paulhus and Vazire, 2007_[10]). Moreover, self-reported scales are often shortened in large-scale assessments to improve their efficiency, increasing both type 1 and type 2 errors (Credé et al., 2012_[11]). For these reasons, researchers advocate to use other methods and sources of information, such as administrative data or reports from others that can be compared with self-reports for validation (Gonyea, 2005_[2]).

Other-reported questionnaires can also measure many of the same constructs measured through self-ratings. This approach assumes that other people may evaluate some individual characteristics more objectively and reliably than the individuals themselves. In fact, 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_[3]). One 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-ratings (Lindqvist and Vestman, 2011_[12]). Ratings by teachers are especially valuable for younger students, whose self-ratings tend to be less reliable. Teachers' ratings have good predictive validity for a number of behavioural markers for school children of all ages (Segal, 2012_[13]).

Other-reported 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 personally related to the individual they are rating) and inconsistency in answers (Connelly and Ones, 2010_[3]). Other-reported ratings are also less appropriate when personal experience, inner feelings and thoughts are the focus of the research, since these subjective states are more difficult to perceive by external observers.

Combining self- and other-reports provides complementary information, leading to a more comprehensive assessment and can also identify and correct for certain types of measurement issues (Connelly and Ones, 2010_[3]; Mathison, 1988_[14]). For example, self- (i.e. student), parent- and teacher-ratings give a more accurate overall picture of the student; they provide unique incremental information as they all know the student in a different context. This is a critical aspect as students may behave differently at home, in school or with their peers and choosing information from any one of these settings may provide a somewhat biased representation of students' social and emotional skills. Even when reports from students and others differ, these inconsistent findings can provide education researchers with a more holistic and meaningful picture of the phenomenon being studied (Mathison, 1988_[14]; Perlesz and Lindsay, 2003_[15]). Therefore, collecting information on students' skills from personal (student), school (teacher) and family (parent) perspectives yields

a better coverage and understanding of students' behaviours in the most important contexts.

Other studies have investigated the inclusion of parent reports when assessing students. In a study by Tackett (2011^[16]), both parents completed an assessment on their child's personality traits and problem behaviours. Results showed that mothers and fathers differed in their reports on their children's internal personality traits (traits difficult to observe from the outside), such as neuroticism and agreeableness. But these incongruent reports still provided unique information about the child's personality, showing differences in settings or relationships which are important in understanding the child. A different study comparing parent and child reports found that the differences in ratings were small, but tended to increase during adolescence (Göllner et al., 2017^[17]). Moreover, parent and child reports corresponded less when it came to neuroticism, agreeableness and extraversion, highlighting that these differences likely depend on the measured constructs. Again, certain personality traits can be easier to see through children's behaviours, making it useful to collect information from both the student and one or more adults familiar with the child.

There have also been positive results when it comes to including teacher reports. In a longitudinal study on children's problem behaviours, Verhulst, Koot and Van der Ende (1994^[18]) recommended using both teacher and parent reports to predict negative outcomes of children, such as mental health concerns or contact with the police; teacher reports, especially, are more likely to predict problem behaviours. Another study used both teacher and parent reports to diagnose Attention Deficit Hyperactivity Disorder (ADHD) (Power et al., 1998^[19]). The researchers carrying out the study also recommended the combined approach because each group provided different information depending on the aspect and context of ADHD, which helped to diagnose this disorder more accurately in children than information from only one informant. Teachers and students have also been found to agree on internal traits. Even though teachers' responses were shown to agree with students' responses to a higher degree on external characteristics, such as physical ability and peer relations, teachers also displayed moderate accuracy when assessing internal characteristics, such as self-concept (Marsh, 1990^[20]). Only one study by Tsolou and Margaritis (2013^[21]) directly examined how teacher reports compared to other reports on students' social and emotional skills. Teachers and students assessed students' self-awareness, self-management, emotional support, relationship building, communication and critical thinking skills. For the most part, the two respondents rated the students similarly in terms of social and emotional skills, except that the students rated themselves lower than the teachers did on students' self-management skills. Despite the different research areas, these results do show that teachers can assess their students' overall development and not just student achievement.

Few studies combine information from teachers, parents and students. Stanger and Lewis (1993^[22]) assessed students' behavioural problems to find the overlap between all three informants. Surprisingly, children reported more behavioural problems than their teachers did and mothers

and fathers had similar responses to each other. Overall, though, the students, teachers and parents responded similarly to both internalising and externalising behaviours. However, there was a strong rater effect, indicating the need for multiple reports (Stanger and Lewis, 1993^[22]). These results contrast those found by Newgent et al. (2009^[23]) where students reported *better* outcomes compared to parent and teacher reports. These discrepancies are likely due to the differences in environments between the various informants. For example, students, teachers and parents differed in their ability to assess the social functioning of students with Autism Spectrum Disorder or a learning disability because behaviours differ between a home environment and a large classroom (Haager, Watson and Willows, 2008^[24]; Jepsen, Gray and Taffe, 2012^[25]). To adjust for the fact that informants cannot interact with the student in all possible environments, using multiple reports to create a composite can increase the reliability of the estimates (Renk and Phares, 2004^[26]). As such, researchers should examine multiple reports to clearly understand student behaviours (Miller et al., 2018^[27]).

2.2. Triangulation in the Study on Social and Emotional Skills

The Study on Social and Emotional Skills (see Box 2.1) uses triangulation of assessment methods by asking three groups of respondents – students, parents and teachers – to report on students’ social and emotional skills. Parents are valuable sources of information regarding their children’s typical behaviours, thoughts and feelings, 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 peers.

Teachers’ reports are highly valuable because they interact with many children and can evaluate students’ social and emotional skills as a reasonably objective non-family member. 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.

All assessment instruments follow a standard Likert-type format, with statements describing typical behavioural patterns and response categories representing various degrees of agreement. Additionally, the instruments consist of the same assessment items for all three informants. However, not all informants have the same number of items per scale (see method section for more details). By using the same instruments for all informants, we can compare estimates of students’ social and emotional skills as well as the properties of the instruments across the three groups.

By obtaining information on students’ skills from three independent sources, it is possible to triangulate the estimates of students’ skills,

providing an additional means of evaluating their construct validities² (Conway and Lance, 2010_[28]). In addition, assessment information on student behaviours in various contexts obtained from separate sources could improve content validity³ of assessment scales.

Box 2.1. The Study on Social and Emotional Skills

The Study on Social and Emotional Skills is a large-scale international survey that assesses the social and emotional skills of 10- and 15-year-old students in a number of cities and countries around the world, identifying the conditions and practices that foster or hinder the development of these critical skills.

Beginning in mid-2017, the study takes place over a three-year period with a field test (pilot study) in October-November 2018 and the main study scheduled for October-November 2019. The findings will be released in September 2020.

Nineteen social and emotional skills were assessed in the field test. The selected skills draw on a well-known framework in the field of social and emotional skills: the Big Five model, which outlines how these skills should be organised.

The broad categories of the Big Five are:

- **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): These individuals can 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): Extraverted people 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

² Construct validity represents the degree to which a test measures what it intends to measure. It is an overarching term that incorporates different aspects of validity, such as convergent and divergent validity.

³ Content validity refers to the extent to which a measure represents all facets of a given construct.

quality relationships, more prosocial behaviours and, at least for children, fewer behavioural issues.

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

Each of the dimensions encompasses a cluster of mutually related social and emotional skills. For example, task performance includes achievement orientation, reliability, self-control and persistence. Apart from demonstrating their mutual similarity, these groupings also ensure a systematic, comprehensive and balanced consideration of individuals' social and emotional skills.

More information about the Study on Social and Emotional Skills can be found on the [study's website](#). This information includes [the conceptual framework](#) underpinning the study as well as the assessment framework, explaining the study design and assessment approach.

3. Data and methodology

This section discusses the sample used in the Study on Social and Emotional Skills field test followed by the descriptive statistics of the main variables. Finally, we address potential selection issues.

3.1. Sample

Participating sites in the study's field test included one country, Turkey and ten cities: Bogota, Colombia (BOG); Daegu, Korea (DAE); Helsinki, Finland (HEL); Houston, Texas, United States (HOU); Manizales, Colombia (MAN); Moscow, the Russian Federation (MOS); Ottawa, Ontario, Canada (OTT); Rome, Italy (ROM); Sintra, Portugal (SIN); and Suzhou, the People's Republic of China (SUZ). The study used a two-stage stratified sampling model to find a representative sample for the participating sites. The first stage included the selection of a random sample of schools and the second stage included the selection of a random sample of individual students within the selected schools. The same sampling procedure was applied for both age cohorts. Approximately 500 students were selected for participation per cohort and per site. Table 3.1 presents the number of participating students across the two cohorts and the 11 participating sites.

Table 3.1. Participating students and sites

SITE	Younger Cohort	Older Cohort	Total	Percent
BOG	732	686	1 418	10.30
DAE	661	771	1 432	10.40
HEL	432	322	754	5.48
HOU	709	622	1 331	9.67
MAN	612	659	1 271	9.23
MOS	684	596	1 280	9.30
OTT	735	539	1 274	9.25
ROM	636	726	1 362	9.89
SIN	490	455	945	6.86
SUZ	723	731	1 454	10.56
TUR	599	648	1 247	9.06
Total	7 013	6 755	13 768	100

This resulted in a final sample of 13 768 students, of which 7 013 students were in the younger cohort and 6 755 students were in the older cohort. Table 3.2 provides information regarding the control variables for students by cohort. There was an equal balance of male and female students. Students were approximately 15.9 years old in the older cohort and 11.0 years old in the younger cohort. Students in the older cohort were, on average, in grade 9 or 10 and students in the younger cohort were, on average, in grade 5. On average, students spent approximately 4 years in their current school. For both cohorts, the large majority of students and

their parents were from the country where they presently reside. More parents were born abroad compared to their children. The majority of students spoke the assessment language at home.

Table 3.2. Descriptive statistics for student respondents

	Obs		Mean		Std. Dev.		Min		Max	
	Younger	Older	Younger	Older	Younger	Older	Younger	Older	Younger	Older
Gender	6 802	6 681	0.50	0.50	0.50	0.50	0.00	0.00	1.00	1.00
Age in months	6 793	6 698	132.46	191.80	10.34	9.41	109.00	109.00	228.00	227.00
Grade	6 803	6 687	4.93	9.56	1.21	1.46	1.00	1.00	12.00	12.00
Years in this school	6 723	6 608	4.68	3.71	2.66	3.12	1.00	1.00	10.00	10.00
Birth country - student	6 639	6 600	1.06	1.07	0.23	0.25	1.00	1.00	2.00	2.00
Birth country - mother	6 658	6 653	1.18	1.17	0.38	0.37	1.00	1.00	2.00	2.00
Birth country - father	6 609	6 607	1.18	1.17	0.38	0.38	1.00	1.00	2.00	2.00
Socio-economic status (SES)	5 875	6 377	-0.07	0.03	0.99	1.01	-3.96	-4.08	2.09	2.42
Language at home	6 600	6 689	1.14	1.15	0.35	0.36	1.00	1.00	2.00	2.00

Note: The variables in this table are based on the total sample of 13 768 students. For gender, 1 indicates females and 2 indicates males. For each country of birth variable, a value of 1 indicates the person was born in the country where the student was currently enrolled while a 2 indicates the person was born outside the country. Socio-economic status (SES) is a factor score that is based on parents' education level, job occupation and wealth and cultural home possessions. If the language at home was the same as the assessment language it was coded as 1 and 2 if it was not.

For each sampled student, we asked parents or legal guardians to participate in the study by filling out a contextual questionnaire and an assessment instrument of their children's social and emotional skills. The parents/guardians decided which parent/guardian answered the questionnaire. Table 3.3 provides an overview of the parents who responded to the survey. In total, 7 007 (50.9%) parents responded to the questionnaire. Approximately 75% of the caregivers who completed the survey were the mother or female guardians. Less than 1% of the group was neither a parent nor a guardian. Mothers were, on average, between 35-39 years old for the younger cohort and between 40-44 years old for the older cohort. Fathers were, on average, slightly older than mothers. On average, parents completed upper secondary education (International Standard Classification of Education or ISCED level 3) or post-secondary non-tertiary education (ISCED level 4) which includes technical or professional education diplomas⁴.

⁴ More information regarding ISCED levels can be found here: <http://uis.unesco.org/sites/default/files/documents/international-standard-classification-of-education-isced-2011-en.pdf>

Table 3.3. Descriptive statistics for parent respondents

	Obs		Mean		Std. Dev.		Min	Max
	Younger	Older	Younger	Older	Younger	Older		
Survey completer	3 777	3 230	1.47	1.62	0.88	0.98	1	5
Age of mother	3 587	3 011	4.33	5.21	1.24	1.21	1	8
Age of father	2 857	2 506	4.92	5.71	1.34	1.17	1	8
Highest level of education- mother	3 715	3 133	4.26	3.94	2.04	2.03	1	8
Highest level of education- father	3 374	2 864	4.11	3.95	1.99	2.00	1	8

Note: The variable *survey completer* consists of 5 categories: 1. Mother 2. Other female guardian 3. Father 4. Other male guardian 5. Other. The variable *age of mother and father* consists of 8 different categories. The variable *highest level of education* is measured by the 2011 ISCED codes which represent the level of education completed by the parent/guardian from levels 1 to 8: 1. Primary education, 3. Upper secondary, 4. Post-secondary non-tertiary education. 8. Doctoral or equivalent.

The teacher who knew each sampled student best or with whom the student had spent the most time was asked to fill out the teacher contextual questionnaire and a short assessment report about the students' social and emotional skills. In total, 4 334 teachers responded to the questionnaire and assessed the social and emotional skills of one or more of the 13 768 students in the study. The majority of the teachers participating in the field test were female. The percentage of male teachers was slightly higher for the older cohort. On average, teachers were about 40 and 42 years old for the younger and older cohort, respectively. Over 90% of teachers reported being full time employees. On average, teachers of the younger cohort had been teaching for approximately 15 years and 16 years for the older cohort. In both cohorts, teachers had, on average, attained at least ISCED level 6 – equivalent to a bachelor degree. Table 3.4 outlines these teacher characteristics.

Table 3.4. Descriptive statistics for teacher respondents

	Obs		Mean		Std. Dev.		Min	Max
	Younger	Older	Younger	Older	Younger	Older		
Gender	1 857	2 347	1.21	1.34	0.41	0.47	1	2
Age	1 850	2 328	40.36	42.49	10.48	10.02	19	70
Employment status	1 861	2 356	1.15	1.12	0.47	0.41	1	4
Years as a teacher	1 819	2 279	15.19	16.27	10.29	9.68	0	51
Highest level of education	1 867	2 346	5.07	5.36	1.02	0.82	1	7

Note: The table is based on 4 334 individual teachers. For gender, 1 indicates females and 2 indicates males. The variable employment status has 4 categories: 1. Full time (+90%) 2. Part time (71-90%) 3. Part time (50-70%) 4. Part time (<50%). The highest level of education variable is measured by the 2011 ISCED codes which represent the level of education completed by the parent/guardian from levels 1 to 8: 1. Primary education, 3. Upper secondary, 4. Post-secondary non-tertiary education. 8. Doctoral or equivalent.

3.2. Assessment scales

Skill assessment scales for all 19 social and emotional skills were developed for the study. Wherever possible, existing scales or questions were used either in their original form or somewhat modified. The

International Personality Item Pool served as the main source for assessment items but a number of other scales were used, such as the Big Five Inventory-2. A number of items were also newly drafted.

The study's instrument development process included multiple rounds of empirical testing in various formats (both qualitative and quantitative) which were wide in scope in order to produce reliable, valid and comparable assessment instruments. The assessment scales were developed specifically for students in the target age groups, with particular focus on items being simple, clear and at an appropriate reading level. Additionally, in order to collect internationally comparable data, translations used by participating cities and countries had to meet stringent quality standards.

The study's instruments are divided into three sets of assessment scales: those designed for students to report on their behaviours, thoughts and feelings as well as those asking their parents and teachers to report on students' behaviours, thoughts and feelings. For comparison purposes, the same items are used in all types of the instruments, although the number of items per scale varies depending on the respondents. For students, the older cohort scales consist of ten items and eight items for the younger cohort. The parent assessment scales consist of eight items per scale while teachers' reports have three items per scale. Table 3.5 provides the abbreviations for each of the skills along with instrument examples.

All items are measured on a 5-point Likert scale. For further analyses, we created factor scores for all 19 skill scales using principal axis factoring. Therefore, each scale has a mean of zero and a standard deviation of one. Descriptive information for skill assessments for students by all three informants can be found in Table 7.1 in the Appendix.

Table 3.5. Overview of included social and emotional skills

Abbreviation	Skill	Example student item
AST	Assertiveness	I am a leader.
COO	Co-operation	I like to help others.
CRE	Creativity	I am original, come up with new ideas.
CRI	Critical thinking	I check if something is true before I believe in it.
CUR	Curiosity	I like learning new things.
EFF	Self-efficacy	I am confident in my abilities.
EMO	Emotional control	I keep my emotions under control.
EMP	Empathy	I care about what happens to others.
ENE	Energy	I am full of energy.
MET	Metacognition	I often stop to check how I am feeling.
MOT	Achievement motivation	I accept challenging tasks.
OPT	Optimism	I am always positive about the future.
PER	Persistence	I keep working on a task until it is finished.
RES	Responsibility	I keep my promises.
SEL	Self-control	I think carefully before doing something.
SOC	Sociability	I have many friends.
STR	Stress resistance	I worry about many things.
TOL	Tolerance	I ask questions about other cultures.
TRU	Trust	I think most of my classmates keep their promises.

3.3. Missing data and selection

We compared differences across a number of demographic and socio-economic variables, including the Study on Social and Emotional Skills index of socio-economic status, between students whose parents and teachers participated in the study with those students whose parents and/or teachers did not participate. Table 7.2 and Table 7.3 in the Appendix show comparisons of student characteristics across the three groups of informants. Informants were included in the sample if they had provided responses for at least one skill. If they had not provided responses for any skill they were not included. Of the 13 768 students, 97% of students, 53% of parents and 77% of teachers provided answers to at least one skill scale. Table 7.2 shows the means, the number of observations and the difference between the in and out of sample groups for students, parents and teachers.

Table 7.3 shows that across all informants there were few substantial differences in student characteristics between the in and out of sample groups. There appears to be some difference regarding the country of birth of students and parents, with parents of immigrant background being less likely to take part in the study. However, overall there appears to be little differences between samples of participating and non-participating respondents. Additionally, Table 7.4 in the Appendix shows there are few significant differences in student characteristics when we compared the sample of students, parents and teachers that we used for analyses. The means across the informants seem similar and any statistically significant differences seem mainly due to the large sample size.

3.4. Methodology

First, we discuss the psychometric properties of the assessment scales, including reliability and discrimination. Second, we discuss the bivariate correlations between students', parents' and teachers' skill estimates for all 19 skills to show to what extent the different estimates overlap. As the scales contain measurement error, we adjusted the correlations for attenuation by using the Cronbach alpha reliability coefficient. Third, in a study that uses different methods to assess the same and different psychological characteristics, it is important to determine the amount of overlap between skill estimates measuring the same construct using different methods (convergent validity) and between estimates measuring different constructs using the same method (divergent validity).

A Multi-Trait Multi-Method (MTMM) matrix is a method developed to simultaneously investigate convergent and divergent validity of multiple measures that have used different assessment methods. The MTMM matrix examines how correlations among scales that measure the same skills but come from different sources compare to correlations among scales that measure different skills but come from the same sources. For example, we investigated whether the correlations between estimates of students' creativity obtained from all three sources (students, teachers and parents) were higher than correlations between scales measuring creativity, persistence and energy, as reported by one same source (e.g. by students). The MTMM models allow us to determine to which degree variation in skill estimates is related to 'method' factors, i.e. to common source variation and to which degree it is related to 'skill' factors, i.e. to valid information about assessed social and emotional skills.

In our context, the different 'methods' of assessment are represented by three different informants – students, teachers and parents – who report on the social and emotional skills of the same students. Each of these methods have their own positive and negative aspects in terms of the accuracy and validity of information obtained about students' social and emotional skills. Every informant does not only contribute overlapping information but also complementary unique information. Each informant also has its own measurement error, only some of which is shared across methods.

We employed the MTMM approach by using Structural Equation Models (SEM) in which we modelled for two sets of latent factors: one set representing the skill factors and the other, representing the three method factors. Both sets were allowed to vary across sites and were estimated using skill estimates in the form of factor scores⁵ that were previously obtained from each of the skill scales. The MTMM models were estimated separately for the two age cohorts.

After establishing there was a considerable amount of convergent validity (overlap) across the three informants, we created a triangulated measure of the social and emotional skills. We used factor analysis to create 19

⁵ Another option was to use simple scale sum-scores (e.g. see (Primi et al., 2019^[51]).

triangulated social and emotional skill measures based on the commonality in the skill estimates of students, parents and teachers. We then compared the strength of the relationship between the three individual skill estimates and the triangulated skill estimate on the life outcomes of students, while we controlled for a standard set of background characteristics. These life outcomes included behavioural indicators, well-being, test anxiety and academic achievement in the form of school grades in reading, science, math and arts. Grades were an especially interesting case to use here as grades are not subject to common method bias because they are obtained via administrative records.

4. Results

This section presents the results of our analyses. First, we discuss the psychometric properties of the scales assessing social and emotional skills by students, parents and teachers. Second, we present correlations between estimates of the same skills across three different raters. Third, we examine convergent and divergent validities of the assessment scales provided by three informants using Multi Trait Multi Method models. Finally, for each skill, we create a triangulated skill estimate based on the information from all three informants and observe how this triangulated measure compares to the three individual measures in terms of criterion validity.

4.1. Psychometric properties of assessment scales

The key findings of the skill scales analysis are presented in the summary tables which outline the results for each informant separately. Each table provides the following information per skill:

- The Cronbach alpha reliability: an overall indication of the internal consistency of the assessment scales.
- Latent reliability: an IRT estimate of the reliability of the latent variable. These are the IRT analogues of the Cronbach alpha estimates.
- Low discrimination: The ‘item-rest correlation’ is a classical test theory estimate of the discrimination of the item (i.e. the correlation between performance on the test item and the performance on all other items). For each cohort, the number of items within the scale with item-rest correlations lower than 0.2 are reported.

The results in Table 4.1 show that the reliabilities of the scales are moderate to high, with most of them varying between 0.7 and 0.8. The reliability estimates for students from the younger cohort and parents are somewhat lower than the estimates for students from the older cohort. Furthermore, teacher-reported assessment scales have the lowest reliabilities. These results could be expected given that scales for teachers have only three items per scale, compared to ten items per scale for older students and eight items per scale for younger students and parents. The lower reliabilities of the teacher scales can also be due to teachers’ limited knowledge of some of students’ characteristics. The estimate for critical thinking is low across all three informants. For almost all skill scales there is a maximum of one item that has an item-rest correlation that is lower than 0.2.

Table 4.1. Psychometric properties of skill scales

	Students				Parents				Teachers			
	Alpha		Latent Reliability		Alpha		Latent Reliability		Alpha		Latent Reliability	
	Y	O	Y	O	Y	O	Y	O	Y	O	Y	O
AST	0.8	0.88	0.8	0.88	0.81	0.82	0.81	0.82	0.88	0.87	0.88	0.87
COO	0.8	0.8	0.8	0.8	0.68	0.68	0.68	0.68	0.56	0.46	0.56	0.46
CRE	0.75	0.83	0.75	0.83	0.75	0.77	0.75	0.77	0.72	0.72	0.72	0.72
CRI	0.3	0.6	0.3	0.6	0.42	0.39	0.42	0.39	0.14	-0.02	0.14	-0.02
CUR	0.77	0.8	0.77	0.8	0.61	0.63	0.61	0.63	0.64	0.64	0.64	0.64
EFF	0.76	0.83	0.76	0.83	0.66	0.64	0.66	0.64	0.58	0.57	0.58	0.57
EMO	0.74	0.85	0.74	0.85	0.7	0.67	0.7	0.67	0.63	0.6	0.63	0.6
EMP	0.66	0.77	0.66	0.77	0.55	0.51	0.55	0.51	0.63	0.58	0.63	0.58
ENE	0.68	0.79	0.68	0.79	0.66	0.69	0.66	0.69	0.73	0.71	0.73	0.71
MET	0.67	0.69	0.67	0.69	0.63	0.55	0.63	0.55	0.65	0.57	0.65	0.57
MOT	0.68	0.78	0.68	0.78	0.69	0.68	0.69	0.68	0.75	0.74	0.75	0.74
OPT	0.75	0.87	0.75	0.87	0.69	0.74	0.69	0.74	0.76	0.75	0.76	0.75
PER	0.75	0.86	0.75	0.86	0.82	0.81	0.82	0.81	0.91	0.9	0.91	0.9
RES	0.76	0.81	0.76	0.81	0.71	0.68	0.71	0.68	0.55	0.56	0.55	0.56
SEL	0.77	0.79	0.77	0.79	0.8	0.76	0.8	0.76	0.58	0.55	0.58	0.55
SOC	0.65	0.81	0.65	0.81	0.53	0.6	0.53	0.6	0.27	0.4	0.27	0.4
STR	0.77	0.87	0.77	0.87	0.76	0.76	0.76	0.76	0.66	0.62	0.66	0.62
TOL	0.77	0.82	0.77	0.82	0.71	0.74	0.71	0.74	0.75	0.76	0.75	0.76
TRU	0.81	0.87	0.81	0.87	0.74	0.77	0.74	0.77	0.65	0.6	0.65	0.6
Average	0.72	0.81	0.72	0.81	0.68	0.68	0.68	0.68	0.63	0.61	0.63	0.61

Note: Y indicates the younger cohort and O indicates the older cohort.

Table 7.5 in the Appendix shows the differences in average responses in skill scales across the three informants. Students tend to provide higher self-ratings (overall 3.60 on average on a 1 to 5 scale) than ratings provided by parents (overall 3.36 on average) and teachers (3.44). However, this order changes for some of the scales. For example, teachers reported that students control their emotions more than the students themselves and much more than their parents reported. And while the differences in average responses across respondents tended to be relatively small (overall average of these differences is around 0.20), in some cases they were rather substantial. For example, the average rating on the students' responsibility scale was 3.66 in student reports but only 2.78 in parent reports.

4.2. Correlations between students', parents' and teachers' assessment estimates

Figure 4.1 and Figure 4.2 depict the correlations between students', parents' and teachers' assessment scales, separately for each cohort. These results show that there are moderate correlations between the skill estimates of the three informants across all 19 skills. The strength of correlations is in line with previous research, which have found these correlations to generally vary between 0.30 – 0.50 (Abrahams L., 2019^[29]; Major, Seabra-Santos and Martin, 2015^[30]; Marsh, 1990^[20]; Marsh and Byrne, 1993^[31]). For example, Major, Seabra-Santos and Martin (2015^[30]) have found that inter-correlations between measures of different social and emotional behaviours based on parent and teacher reports are on average about 0.30. Not surprisingly, their findings show more agreement in externalising behaviours that are, in general, easier to observe than internalising

behaviours. Likewise, Achenbach and colleagues (1987^[32]) have found a correlation of only 0.22 between children's self-reports and reports by adults. In fact, findings of modest correlations between informants have been cited as 'the most robust findings in paediatric clinical research' (De Los Reyes and Kazdin, 2005^[33]).

These results could be expected given that these are multidimensional constructs that are manifested differently depending on the context (Abrahams L., 2019^[29]; Ramsey et al., 2016^[34]). The findings indicate that although the three informants generally tend to perceive students' behaviours, feelings and thoughts in a similar way, they also each value and focus on somewhat different aspects of these constructs and thus give different ratings. Therefore, they are in line with the general recommendation that different informants (e.g. children, parents, teachers, peers, etc.) should be used to assess different aspects of children's psychological characteristics (Achenbach, McConaughy and Howell, 1987^[32]; Grietens et al., 2004^[35]; Kamphaus and Frick, 1996^[36]; Treutler and Epkins, 2003^[37]; Winsler and Wallace, 2002^[38]). This recommendation is based on the knowledge that differences in informants' characteristics as well as in the setting in which behaviours are observed can affect the ratings that are made (De Los Reyes and Kazdin, 2005^[33]; Gagnon, Nagle and Nickerson, 2007^[39]; Major, Seabra-Santos and Martin, 2015^[30]; Merrel, 2008^[40]; Strickland, Hopkins and Keenan, 2012^[41]).

At the same time, levels of inter-correlation between corresponding scales from different informants could be reduced not only due to differences in behaviours across contexts but to methodological factors as well. In particular, using different informants to assess the same construct tends to lead to underestimation of the actual association between the scales even if the observed behaviours are completely the same (Conway, 2002^[42]). This underestimation is due to informant-specific measurement errors (both random and systematic) across the different informants. These measurement errors that are unique to each informant and the way informants respond to a given scale reduce correlations between scales measuring the same construct. The rater effect is an informant-specific systematic measurement error, which tends to differ across informants, contributing to the reduction of inter-correlations between corresponding measures of the same construct. Thus, the more rater effects in student, parent and teacher reports are uncorrelated with one another, the more the resulting correlations between their measures will decrease.

It is interesting to note that there is a general trend for the younger cohort to have a lower correlation between the skill estimates of the three informants compared to the older cohort. However, this is not consistent across all skills. The correlation between students and parents is generally higher than the correlation between students and teachers and teachers and parents.

Figure 4.1. Correlation between corresponding assessment scales across different informants – older cohort

Note: The graph presents pairwise correlations between students, parents and teachers using pooled data. Correlations are corrected for attenuation.

Figure 4.2. Correlation between corresponding assessment scales across different informants – younger cohort

Note: The graph presents pairwise correlations between students, parents and teachers using pooled data. Correlations are corrected for attenuation.

The full list of inter-correlations between corresponding and non-corresponding scales across the three informants can be found in the Appendix in Table 7.6 through Table 7.11.

Correlations between assessment scales presented in Figure 4.1 and Figure 4.2, were corrected for attenuation⁶ by using the Cronbach's alpha reliability coefficients of assessment scales. By doing so, in accordance with classical test theory, we estimate the association between true scores of the two scales, i.e. one that is corrected for its measurement error. Figure 4.3 presents average correlations between scales measuring the same skills across the three informants, before correction for attenuation. As expected, after correcting for attenuation, the average correlation between skill estimates increases. For example, the average correlation of students' and parents' skill estimates across all skills, rises from 0.34 and 0.24 for older and younger cohorts to 0.47 and 0.35 for the two cohorts, respectively. The average corrected correlation between students' and teachers' skill estimates is substantially lower across all skills: 0.24 and 0.23 for older and younger cohorts, respectively. Finally, the average corrected correlation is 0.30 between parents' and teachers' skill estimates across all skills and 0.30 for older and younger cohorts.

Apart from the correlations between scales measuring the same skills that inform us about convergent validities⁷ of given estimates, we have also calculated correlations between scales from different informants measuring different skills. Such correlations inform us about divergent validities⁸ of the assessment scales, i.e. the degree to which they are able to distinguish between measures of different constructs. As can be seen from Figure 4.3, these correlations generally vary around 0.10, indicating a satisfactory level of divergent validity. It should also be noted that selected social and emotional skills are even theoretically expected to co-vary to some degree in the case that the skills belong to the same Big Five domain.

⁶ Correction for attenuation allows for a better estimate of true strengths of association between two measures by taking into account (i.e. correcting for) the observed amount of measurement error in the two scales.

⁷ Convergent validity refers to the degree to which two measures of constructs that theoretically should be related, are in fact related.

⁸ Divergent validity represent the degree to which two measures of constructs that theoretically should not be related, are in fact related.

Figure 4.3. Average correlation between student, parent and teacher assessment scales measuring the same or different skills, before and after corrections for attenuation

4.3. Multi-Trait Multi-Method analysis

Given that there is a complex structure of interrelated constructs within the wider Big Five domains, a number of different structures of MTMM models were possible to estimate. First, we could choose to estimate measures of skills coming from the same Big Five domain (in which case they would be allowed to correlate) versus estimating measures of skills from different Big Five domains (where they would be independent from one another). Secondly, it is possible to estimate a different number of skill factors in an MTMM model, with the number of skill factors varying between three and five. We present results on these variations of the MTMM structures to show how these different model set-ups affect results.

4.3.1. MTMM models with correlated skills from the same Big Five domains

In this section, we present results for the MTMM models with skill factors from the same Big Five domain. Given that these skills are conceptually and empirically related to one another, they are allowed to correlate. We estimated five models, one for each of the Big Five domains. An example of the structure of such a model for the domain of Working with Others (Agreeableness) is presented in Figure 4.4.

Figure 4.4. Example of an MTMM model with correlated skills from the same domain

Figure 4.5 presents the results of these five MTMM models. The average factor loadings of the skill factors are .41 and .38 for older and younger cohorts respectively. At the same time, the average factor loadings for the method factors are .52 for the older cohort and .54 for the younger cohort. Structural relations (correlations between skill factors) were .71 and .74 on average for older and younger cohort respectively (see Table 7.6 in the Appendix for a full list of these correlations). Obviously, such substantial structural relations between related skills have to be incorporated in the MTMM models for more valid estimation of all model parameters. The strength of the structural relations between measures of the same skill across three informants, as represented by the size of the factor loading of skill factors in these MTMM models, indicates a moderate degree of overlap in the information captured by scales from different informants measuring same construct. This level of convergence between the skill estimates from different sources indicates that they do assess the same underlying construct but that they also incorporate a substantial amount of information that is unique to each of the three types of measures.

Apart from investigating MTMM models with skills belonging to the same Big Five domain, which are conceptually related and mutually correlated, we also examined models that include skills that belong to different domains. Skills belonging to different Big Five domains are, by definition, (largely) uncorrelated. Therefore, including skills from different domains in the MTMM models allows us to examine divergent validity of assessment scales in a more direct way.

Figure 4.5. Factor loadings of MTMM models with correlated skills from the same domain by cohort

4.3.2. MTMM models with uncorrelated skills from different Big Five domains

The MTMM models with skill scales belonging to different Big Five domains are generally assumed to be independent from one another, although there are some conceptual and empirical exceptions. In this section, we present results of a number of MTMM models in which the number of skills included in the models come from three to five different Big Five domains. In these models, the skill factors were not assumed to correlate with one another. Among the large number of possible combinations of 19 skill measures from five domains, we have estimated six models: three models with three (largely) independent skills and three models with five (largely) independent skills. An example of the structure of such a model with three independent skills belonging to different Big Five domains is presented in Figure 4.6.

Figure 4.6. Example of an MTMM model with uncorrelated skills from different domains

The factor loadings of the skill and method factors in these six MTMM models are presented in Figure 4.7, separately for the older and younger cohort.

Figure 4.7. Selected examples of factor loadings of MTMM models with uncorrelated skills from different domains, by cohort

These results show a similar magnitude of the skill factor loadings as in the case of MTMM models with correlated skill factors, indicating there is a relatively moderate degree of convergent validity of the skill measures. Across these six models, the average factor loadings of the skill factors are .37 and .33 for the older and younger cohort, respectively. The average factor loadings of the method factors are .49 and .50 for the older and younger cohort, respectively. These results support the conclusions about the moderate overlap between the skill measures obtained from the three different sources. At the same time, the moderate size of factor loadings of skill factors indicate that each of the three skill measures captures additional information that is not shared with the other two measures. As discussed before, some of this additional information might be due to valid observations of different aspects of students' behaviours across different contexts (McCrae, 2015^[43]). Another part of this informant-specific information is due to measurement errors that reduce the overlaps across measures from different sources.

In this sense, it is important to note that the observed factor loadings of the skill factors are strongly related to the reliabilities of the skill measures. An overview of the relations between the scale reliability and the MTMM factor loadings for all skill scales is presented in Table 4.2. In particular, the correlations between the individual skill factor loadings and the scale reliabilities of the corresponding measures are, on average, .63, .53 and .58 for student, parent and teacher reports, respectively. These results indicate that, as expected, the random measurement error in the skill scales reduces the magnitude of the skill overlaps and consequently reduces the convergent validity of the measures.

These results show considerable differences exist among the skill scales in terms of their loadings on the corresponding skill factor, despite measurement error or scale specificities that can influence the convergent validity of the scales. The skills that have the highest factor loadings are relatively easy to observe by all informants, such as sociability, assertiveness or energy. Moreover, they do not necessarily have especially high reliability estimates. Nevertheless, the skill scales that assess more subjective states and mental processes, such as metacognition, trust or critical thinking, have much lower levels of convergent validity across the informants. Thus, the degree to which certain skills are observable in students behaviours and in different contexts seem to influence the convergent validity of the information provided by the three groups of respondents. These patterns of results are in agreement with previous research (e.g. (Marsh and Byrne, 1993^[31]; McCrae and Costa, 1987^[44]; McCrae and Costa, 1988^[45])).

Table 4.2. Relationship between scale reliabilities and their MTMM skill factor loadings for student, parent and teacher scales

Skills	Student scales		Parent scales		Teacher scales	
	Scale reliability	MTMM factor loading	Scale reliability	MTMM factor loading	Scale reliability	MTMM factor loading
Assertiveness	0.88	0.53	0.82	0.53	0.87	0.53
Co-operation	0.80	0.40	0.68	0.40	0.46	0.40
Creativity	0.83	0.33	0.77	0.33	0.72	0.33
Critical thinking	0.60	0.22	0.39	0.22	-0.02	0.22
Curiosity	0.80	0.41	0.63	0.41	0.64	0.41
Self-efficacy	0.83	0.40	0.64	0.40	0.57	0.40
Emotional control	0.85	0.44	0.67	0.44	0.60	0.44
Empathy	0.77	0.37	0.51	0.37	0.58	0.37
Energy	0.79	0.45	0.69	0.45	0.71	0.45
Metacognition	0.69	0.31	0.55	0.31	0.57	0.31
Achievement motivation	0.78	0.48	0.68	0.48	0.74	0.48
Optimism	0.87	0.43	0.74	0.43	0.75	0.43
Persistence	0.86	0.48	0.81	0.48	0.90	0.48
Responsibility	0.81	0.43	0.68	0.43	0.56	0.43
Self-control	0.79	0.41	0.76	0.41	0.55	0.41
Sociability	0.81	0.52	0.60	0.52	0.40	0.52
Stress resistance	0.87	0.45	0.76	0.45	0.62	0.45
Tolerance	0.82	0.39	0.74	0.39	0.76	0.39
Trust	0.87	0.31	0.77	0.31	0.60	0.31
Average correlation	0.63		0.53		0.58	

4.3.3. Summary results of the MTMM models

The aggregate results of the two sets of MTMM models are presented in Figure 4.8. They indicate the presence of strong method factors or a ‘rater effect’, as the reports by the same raters on different skills strongly correlate. In other words, raters who indicate positive behaviours on one skill tend to indicate positive behaviours on other skills as well, irrespective of the actual skill pattern of the student in question and vice versa. Such strong method factors reduce the divergent validity of the skill scales, i.e. their ability to capture skill-specific variability across students.

The method factor is strongest for teachers, which is expected given their shorter scales (three items per scale). And consequently, their scales could contain more measurement error and ‘common source’ bias. Moreover, teachers are the only one of the three groups of raters that assessed multiple students – on average, 2.6 students per teacher, which could have triggered a higher rater effect in teacher estimates. Conversely, the older cohort has somewhat higher skill factor loadings and lower method factor loadings compared to the younger cohort. This could be partly attributed to their longer scales (ten items per scale compared to eight items for the younger cohort), although reliabilities of the assessment scales for the older cohort are still slightly higher than those for the younger cohort even after accounting for their different length using the Spearman-Brown formula. Higher convergent validities of assessment scales of older students might also have to do with their slightly better ability to provide more objective self-reflection (Marsh and Byrne, 1993^[31]).

The MTMM results show a relatively moderate degree of convergent validity across the three groups of respondents. A similar level of convergence is evident both in the models

with skills from the different Big Five domains as well as in models with correlated skills from the same Big Five domain. These results are in line with previous research that reported skill factor loadings between students, parents and teachers in MTMM models ranging between .30 and .50 (Marsh, 1990^[20]). Apart from the existence of overlapping information, the relatively moderate size of factor loadings of skill factors indicates that each informant provides unique information about students' skills.

Given that the factor loadings of skill scales are larger than 0.30⁹, it is possible to create an overall, higher-order measure of the assessed skills based on the information from all three individual respondents. This new, triangulated scale is examined in the following section.

Figure 4.8. Average factor loadings of MTMM models across the three sets of models, by cohort

4.4. Creating a 'triangulated' measure of social and emotional skills

Using different informants to estimate one construct inevitably raises the issue of how to treat these corresponding measures in consequent analyses, i.e. whether or not to try to aggregate the reports from different informants or to treat them separately. In this section, we will look at the conditions to create such aggregated measure as well as its validity in comparison to the three individual skill estimates.

Decisions on how to treat skill estimates from multiple informants are driven by both conceptual and statistical considerations. First, as discussed before, it is important to stress that differences between informants do not necessarily represent measurement error but are to be expected even if each of the informants reports valid observations of students' skills. In particular, it could be expected that assessed students behave somewhat differently across different contexts and in relation with different people, such as teachers and parents. Thus, the differing information captured in reports from different informants could be equally valid as the information captured that is similar in all of the reports (Achenback et al., 2008^[46]; Major, Seabra-Santos and Martin, 2015^[30]). Therefore, one way to treat the available skill estimates is to keep all three individual skill scales and conduct follow-up

⁹ In order for an item or scale to be considered an indicator of a higher order factor, corresponding factor loadings of 0.30 and above are generally required (e.g. (Grimm and Yarnold, 1995^[52])).

analyses with each of them. In this way, one could make use of the valid information on student skills from all three informants, both from the part that overlaps among their reports as well as from the part that is uniquely provided by each informant.

On the other hand, there are also benefits with aggregating results from different sources. First, it allows for a much more parsimonious approach, where one unified skill estimate is used instead of a number of individual ones. It also avoids situations in which different skill measures have relationships with other variables that substantially vary from each other and are thus difficult to interpret. Aggregating results by using some of the latent variable models, such as factor analysis, also assures that informant-specific measurement errors of individual scales are removed from the unified measures, improving their reliabilities and validities. Yet, in order to aggregate different informants' reports, they must overlap to a certain degree.

Thus, a number of conceptual and statistical considerations will determine the way triangulated data are used. Researchers must decide whether the overlapping information among the three sources is conceptually the most relevant and how important it is to use the valid information about the given construct, captured only in individual reports. In addition, the amount of observed statistical relationships between the measures from different sources will indicate how they can be used. For example, in a situation of moderate to high inter-correlations between these scales, it would make most sense to use one unified measure. On the other hand, in situations of low or no inter-correlations, using three individual measures is statistically the only available option since creating a unified variable requires at least moderate inter-correlations between skill measures.

The results from the MTMM models show that students', parents' and teachers' reports on students' skills overlap to a relatively moderate degree, even after controlling for the 'common method' bias. Similar conclusions can be reached after analysing the factor loadings of models in which estimates of students' skills from the three respondents were factor analysed. These analyses consisted of two steps. In the first step, 57 factor analyses of all 19 scales from three sources – student, parent and teacher reports (19 x 3) – were conducted (separately for older and younger cohorts). Factor scores of these analyses were then saved as separate variables, representing skill estimates for each skill by each of the three informants. In the second step, 19 factor analyses were conducted, each with three factor score variables saved in step one that represented the estimates of the same skills obtained from the three different respondents. For example, three estimated factors of students' emotional control, from student, parent and teacher scales, were factor analysed to obtain one unified (triangulated) estimate of students' emotional control. This was repeated for all other assessed skills. A schematic representation of these two steps is presented in Figure 4.9. The full set of factor loadings obtained from the second step, as well as their averages across skills and informants, is presented in Table 4.3, separately for the two cohorts.

Figure 4.9. Estimation of triangulated measures of social and emotional skills from three respondents

Table 4.3. Factor loadings of student, parent and teacher scales on the triangulated factor, by cohort

	Older cohort				Younger cohort			
	Students	Parents	Teachers	Average	Students	Parents	Teachers	Average
ASS	0.66	0.70	0.39	0.58	0.50	0.72	0.43	0.55
COO	0.50	0.59	0.24	0.44	0.47	0.47	0.35	0.43
CRE	0.47	0.67	0.18	0.44	0.48	0.50	0.30	0.42
CRI	0.43	0.53	0.08	0.35	0.35	0.37	0.09	0.27
CUR	0.61	0.70	0.32	0.54	0.54	0.58	0.32	0.48
EFF	0.44	0.70	0.29	0.48	0.41	0.62	0.36	0.46
EMO	0.50	0.73	0.27	0.50	0.40	0.64	0.34	0.46
EMP	0.48	0.60	0.21	0.43	0.46	0.36	0.26	0.36
ENE	0.63	0.63	0.29	0.52	0.46	0.60	0.38	0.48
MET	0.37	0.57	0.27	0.41	0.43	0.47	0.29	0.40
MOT	0.57	0.74	0.35	0.55	0.50	0.60	0.43	0.51
OPT	0.58	0.59	0.26	0.48	0.48	0.54	0.31	0.44
PER	0.51	0.70	0.39	0.53	0.45	0.65	0.49	0.53
RES	0.44	0.70	0.34	0.50	0.41	0.69	0.38	0.49
SEL	0.50	0.66	0.32	0.50	0.40	0.72	0.31	0.47
SOC	0.68	0.68	0.30	0.55	0.57	0.56	0.34	0.49
STR	0.54	0.69	0.22	0.48	0.35	0.64	0.27	0.42
TOL	0.67	0.55	0.22	0.48	0.39	0.52	0.22	0.38
TRU	0.59	0.54	0.25	0.46	0.52	0.41	0.25	0.40
Average	0.54	0.65	0.27	0.46	0.45	0.56	0.32	0.44

These results confirm the considerable overlap between the skill estimates from different sources, with an average factor loading of .46 and .44 for older and younger cohorts, respectively. As expected, these factor loadings indicate a higher degree of correspondence between the groups of informants than what could be concluded based on the initial analysis of their inter-correlations. This is because factor loadings exclude unique variations within the scale and only represent the strength of the relationship between given skill estimates for the overlapping part of information among the three types of scales. Factor loadings of this size imply that there is enough of an overlap (communality) between reports from individual informants to justify creating unified factor measures. The variation of factor loadings across the skills illustrates that the skills that are behaviourally easier to observe are also more similarly estimated by the three respondents. Interestingly, parent reports overlap with the other two reports most, with average factor loadings of .65 and .56 for older and younger cohorts respectively. Teacher reports, however, have the least communality with the other two reports, with average factor loadings of .27 and .32 for the younger and older cohorts respectively. These results are in line with the correlations between the scales presented in the Appendix in Table 7.6 through Table 7.11. Lower communalities of teacher-reported skill estimates could, at least partially, be the result of a higher level of measurement error in these scales due to their brevity (only three items per scale).

Based on the observed overlap between skill estimates and in line with the results obtained from the MTMM models, we proceeded by creating ‘triangulated’ skill estimates. These estimates represent factor scores of each of the 19 factor analytical models, created from the responses from all three participants. For example, the ‘triangulated’ measure of assertiveness was created by factor analysing skill estimates obtained from the student, parent and teacher assertiveness scales. Thus, these triangulated measures represent the part of concept-relevant data that were captured in reports from all three informants. In the following sections we examine how the triangulated measures compare with the three individual measures in regard to their criterion validity (the most important aspect of their use in policy analyses). At the same time, we investigate the degree to which triangulated measures are able to alleviate some of the methodological constraints that are common for these types of reports from individual informants.

4.4.1. Comparing criterion validities of ‘triangulated’ measure of social and emotional skills with measures obtained from individual informants

One of the basic premises of combining assessment sources – triangulation – is that it brings new insights and overcomes some of the drawbacks of each of the individual methods. It can do so either by using individual skill estimates from different informants in parallel or by creating an aggregated measure based on the combination of individual estimates. In this section, we will compare the triangulated skill estimate with the three individual skill estimates across some of the most relevant psychometric properties in order to check which of the two approaches to dealing with triangulated data is more appropriate in this study.

One of the most critical aspects of these estimates’ validity is their ability to identify and estimate the strength of relationships between students’ assessed skills and life outcomes (criterion validity). It is important to note that in order to compare and evaluate the criterion validities of different skill estimates, it is necessary to take into account theoretical expectations and previous research on the relationship between the two variables: skill predictor and its criterion/outcome. These findings are then used as reference points on the type and strength of a criterion validity coefficients between the two variables against which all criterion validity coefficients of various skill measures are compared.

Comparing the results on the criterion validity of the ‘triangulated’ scales with the three individual skill scales from students, parents and teachers produces a large number of possible relationships to investigate. Below is a selection that is the most illustrative for our purposes.

The study assesses behavioural indicators which are a set of questions regarding students’ behaviours at home, in class and in school, such as class behaviours, school absenteeism, behaviours with friends and parents, disruptive conduct and health-related behaviours. Each of the behavioural indicators can be seen to a certain extent as a concrete manifestation of one or more assessed social and emotional skills. Of course, any particular behaviour might (and usually is) caused by a multitude of psychological factors, so the relationship between any social and emotional skill and its related behavioural indicator is probabilistic rather than deterministic. Therefore, the strength of the relationship between a skill estimate and the related behavioural indicator might be considered as a measure of its criterion/concurrent validity¹⁰. Likewise, comparison of the strengths of these relationships between different skill estimates could serve as an indicator of their comparative criterion validity and, consequently, their analytical value.

However, two important caveats, one methodological and one conceptual, needs to be taken into account when considering these behavioural indicators as skill criteria. First, from a methodological point of view, there is a possible biasing effect of common source of information for both measures. Second, there is the issue of a conceptual overlap between skills and related behavioural indicators. In particular, while the distinction between a particular social and emotional skill and its potential behavioural indicator is rather clear in some cases, in others this distinction is much more blurred. For example, stress resistance, defined as the ability to deal with stress and anxiety and calmly solve problems, is conceptually different from its behavioural indicator of reporting physical problems without known medical cause. On the other hand, in the case of co-operation, defined as the ability to live in harmony with others and its related behavioural indicator of being polite, this conceptual distinction between the two is not as clear.

Table 4.4 presents standardised regression coefficients of the relation between four different measures of co-operation (three created from individual student, parent and teacher reports and the fourth, ‘triangulated’ created from a combination of these three) and the behavioural indicator that is most relevant to this skill: students’ politeness (“Is polite to others”). These relationship estimates are obtained after controlling for a set of background variables: students’ gender, grade and their socio-economic status (a combined measure including parents’ education, occupational status, home possessions and cultural capital).

¹⁰ Criterion validity represents the extent to which a measure is related to an outcome. If the outcome is assessed at a later time, it is called predictive validity. But when the outcome is assessed at the same time as the measure, it is also called concurrent validity.

Table 4.4. Relation between co-operation and politeness across four measures of informants

		SKILL ESTIMATES			
OLDER COHORT		Co-operation			
Politeness	Students	Parents	Teachers	Triangulated	
Students	0.39	0.12	0.09		0.29
Parents	0.06	0.20	0.01		0.20
Teachers	0.09	0.09	0.60		0.21
Triangulated	0.28	0.20	0.41		0.39
YOUNGER COHORT					
Politeness	Students	Parents	Teachers	Triangulated	
Students	0.26	0.07	0.09		0.22
Parents	0.03	0.23	0.06		0.17
Teachers	0.09	0.16	0.65		0.37
Triangulated	0.22	0.25	0.50		0.44

These findings indicate that a ‘common method’ bias or rater effect exists between the students’ co-operation skill estimate and the students’ politeness behavioural indicator. In other words, the magnitude of the coefficients regarding the relation between co-operation and politeness varies depending on whether the estimates of these two variables are obtained from the same or from different raters. The estimated coefficient for the relation between co-operation and politeness based on older student reports shows a strong relation (coefficient of .39). At the same time, co-operation has a small association with politeness when based on reports provided by the other two raters (.06 and .09 for parents and teachers, respectively). But the associations between co-operation and the politeness behavioural indicator are very strong when both are reported by teachers (.60 and .65 for the older and younger cohorts, respectively) and very weak when reported by students and parents. As could be expected, the triangulated estimate of co-operation has a more stable relationship with all raters of this behavioural indicator compared to the individual raters.

One ‘triangulated’ behavioural indicator representing the overlapping information of the three reports obtained from students, parents and teachers was created in order to reduce the effects of ‘common method’ bias and to create a more valid and objective criterion measure. Correlations of the skill estimates with this criterion should, therefore, be less prone to the issue of ‘common method’ bias, allowing for a more straightforward comparison of strengths of associations. Evaluating the strengths of these relationships with the four measures of co-operation reveals that while student and parent estimates of co-operation are associated with a triangulated indicator of politeness between .20 and .28, the size of this relationship is much bigger in the case of the teacher and the triangulated variable, which vary between .39 and .50. In other words, the triangulated skill measure, constructed from the triangulated reports from students, parents and teachers, better predicts of this particular behaviour (students’ politeness) than student and parent reports on their own and is almost on par with teacher reports.

In order to evaluate whether such findings hold for other behavioural indicators and the most related skills, we created triangulated variables for the nine behavioural indicators that were asked to all three informants. Then, using factor analysis to identify their commonalities, we kept their factor scores. We then ran 36 linear regression models: four models for each of the nine behavioural indicators, with each model estimating the predictive power of one of the four skill estimates (informants). We selected one social and emotional skill that is the most conceptually related to each of the nine behavioural

indicators. Table 4.5 presents regression coefficients of the four skill estimates on these nine behavioural indicators. Coefficients are obtained after accounting for differences in student background variables (gender, grade and family socio-economic status), separately for older and younger cohorts.

Table 4.5. Strength of relationship between skill estimates and behavioural indicators

Behavioural Indicator	Skill Predictor	Students		Parents		Teachers		Triangulated	
		Older	Younger	Older	Younger	Older	Younger	Older	Younger
Can't focus and pay attention for long	Self-control	0.31	0.19	0.19	0.27	0.20	0.54	0.30	0.38
Secretive, does not talk about himself/herself	Sociability	0.13	0.12	0.21	0.13	0.07	0.16	0.23	0.18
Does not like school	Creativity	0.10	0.18	0.11	0.09	0.13	0.15	0.15	0.19
Never gets in fights	Co-operation	0.12	0.10	0.05	0.07	0.34	0.17	0.14	0.16
Can't sit still, is restless	Self-control	0.16	0.06	0.12	0.15	0.19	0.36	0.17	0.21
Gets along with other kids	Co-operation	0.30	0.28	0.23	0.20	0.22	0.19	0.36	0.36
Obedient to parents	Responsibility	0.20	0.25	0.29	0.27	0.34	0.38	0.37	0.38
Is polite to others	Co-operation	0.28	0.22	0.20	0.25	0.40	0.50	0.39	0.44
Physical problems without known medical cause	Stress resistance	0.15	0.08	0.15	0.13	0.08	0.16	0.20	0.17

Note: Estimates are standardised regression coefficients.

These findings are similar to those obtained in the first example in Table 4.4 of the relationship between students' co-operation and their tendency to be polite to others. In particular, the triangulated measure is, on average, the best predictor of the nine behavioural indicators for the older cohort and the second best predictor for the younger cohort. Interestingly, teacher reports, although based on very short scales, consistently show relatively high predictive value related to this set of behavioural indicators. Student and parent reports lag behind with substantially lower associations between their skill estimates and the behavioural indicators.

There are a number of other criterion variables in the Study on Social and Emotional Skills dataset. Figure 4.10 shows the relationship between students' optimism, estimated by four methods and their personal well-being (as reported by students themselves). The strength of the relationship was calculated after controlling for the standard set of background characteristics and a separate set of estimates for each age cohort is included.

Figure 4.10. Relationship between optimism and personal well-being, by type of skill estimate and cohort

In line with previous results, relationships between estimates of optimism and personal well-being are strongest when both measures are based on the reports of the same rater. In this study, students are the only respondents that report on personal well-being. This result indicates the presence of ‘common method’ bias which could then be controlled for by using the triangulated measure of reports from all three informants which lowers the strength of association but is still considerably higher than for parents and teachers. Moreover, associations between the optimism estimate, based on parent and teacher reports and student-reported life satisfaction are substantially lower and likely attenuated due to the multi-source measurement of the two variables (Conway and Lance, 2010^[28]; Conway, 2002^[42]). The triangulated estimate largely avoids the issue of attenuation of association coefficients because the measure is obtained from different sources. Thus, the triangulated estimate could be a more valid estimate of the true relationship between these two constructs in this population. Similar findings are obtained after examining the relationship between students’ optimism and well-being (see Appendix, Table 7.10).

Test anxiety – specifically stress resistance – is another criterion of students’ social and emotional skills. Figure 4.11 presents the size of relationships between stress resistance of students and their school test anxiety (as reported by the students). Here again, measures of association were obtained after controlling for the standard set of background characteristics, separately for each age cohort. Again, these results show that ‘common method’ bias in the form of a common rater effect might lead to an overestimation of the true strength of the relationship between these two constructs. At the same time, potential use of different informants for two constructs could lead to underestimation of the actual association between the two variables (Conway and Lance, 2010^[28]). Such underestimation could be expected due to accumulation of unshared method effects and random errors of the two measures (Conway, 2002^[42]). In other words, the more method biases, such as rater effects, in student, parent and teacher reports are uncorrelated with one another, the more the resulting correlations between their measures will be attenuated. The use of the triangulated skill estimate could significantly reduce the ‘common method’ bias of student

reports, while at the same time it can help to overcome inherent limitations of reports provided by parents and teachers.

Figure 4.11. Relationship between stress resistance and school test anxiety, by type of skill estimate and cohort

Finally, we also examined the relationship between the four measures of students' social and emotional skills and their academic achievement, operationalised through their school grades in four academic subjects – reading/language, math, science and arts. In particular, achievement motivation, curiosity and persistence have the strongest empirical correlations to school grades. Students' school grades in the four different subjects were factor analysed and one generalised measure of school grades was calculated for each student and used in subsequent analyses.

These analyses are especially important when comparing criterion validities of the four skill measures because they circumvent the issue of 'common method' bias. School grades were obtained from administrative records – an independent source of information – rather than from informant reports¹¹. Figure 4.12 presents the results of models including academic achievement and school grades. The relationships between school grades and measures of curiosity and persistence are presented in the Appendix in Figure 7.1 and Figure 7.2. As in the previously presented regression models, the strength of the relationship was calculated after controlling for the standard set of background characteristics, again separately for each cohort.

¹¹ Administrative records, including grades, were provided by School Coordinators, which were usually members of schools' administration.

Figure 4.12. Relationship between achievement motivation and school grades, by type of skill estimate and cohort

These findings again illustrate the validity of the triangulated skill estimates obtained from the three individual reports. The strengths of the relationships between the triangulated measures and school grades are stronger than from any of the three individual skill estimates. Teachers' reports on these three skills have very close criterion validities to those of the triangulated measures but this could be expected as teachers are familiar with the criterion variable – students' school achievement — especially in the subjects that they teach these children in¹². In other words, knowing the students' grades could lead teachers to evaluate their social and emotional skills in a similar way (the “halo effect”), artificially increasing the associations between the two variables in their reports. Also, teachers' grades can be a reflection of teachers' evaluations of students' social and emotional skills, such as their motivation, persistence, curiosity, etc. This common factor in the two measures in question can additionally increase correlations between teacher-reported skill estimates of students and their school grades.

It is also worth noting the difference in the criterion validity of student and parent skill scales across the two cohorts. In particular, skill estimates based on reports from younger students and, especially, their parents are substantially less predictive of students' academic achievement than those from older students and their parents. These results are somewhat different from the findings obtained in previous research, which indicate that achievement motivation is a stronger predictor of academic achievement for younger students compared to older ones (Orhan-Ozen, 207_[47]). One reason might be greater difficulty in providing objective self-reports in the case of 10-year-olds. At the same time, parents of 10-year-olds might be less able to discern their children's motivation than parents of 15-year-olds.

¹² Teachers who report on students' skills are selected on the basis of the degree to which they are involved with students, with priority given to recruiting teachers who best know each of the sampled students.

5. Discussion

The data show that the assessment scales of students, parents and teachers have satisfactory psychometric properties. The reliability coefficients of the scales are generally solid, largely varying between 0.70 to 0.80. The psychometric results for both the younger and the older cohort are similar suggesting that even the younger students are able to provide reliable and consistent self-reports of their social and emotional skills. The only exception, across all three informants, is the scale measuring critical thinking which has unsatisfactory properties. The psychometric characteristics of the scales reported by teachers are remarkably good considering the scales in the teacher questionnaire only consist of three items compared to eight items per scale for the younger students and parents and ten items per scale for the older students.

The correlations between students', parents' and teachers' responses show that there is a relatively moderate overlap in how the different informants report students' social and emotional skills. The responses from all three informants contain measurement error which can be observed from the correlation increase after correction for attenuation. We observe that the correlation is highest between the responses of students and parents, especially for the older students which is consistent with other research.

The correlations of student, parent and teacher assessment scales measuring the same skill indicate a moderate degree of convergent validity. At the same time, correlations of the assessment scales with different scales reported by different raters are rather small, around 0.10, indicating relatively high divergent validity. However, correlations between scales measuring different skills from the same rater are substantially larger (e.g. .33 in the case of older students), indicating a common rater effect.

Multi-Trait Multi-Method matrices were used to examine convergent and divergent validities of the assessment scales while at the same time assessing the size of the method bias in the form of a common rater effect. Two groups of MTMM models were examined, depending on whether or not skill factors were allowed to correlate and whether or not they were coming from the same Big Five domain. Each set of these models provided a number of important insights. Most importantly, they confirmed moderate overlaps between assessment scales across different raters, even after controlling for method bias. This overlap is shown both in models that specify correlations between skill scales coming from the same Big Five domain and in models that mainly contain uncorrelated scales from different domains. These results are largely in line with previous research that uses triangulated reports of students, parents and teachers such as the work by Marsh (2008^[48]). These results indicate a substantial degree of convergent validity which would allow for the combination of the shared information from the three individual reports on each of the student skills.

The MTMM results also indicate the presence of a strong method factor, which in this context represents a common rater effect. Such method bias introduces systematic measurement error in the assessment scales which reduces the divergent validity across scales that are reported by the same rater and are intended to assess different skills. The presence of this strong rater effect justifies the use of the triangulated assessment approach because it allows researchers to detect this method bias and control for it.

Teachers are found to have the strongest method effect, which is probably due to their short assessment scales. Shorter scales reduce scale reliabilities and the amount of concept-

relevant data that is captured by the scales (Credé et al., 2012^[11]). This can lead to the prevalence of the method factor in the variability of these scales.

As expected, older students reported a somewhat larger amount of valid information and have a somewhat lower rater effect than younger students. However, this difference was relatively small and could also be, at least partially, attributed to the longer assessment scales administered to the older cohort. These results show that although older students are able to report slightly more valid information on their social and emotional skills, the difference with their younger counterparts is relatively small. This indicates that 10-year-old students are generally able to provide valid self-report data on their behaviours, thoughts and emotions.

The MTMM findings also show that the scales' convergent validity is strongly related to their reliabilities with correlations of around 0.60. This means that the more reliable student, parent and teacher scales are, the higher their mutual correlations are (i.e. their convergent validities). This finding is in accordance with the general psychometric principle of the reliability of scales being a pre-condition and a ceiling of their validity (Guba, 1981^[49]). However, this relationship with reliability, although strong, still leaves a large amount of variation in scales' convergent validities unexplained. One factor that seems to account for some of the observed differences in convergent validities of scales measuring different skills is their observability in behaviours of students. For example, skills that have the highest convergent validities are sociability and energy and those with lowest convergent validities are metacognition, trust or critical thinking. These results are also in line with previous research and with the general expectation of a higher rater correspondence in cases of skills that are easier to observe in different contexts and by different observers (e.g. (Marsh and Byrne, 1993^[31]; McCrae and Costa, 1987^[44]; McCrae and Costa, 1988^[45])).

Observed convergent validities between student, parent and teacher assessment scales allow us to create a triangulated skill estimate. This triangulated measure mostly related to parent scales (with an average factor loading of around 0.60) and, somewhat less so to student scales (with factor loadings around 0.50) and to teacher estimates (with factor loadings around 0.30). Such results again confirm that teacher reports have the least amount of valid information of the three reports which is expected given the low number of items in the scales. However, it should be noted that in teacher scales were more valid for the younger cohort than for the older while student and parent scales were substantially less valid for the younger cohort in comparison with these scales for the older cohort. This pattern of results indicates that teacher reports might add valuable information in the case of 10-year-old students, especially as they are less able to provide valid self-reports. Also, it should be taken into account that teachers of younger students usually spend more time with them (by teaching multiple subjects) and are hence in a better position to evaluate students' behaviours and skills.

Analysing the criterion validity of the triangulated measure in comparison with the individual student, parent and teacher assessment scales revealed important insights. First, it showed again that there was a strong presence of a common rater effect in all cases where the skill estimate and criterion variable were reported by the same rater. In particular, criterion validities of skill scales were three to five times larger in cases of common raters compared to cases where the two estimates were provided by different raters. Similar patterns of results were obtained in the case of a variety of criterion variables, including behavioural indicators, life satisfaction, personal well-being and test anxiety. The common rater effect was especially pronounced in the case of teachers which is in line with results

of the MTMM analyses. However, the triangulated measure seemed to provide a more balanced estimation of the relationship between skills and their criteria. In particular, the triangulated measure shows lower associations compared to the associations between variables provided by the same raters and higher associations compared to those variables reported by different raters. In this way, the triangulated measure substantially reduces the common rater effect in reports from the same rater. At the same time, it reduces the attenuating effect of multi-source random measurement error in the case of reports provided by different raters which consequently amplifies the signal (i.e. the true relationship between these variables) (Conway and Lance, 2010^[28]; Conway, 2002^[42]). Thus, the triangulated measure seems to help alleviate two undesired sources of measurement error in the analysis of criterion validities: one systematic (common rater effect among the examined variables) that tends to overestimate criterion validity and the other random (multi-source random measurement error) which tends to underestimate the criterion validity of skill scales.

Furthermore, when both the skill measure and criterion variable (i.e. behavioural indicators) are triangulated, the criterion validity of the triangulated measures performs comparatively better than that of individual student and parent reports and is on par with the teacher reports. This result indicates that the triangulated skill estimate better captures the measured skills and, consequently, has a higher criterion (predictive) validity compared to skill estimates based on individual student and parent reports. On the other hand, the fact that teacher reports provide comparable levels of criterion validity might be due to the strong common rater effect present in their reports. This effect is reduced by triangulation of the skill estimates and the criterion variables but is still present because teacher ratings are part of the triangulated measure. In sum, a triangulated estimate of social and emotional skills seems to provide stronger and more valid estimates of criterion validity compared to the three skill estimates based on individual rater reports.

Examining the criterion validity of skill scales in relation to students' school grades provides additional indication of the usefulness of the triangulated assessment approach. These results are especially important because as a criterion they use information provided from administrative school data independent from the three raters. Here again, the triangulated skill measure showed the highest criterion validities followed by teacher and parent reports. Interestingly, these findings indicate that the triangulation approach is especially useful in the case of younger students where the difference in criterion validity estimates is much more pronounced compared to the older students. This is in line with previously reported results which indicate that, in the case of younger students, teacher reports are more valuable while student and parent reports are less valuable. The presented results show that the triangulated measure allowed for alleviation of some of the inherent biases of self- and other-reports, resulting in improved validity of assessment estimates and higher criterion validity estimates. In addition, triangulation has allowed us to examine assessment data in much greater detail and to make more informed choices about the construction of overall skill estimates.

However, as discussed previously, triangulation does not necessarily involve creating a unified measure from three separate methods or sources. For a triangulated approach to be justified (i.e. in terms of investing additional resources), it is not necessary for corresponding scales to significantly overlap in order to create a unified variable that would capture most of the information from the three separate informants. Actually, in such a situation one would not need three separate sources of information, and there would be less benefits for the additional resources invested. Triangulation is more useful when there is no overlap between different sources (as long as each scale provides valid information) or

when there is partial, moderate overlap (which is most often the case). Results observed in our field test data validate the triangulation approach. Scales based on different sources are providing large amount of unique relevant information, and the overlapping part of information is evidently valid, and provides additional information. In such a situation, triangulation offers researchers a wide range of options, either in choosing to use only one, triangulated skill scale or three of them individually, or all four of them. Importantly, apart from the ability to gather the largest amount of valid information on assessed constructs, triangulation also allows for better control of various method biases and random measurement errors, additionally improving the reliability and validity of obtained skill estimates.

At the same time, it is important to take into account a number of considerations and constraints regarding this assessment design. First, this approach requires more resources as it necessitates the administration of assessment scales to multiple groups of raters. In study designs where these groups of respondents would be included anyway (e.g. where parents and teachers are asked to provide contextual information on students' home and school environment), additional costs of asking them to provide ratings of students' skills might not be too large. However, in cases where other groups of respondents are included only for the purposes of providing additional skill ratings, the costs of these additional assessment sources can be substantial. One should also take into account the additional work needed in preparing multiple versions of the assessment scales, handling and analysing the data which is much more intensive compared to assessment scales from one source.

Another important consideration regarding the triangulated approach is the fact that relying on multiple sources of information will inevitably lead to a higher proportion of missing values, both at the respondent and item levels. In the field test of the Study on Social and Emotional Skills, participation rates of students, teachers and parents were 90%, 80% and 50%. This resulted in only around a third of the participating students having information from all three sources. Our analyses of non-response bias (see Method section) shows that missing respondents were not substantially different from those that participated, indicating that obtained results with the triangulated measure would hold for the entire sample. This may not be the case in other situations and if there is a large amount of missing values coupled with a substantial non-response bias, the value of a triangulated approach is reduced.

Finally, it is important to note that the triangulation approach used in this study relies on different sources/informants that might not be considered different methods of assessment as they all rely on providing ratings of typical behaviours. There are other triangulation approaches where assessment methods are fully distinct from one another, e.g. through using direct observation, performance tests, biodata, administrative data, log files ("paradata"), etc.

6. Conclusion

Triangulation – gathering assessment information from three different sources in order to evaluate the same psychological construct – is an approach rarely used, especially in large-scale international assessments. Evidence from the Study on Social and Emotional Skills field test shows that the triangulation of assessment methods can benefit both methodological and substantive outcomes of the study.

Students', parents' and teachers' reports on students' skills overlap to a moderate degree.

Field test data show that the assessment scales of students, parents and teachers have satisfactory psychometric properties, with moderate to high scale reliabilities and internal consistencies. The correlations between the reports from students, parents and teachers show that there is a moderate overlap – in that the different informants report in similar way on students' social and emotional skills. The Multi-Trait Multi-Method (MTMM) matrices' results provide further insights into the interrelations between skill measures across different raters. They confirm relatively moderate overlaps between the assessment scales across the different raters, even after controlling for method bias. These results allow us to combine the shared information from the three individual reports on each of the student skills.

Students', parents' and teachers' reports on students' skills contain a strong rater bias.

The MTMM results indicate a strong method factor, which in this context represents a common rater effect – the increased tendency of a rater to provide similar ratings across different measures. It is strongest in the case of teachers. This strong rater effect justifies the triangulated assessment approach because it allows us to detect this method bias and to control for it.

The agreement between raters relates to the scale's reliability and the observability of the measured construct.

Some of the observed differences in convergent validities of scales measuring different skills are strongly related to the reliabilities of the scales. In addition, the degree of overlap between scales seems to depend on how easy it is to observe certain behaviours.

The triangulation approach significantly reduces the inherent limitations of self- and other-reports, thus providing a more valid estimate of the relationship between skills and life outcomes.

After creating the triangulated skill estimate, comparing the criterion validity of the triangulated measure with the individual student, parent and teacher assessment scales reveal important insights. The triangulated measure seems to control for two undesired sources of measurement error while analysing criterion validities. The first one is systematic, in the form of a common rater effect among the examined variables that tends to overestimate criterion validity when both variables are reported by the same rater. The second one is random (i.e. multi-source random measurement error) which tends to underestimate the criterion validity of skill scales when measured by different raters. In sum, a triangulated estimate of social and emotional skills provides stronger and more valid estimates of criterion validity compared to the three skill estimates based on individual rater reports.

These results show that the triangulated approach implemented in the field test identified and reduced some of the inherent biases of self- and other-reports, resulting in improved validity of assessment estimates and higher criterion validity estimates. In addition, triangulation has allowed us to examine assessment data in much greater detail and to make more informed choices about the construction of overall skill estimates. Importantly, triangulation still allows the separate use of all skill estimates obtained from independent sources. This allows a high degree of flexibility for researchers to use independent or triangulated estimates in the construction of a final skill estimate or comparing measures across studies.

Although there are benefits to triangulation, there are still certain conditions and costs to keep in mind.

The triangulation method requires more resources and accrues more cost as it is necessary to administer the assessment scales to multiple groups of raters, especially if they are only added to the study for the purpose of providing assessment data. There is also the additional work needed in terms of assessment preparation, data collection and data analysis. Another constraint that needs to be taken into account is the possibility of a large amount of missing values. Missing data, when coupled with a possible non-response bias, potentially reduces the applicability of a triangulated approach. Achieving high response rates is especially difficult with some groups of raters, such as parents. These are all important aspects to take into consideration before trying the triangulation approach.

Directions for future research.

The triangulation of assessment approach offers an important avenue of exploration and possible improvement of the psychometric quality of assessment measures. However, the relatively high costs and resource requirements have limited the applications of this methodology and in consequence learning about its features and benefits. Increasing the number and scope of studies that implement the triangulation approach by using different raters, assessing different constructs (e.g. skills, knowledge, attitudes, values, interests, etc.) and/or different subpopulations (e.g. children, students, adults, employees.) is needed. Using different assessment methods (rather than just different informants using the same method), such as direct observation, performance tests, ratings, biodata, “paradata”, administrative records, etc. could further enrich the analysis, as the method-specific biases would be easier to identify and control. These studies could provide valuable insights and offer new opportunities to develop better assessment instruments and approaches.

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. [29]
- Achenback, T. et al. (2008), “Multicultural assessment of child and adolescent psychopathology with ASEBA and SDQ instruments: Research findings, applications, and future direction”, *Journal of Child Psychology and Psychiatry*, Vol. 49, pp. 251-275. [46]
- Achenback, T., S. McConaughy and C. Howell (1987), “Child/adolescent behavioural and emotional problems: Implications of cross-informant correlations for situational specificity.”, *Psychological Bulletin*, Vol. 101, pp. 213-232. [32]
- Carlson, E., S. Vazire and R. Furr (2011), “Meta-Insight: Do People Really Know How Others See Them?”, *Journal of Personality and Social Psychology*, Vol. 101/4, pp. 831-846, <http://dx.doi.org/10.1037/a0024297>. [7]
- Cohen, L., L. Manion and K. Morrison (2002), *Research Methods in Education*, Routledge, London, <https://www.taylorfrancis.com/books/9780203224342> (accessed on 19 July 2019). [1]
- 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>. [3]
- Conway, J. (2002), “Method variance and method bias in industrial and organizational psychology”, in S.G. Rogelberg (ed.), *Handbook of Research Methods in Industrial and Organizational Psychology*, Blackwell Publishing, <https://psycnet.apa.org/record/2002-02945-017> (accessed on 16 July 2019). [42]
- Conway, J. and C. Lance (2010), “What reviewers should expect from authors regarding common method bias in organizational research”, *Journal of Business and Psychology*, Vol. 25/3, pp. 325-334, <http://dx.doi.org/10.1007/s10869-010-9181-6>. [28]
- Credé, M. et al. (2012), “An evaluation of the consequences of using short measures of the Big Five personality traits”, *Journal of Personality and Social Psychology*, Vol. 102/4, pp. 874-888, <https://psycnet.apa.org/buy/2012-04359-001> (accessed on 19 July 2019). [11]
- De Los Reyes, A. and A. Kazdin (2005), “Informant discrepancies in the assessment of childhood psychopathology: A critical review, theoretical framework, and recommendations for further study”, *Psychological Bulletin*, Vol. 131, pp. 483-509. [33]
- Gagnon, S., R. Nagle and A. Nickerson (2007), “Parent and teacher ratings of peer interactive play and social-emotional development of preschool children at risk”, *Journal of Early Intervention*, Vol. 29, pp. 228-242. [39]

- Göllner, R. et al. (2017), “Whose “Storm and Stress” Is It? Parent and Child Reports of Personality Development in the Transition to Early Adolescence”, *Journal of Personality*, Vol. 85/3, pp. 376-387, <http://dx.doi.org/10.1111/jopy.12246>. [17]
- Gonyea, R. (2005), “Self-reported data in institutional research: Review and recommendations”, *New Directions for Institutional Research*, Vol. 2005/127, pp. 73-89, <http://dx.doi.org/10.1002/ir.156>. [2]
- Grietens, H. et al. (2004), “Comparison of mothers’, fathers’, and teachers’ reports on problem behaviour in 5- to 6-year-old children”, *Journal of Psychopathology and Behavioural Assessment*, Vol. 26, pp. 137-146. [35]
- Grimm, L. and P. Yarnold (1995), *Reading and understanding multivariate statistics*, American Psychological Association. [52]
- Guba, E. (1981), “Criteria for Assessing the Trustworthiness of Naturalistic Inquiries”, *Educational Technology Research and Development*, Vol. 29/2, pp. 75-91. [49]
- Haager, D., C. Watson and D. Willows (2008), “Parent, Teacher, Peer, and Self-Reports of the Social Competence of Students with Learning Disabilities”, *Journal of Learning Disabilities*, Vol. 28/4, pp. 205-215, <http://dx.doi.org/10.1177/002221949502800403>. [24]
- Heine, S., E. Buchtel and A. Norenzayan (2008), “What do cross-national comparisons of personality traits tell us? The case of conscientiousness: Research report”, *Psychological Science*, Vol. 19/4, pp. 309-313, <http://dx.doi.org/10.1111/j.1467-9280.2008.02085.x>. [4]
- Jepsen, M., K. Gray and J. Taffe (2012), “Agreement in multi-informant assessment of behaviour and emotional problems and social functioning in adolescents with autistic and Asperger’s disorder”, *Research in Autism Spectrum Disorders*, Vol. 6/3, pp. 1091-1098, <http://dx.doi.org/10.1016/j.rasd.2012.02.008>. [25]
- Kamphaus, R. and P. Frick (1996), *Clinical assessment of child and adolescent personality and behavior*, Allyn and Bacon. [36]
- 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>. [9]
- Karadag, E. (ed.) (2007), *The effect of motivation on student achievement*, Springer. [47]
- Krosnick, J. (1999), “Survey research.”, *Annual review of psychology*, Vol. 50, pp. 537-567, <http://dx.doi.org/10.1146/annurev.psych.50.1.537>. [5]
- Lindqvist, E. and R. Vestman (2011), “The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the The Labor Market Returns to Cognitive and Noncognitive Ability: Evidence from the Swedish Enlistment”, *Source: American Economic Journal: Applied Economics American Economic Journal: Applied Economics*, <http://dx.doi.org/doi:10.1257/app.3.1.101>. [12]

- Major, S., M. Seabra-Santos and R. Martin (2015), “Are we talking about the same child? Parent-teacher ratings of preschoolers’ social-emotional behaviours”, *Psychology in the Schools*, Vol. 58/2, pp. 789-799. [30]
- Marsh, H. (2008), “A multidimensional, hierarchical model of self-concept: An important facet of personality”, in G.J. Boyle, G. Matthews, D. (ed.), *The SAGE Handbook of Personality Theory and Assessment, Volume I: Personality Theories and Models*, SAGE Publications, Thousand Oaks, CA, [https://books.google.fr/books?hl=fr&lr=&id=s2f8jIpPeU4C&oi=fnd&pg=PA447&dq=\).+A+multiparametric,+hierarchical+model+of+self-concept:+An+important+facet+of+personality.+In+G.+J.+Boyle,+G.+Matthews,+D.+H.+Saklofske,+G.+J.+Boyle,+G.+Matthews+%26+D.+H.+Saklo](https://books.google.fr/books?hl=fr&lr=&id=s2f8jIpPeU4C&oi=fnd&pg=PA447&dq=).+A+multiparametric,+hierarchical+model+of+self-concept:+An+important+facet+of+personality.+In+G.+J.+Boyle,+G.+Matthews,+D.+H.+Saklofske,+G.+J.+Boyle,+G.+Matthews+%26+D.+H.+Saklo) (accessed on 19 July 2019). [48]
- Marsh, H. (1990), “A multidimensional, hierarchical model of self-concept: Theoretical and empirical justification”, *Educational Psychology Review*, Vol. 2/2, pp. 77-172, <http://dx.doi.org/10.1007/BF01322177>. [20]
- Marsh, H. and B. Byrne (1993), “Confirmatory factor analysis of multitrait-multimethod self-concept data: Between-group and within-group invariance constraints”, *Multivariate Behavioral Research*, Vol. 28/3, pp. 313-449, http://dx.doi.org/10.1207/s15327906mbr2803_2. [31]
- Mathison, S. (1988), “Why Triangulate?”, *Educational Researcher*, Vol. 17/2, pp. 13-17, <http://dx.doi.org/10.3102/0013189X017002013>. [14]
- McCrae, R. (2015), “A more nuanced view of reliability: specificity in the trait hierarchy”, *Personality and Social Psychology Review*, Vol. 19/2, pp. 97-112. [43]
- McCrae, R. and P. Costa (1988), “Recalled parent-child relations and adult personality”, *Journal of Personality*, Vol. 56/2, pp. 417-434, <http://dx.doi.org/10.1111/j.1467-6494.1988.tb00894.x>. [45]
- McCrae, R. and P. Costa (1987), “Validation of the five-factor model of personality across instruments and observers.”, *Journal of Personality and Social Psychology*, Vol. 52/1, pp. 81-90, <http://dx.doi.org/10.1037/0022-3514.52.1.81>. [44]
- Merrel, K. (2008), *Behavioral, social, and emotional assessment of children and adolescents*, Erlbaum. [40]
- Miller, F. et al. (2018), “Methods matter: A multi-trait multi-method analysis of student behavior”, *Journal of School Psychology*, Vol. 68, pp. 53-72, <http://dx.doi.org/DOI:10.1016/j.jsp.2018.01.002> (accessed on 11 July 2019). [27]
- Newgent, R. et al. (2009), “Differential Perceptions of Bullying in the Schools: A Comparison of Student, Parent, Teacher, School Counselor, and Principal Reports”, *Journal of School Counseling*, Vol. 7/38, pp. 1-33. [23]
- Paulhus, D. and S. Vazire (2007), “The Self-Report Method”, in *Handbook of Research Methods in Personality Psychology*, Guilford, New York. [10]

- Perlesz, A. and J. Lindsay (2003), "Methodological triangulation in researching families: Making sense of dissonant data", *International Journal of Social Research Methodology: Theory and Practice*, Vol. 6/1, pp. 25-40, <http://dx.doi.org/10.1080/13645570305056>. [15]
- Power, T. et al. (1998), "Evaluating attention deficit hyperactivity disorder using multiple informants: The incremental utility of combining teacher with parent reports.", *Psychological Assessment*, Vol. 10/3, pp. 250-260, <http://dx.doi.org/10.1037/1040-3590.10.3.250>. [19]
- Ramsey, C. et al. (2016), "School climate: perceptual differences between students, parents, and school staff", *Schools Effectiveness and Schools Improvement*, Vol. 27/4, pp. 629-641. [34]
- Renk, K. and V. Phares (2004), "Cross-informant ratings of social competence in children and adolescents", *Clinical Psychology Review*, Vol. 24/2, pp. 239-254, <http://dx.doi.org/10.1016/j.cpr.2004.01.004>. [26]
- Rothbauer, P. (2008), "Triangulation", *The SAGE Encyclopedia of Qualitative Research Methods*, pp. 892-894, [https://books.google.fr/books?hl=fr&lr=&id=byh1AwAAQBAJ&oi=fnd&pg=PP1&dq=Rothbauer,+P.+\(2008\)+%27Triangulation%27,+in+Lisa+Given+\(Ed.\),+The+SAGE+Encyclopedia+of+Qualitative+Research+Methods.+Thousand+Oaks,+CA:+Sage,+pp.+892%E2%80%93934+&ots=LOVXPJ3G2l&sig=M](https://books.google.fr/books?hl=fr&lr=&id=byh1AwAAQBAJ&oi=fnd&pg=PP1&dq=Rothbauer,+P.+(2008)+%27Triangulation%27,+in+Lisa+Given+(Ed.),+The+SAGE+Encyclopedia+of+Qualitative+Research+Methods.+Thousand+Oaks,+CA:+Sage,+pp.+892%E2%80%93934+&ots=LOVXPJ3G2l&sig=M) (accessed on 19 July 2019). [50]
- Segal, C. (2012), "Working When No One Is Watching: Motivation, Test Scores, and Economic Success", *Management Science*, Vol. 58/8, pp. 1438-1457, <http://dx.doi.org/10.1287/mnsc.1110.1509>. [13]
- Stanger, C. and M. Lewis (1993), "Agreement among Parents, Teachers, and Children on Internalizing and Externalizing Behavior Problems", *Journal of Clinical Child Psychology*, Vol. 22/1, pp. 107-115. [22]
- Strickland, J., J. Hopkins and K. Keenan (2012), "Mother-teacher agreement on preschoolers' symptoms of ODD and CD: Does context matter?", *Journal of Abnormal Child Psychology*, Vol. 40, pp. 933-943. [41]
- Tackett, J. (2011), "Parent informants for child personality: Agreement, discrepancies, and clinical utility", *Journal of Personality Assessment*, Vol. 93/6, pp. 539-544, <http://dx.doi.org/10.1080/00223891.2011.608763>. [16]
- Treutler, C. and C. Epkins (2003), "Are discrepancies among child, mother, and father reports on children's behavior related to parents' psychological symptoms and aspects of parent-child relationships?", *Journal of Abnormal Child Psychology*, Vol. 31, pp. 13-27. [37]
- Tsolou, O. and V. Margaritis (2013), "Social and Emotional Learning Competencies and Cross-Thematic Curriculum-Related Skills of Greek Students: A Multifactorial and Triangulation Analysis", *Journal of Educational Research and Practice*, Vol. 3/1, pp. 65-78, <http://dx.doi.org/10.5590/JERAP.2013.03.1.05>. [21]
- Vazire, S. (2010), "Who Knows What About a Person? The Self-Other Knowledge Asymmetry (SOKA) Model", *Journal of Personality and Social Psychology*, Vol. 98/2, pp. 281-300, <http://dx.doi.org/10.1037/a0017908>. [8]

- Vazire, S. and E. Carlson (2010), "Self-Knowledge of Personality: Do People Know Themselves?", *Social and Personality Psychology Compass*, Vol. 4/8, pp. 605-620, <http://dx.doi.org/10.1111/j.1751-9004.2010.00280.x>. [6]
- Verhulst, F., H. Koot and J. Ende (1994), "Differential predictive value of parents' and teachers' reports of children's problem behaviors: A longitudinal study", *Journal of Abnormal Child Psychology*, Vol. 22/5, pp. 531-546, <http://dx.doi.org/10.1007/BF02168936>. [18]
- Wiberg, S. et al. (eds.) (2019), *Controlling acquiescence Bias with Multidimensional IRT Modeling*, Springer International Publishing. [51]
- Winsler, A. and G. Wallace (2002), "Behavior problems and social skills in preschool children: Parent-teacher agreement and relations with classroom observations", *Early Education & Development*, Vol. 13, pp. 41-58. [38]

7. Appendices

Table 7.1. Raw skill scale descriptive statistics by respondent and cohort

	Students			Parents			Teachers		
	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev	Obs	Mean	Std. Dev
10 year olds									
Assertiveness	6 188	3.00	0.83	3 874	3.36	0.72	5 337	3.03	1.01
Co-operation	5 958	4.04	0.62	3 820	4.11	0.51	4 996	3.85	0.76
Creativity	5 908	3.89	0.63	3 805	3.88	0.56	4 948	3.46	0.82
Critical thinking	5 678	3.36	0.45	3 826	3.42	0.43	4 860	3.41	0.62
Curiosity	5 616	4.02	0.63	3 707	4.12	0.54	4 886	3.87	0.79
Self-efficacy	5 519	3.85	0.62	3 694	3.81	0.60	4 755	3.59	0.77
Emotional control	5 497	2.64	0.71	3 696	2.62	0.68	4 892	2.36	0.88
Empathy	5 433	3.75	0.58	3 693	3.78	0.51	5 033	3.65	0.67
Energy	5 670	2.41	0.65	3 869	2.25	0.58	5 031	2.46	0.90
Metacognition	5 381	3.66	0.58	3 679	3.45	0.55	5 002	3.35	0.78
Achiev. motivation	5 696	3.93	0.60	3 702	3.74	0.63	4 967	3.45	0.91
Optimism	5 840	3.87	0.66	3 866	3.99	0.52	5 320	3.79	0.69
Persistence	5 741	3.93	0.64	3 834	3.61	0.72	5 367	3.64	1.00
Responsibility	5 443	2.30	0.68	3 620	2.38	0.70	4 855	2.30	0.79
Self-control	6 021	3.72	0.66	3 838	3.47	0.66	4 754	3.32	0.83
Sociability	5 859	3.73	0.60	3 697	3.78	0.55	4 869	3.49	0.58
Stress resistance	5 957	2.87	0.77	3 828	2.82	0.67	4 955	2.79	0.79
Tolerance	5 623	3.89	0.66	3 830	3.94	0.56	5 011	3.45	0.83
Trust	5 994	3.57	0.73	3 851	3.62	0.56	4 937	3.48	0.67
15-year-olds									
Assertiveness	6,224	3.13	0.79	3 289	3.30	0.70	5 028	2.90	0.90
Co-operation	5 619	3.93	0.53	3 246	4.06	0.52	4 715	3.80	0.67
Creativity	6 070	3.65	0.59	3 238	3.73	0.59	4 691	3.35	0.73
Critical thinking	6 255	3.69	0.43	3 257	3.58	0.42	4 561	3.30	0.57
Curiosity	6 326	3.76	0.57	3 105	3.90	0.61	4 597	3.70	0.75
Self-efficacy	5 572	3.67	0.60	3 118	3.79	0.60	4 468	3.54	0.70
Emotional control	6 237	2.76	0.72	3 104	2.57	0.69	4 613	2.39	0.79
Empathy	6 325	3.77	0.53	3 090	3.75	0.51	4 731	3.53	0.59
Energy	6 125	2.74	0.66	3 298	2.47	0.61	4 727	2.64	0.85
Metacognition	6 277	3.59	0.49	3 077	3.55	0.52	4 666	3.32	0.68
Achiev. motivation	6 512	3.63	0.55	3 100	3.72	0.63	4 693	3.38	0.84
Optimism	6 469	3.46	0.75	3 285	3.84	0.58	5 002	3.60	0.65
Persistence	6 195	3.65	0.64	3 264	3.68	0.69	5 052	3.56	0.91
Responsibility	6 143	2.35	0.58	3 032	2.29	0.70	4 555	2.37	0.75
Self-control	5 894	3.55	0.58	3 260	3.61	0.61	4 471	3.35	0.74
Sociability	6 362	3.53	0.62	3 093	3.50	0.62	4 567	3.38	0.58
Stress resistance	6 179	3.06	0.78	3 261	2.71	0.66	4 646	2.84	0.70
Tolerance	6 444	3.84	0.60	3 257	3.85	0.60	4 727	3.28	0.76
Trust	6 412	3.14	0.70	3 273	3.48	0.60	4 648	3.34	0.59

Table 7.2. Comparison of student characteristics across informants

	STUDENTS					PARENTS					TEACHERS				
	Mean IN	Mean OUT	Difference (p-value)	n IN	n OUT	Mean IN	Mean OUT	Difference (p-value)	n IN	n OUT	Mean IN	Mean OUT	Difference (p-value)	n IN	n OUT
Gender	0.49	0.55	0.059	13 221	262	0.49	0.51	0.033	7 169	6 314	0.49	0.51	0.156	10 434	3 049
Age in months	162	147	0.000	13 238	253	159	164	0.000	7 161	6 330	161	163	0.000	10 431	3 060
Grade 10	4.93	4.86	0.554	6 611	192	4.89	4.97	0.009	3 851	2 952	4.89	5.08	0.000	5 342	1 461
Grade 15	9.57	8.94	0.004	6 624	63	9.59	9.53	0.078	3 303	3 384	9.54	9.63	0.015	5 098	1 589
Years in this school	4.19	4.71	0.007	13 076	255	4.22	4.18	0.354	7 068	6 263	4.12	4.48	0.000	10 318	3 013
Country born student	1.06	1.13	0.001	12 969	270	1.03	1.10	0.000	7 230	6 009	1.05	1.10	0.000	10 285	2 954
Country born mother	1.17	1.24	0.012	13 041	270	1.08	1.28	0.000	7 234	6 077	1.15	1.25	0.000	10 341	2 970
Country born father	1.17	1.25	0.002	13 953	263	1.08	1.28	0.000	7 197	6 019	1.15	1.26	0.000	10 271	2 945
SES	-0.01	-0.20	0.039	12 115	137	-0.03	0.01	0.016	6 704	5 548	-0.01	-0.03	0.370	9 552	2 700
Language at home	1.14	1.18	0.108	13 039	250	1.11	1.19	0.000	7 105	6 184	1.14	1.17	0.000	10 298	2 991

Note: The table shows comparisons of student characteristics. Mean IN (sample) for students refers to students for whom we have responses on at least one social and emotional skill. Mean OUT (of sample) for students refers to students for whom we do not have any responses on social and emotional skills. The same logic applies to parents and teachers. The difference column shows the results of a t-test that compare the mean of student characteristics between the IN sample and the OUT of sample groups for students, parents and teachers.

Table 7.3. Comparison of student characteristics across informants II

	Students (IN)	Parents (IN)	Teachers (IN)
	margins	margins	margins
Males	-0.032	-0.052*	-0.041
	-0.068	-0.024	-0.026
Age_months	0.000	-0.006**	-0.001
	-0.002	-0.001	-0.001
Grade	0.081**	0.023**	-0.017*
	-0.023	-0.009	-0.010
Years_in_this_school	-0.010	-0.002	-0.028**
	-0.012	-0.004	-0.004
Student is born abroad	-0.002	0.151*	-0.094
	-0.165	-0.063	-0.061
Mother is born abroad	-0.064	-0.544**	-0.167**
	-0.138	-0.050	-0.053
Father is born abroad	-0.046	-0.570**	-0.313**
	-0.135	-0.049	-0.051
SES	0.071*	-0.060**	-0.002
	-0.034	-0.012	-0.013
Language spoken at home is not the same as in the survey	0.071	0.027	0.106*
	-0.115	-0.040	-0.042
Observations	11,570	11,570	11,570

Note: The table shows marginal effects of probit regressions. Standard errors are in parentheses and represent ** p<0.01, * p<0.05. The dependent variable represents whether the respondent is included in the sample or not.

Table 7.4. Comparison of student characteristics across informants III

	STUDENTS		PARENTS			TEACHERS		
	Mean	N	Mean	N	p-value: S-P	Mean	N	p-value: S-T
Gender	0.49	13 221	0.49	7 169	0.307	0.49	10 434	0.741
Age in months	162	13 238	159	7 161	0.000	161	10 431	0.041
Grade 10	4.93	6 611	4.89	3 851	0.135	4.89	5 342	0.049
Grade 15	9.57	6 624	9.59	3 303	0.413	9.54	5 098	0.318
Years in this school	4.19	13 076	4.22	7 068	0.457	4.12	10 318	0.059
Country born student	1.06	12 969	1.03	7 230	0.000	1.05	10 285	0.003
Country born mother	1.17	13 041	1.08	7 234	0.000	1.15	10 341	0.000
Country born father	1.17	12 953	1.08	7 197	0.000	1.15	10 271	0.000
SES	-0.01	12 115	-0.03	6 704	0.151	-0.01	9 552	0.868
Language at home	1.14	13 039	1.11	7 105	0.000	1.14	10 298	0.139

Note: The table shows the mean and the number of observations for student characteristics for participating students, as well as for students of participating parents and teachers. The difference column shows the results of t-tests that compare the means between participating students and students of participating parents (S-P), as well as between participating students and students of participating teachers (S-T).

Table 7.5. Mean responses and their standard deviations in skill scales between all three informants

	STUDENTS			PARENTS			TEACHERS		
	Mean	Standard Deviation	N	Mean	Standard Deviation	N	Mean	Standard Deviation	N
ASS	3.07	0.82	13 701	3.23	0.61	7 559	2.96	0.96	11 692
COO	3.97	0.60	13 705	3.63	0.44	7 566	3.81	0.72	11 703
CRE	3.76	0.63	13 702	3.46	0.44	7 560	3.39	0.79	11 675
CRI	3.53	0.48	13 700	3.31	0.41	7 564	3.35	0.61	11 247
CUR	3.88	0.63	13 710	3.78	0.52	7 571	3.77	0.78	11 670
EFF	3.76	0.63	13 698	3.72	0.49	7 552	3.56	0.75	11 676
EMO	3.29	0.72	13 702	2.94	0.44	7 550	3.60	0.84	11 675
EMP	3.75	0.56	13 704	3.41	0.43	7 562	3.57	0.64	11 680
ENE	3.43	0.68	13 715	3.02	0.45	7 570	3.45	0.88	11 689
MET	3.61	0.55	13 704	3.30	0.54	7 549	3.33	0.75	11 684
MOT	3.77	0.60	13 705	3.47	0.47	7 552	3.40	0.89	11 693
OPT	3.66	0.73	13 713	3.54	0.44	7 562	3.68	0.69	11 671
PER	3.78	0.66	13 701	3.07	0.38	7 557	3.58	0.97	11 692
RES	3.66	0.64	13 703	2.78	0.53	7 554	3.12	0.52	11 647
SEL	3.63	0.63	13 704	3.48	0.58	7 558	3.32	0.80	11 673
SOC	3.63	0.63	13 713	3.58	0.50	7 564	3.57	0.59	11 680
STR	3.04	0.78	13 701	2.95	0.52	7 547	3.18	0.75	11 666
TOL	3.85	0.64	13 704	3.69	0.51	7 557	3.35	0.80	11 680
TRU	3.35	0.75	13 701	3.41	0.50	7 523	3.39	0.64	11 675

Note: The table shows the mean responses to all of the items belonging to individual assessment scales, their standard deviation and the number of observations per assessment scale for the three groups of respondents: students, parents and teachers. The response scale of all assessment items ranges from 1 (“Completely disagree”) to 5 (“Completely agree”). Answers to the negatively formulated items (e.g. “I give up easily”) were reversed before the calculation of the means.

Table 7.6. Correlations between parent (x-axis) and student (y-axis) reports of student skills for the older cohort

	ASS	COO	CRE	CRI	CUR	EFF	EMO	EMP	ENE	MET	MOT	OPT	PER	RES	SEL	SOC	STR	TOL	TRU
ASS	0.54	0.12	0.21	0.26	0.18	0.27	0.06	0.21	0.22	0.15	0.17	0.18	0.09	0.04	0.08	0.27	0.19	0.13	0.04
COO	0.06	0.36	0.11	0.12	0.17	0.19	0.20	0.35	0.21	0.19	0.23	0.21	0.21	0.24	0.20	0.25	0.10	0.13	0.18
CRE	0.25	0.16	0.40	0.32	0.34	0.27	0.01	0.22	0.17	0.21	0.22	0.19	0.08	0.03	0.10	0.16	0.11	0.25	0.00
CRI	0.24	0.05	0.29	0.46	0.32	0.25	-0.02	0.14	0.06	0.22	0.19	0.07	0.11	0.10	0.11	0.01	0.12	0.23	-0.12
CUR	0.23	0.24	0.40	0.45	0.58	0.40	0.10	0.29	0.25	0.39	0.46	0.26	0.27	0.21	0.29	0.12	0.12	0.39	-0.02
EFF	0.31	0.19	0.28	0.30	0.31	0.42	0.18	0.20	0.33	0.30	0.32	0.33	0.24	0.18	0.22	0.25	0.25	0.17	0.03
EMO	0.01	0.10	0.02	0.04	0.06	0.14	0.45	0.02	0.15	0.07	0.06	0.19	0.10	0.14	0.07	0.08	0.33	0.03	0.11
EMP	0.17	0.31	0.15	0.20	0.15	0.17	0.07	0.49	0.15	0.16	0.17	0.12	0.12	0.17	0.15	0.25	0.05	0.16	0.21
ENE	0.27	0.20	0.17	0.12	0.20	0.28	0.17	0.15	0.54	0.18	0.21	0.34	0.16	0.10	0.15	0.38	0.23	0.08	0.05
MET	0.17	0.18	0.23	0.31	0.28	0.26	0.10	0.29	0.17	0.36	0.30	0.15	0.22	0.24	0.27	0.09	0.07	0.19	0.06
MOT	0.26	0.24	0.27	0.29	0.41	0.42	0.14	0.27	0.39	0.43	0.55	0.27	0.42	0.36	0.38	0.14	0.12	0.20	0.01
OPT	0.15	0.18	0.14	0.10	0.20	0.30	0.28	0.11	0.32	0.19	0.20	0.42	0.17	0.13	0.16	0.28	0.27	0.12	0.08
PER	0.20	0.22	0.25	0.30	0.36	0.41	0.19	0.23	0.34	0.37	0.45	0.26	0.43	0.39	0.34	0.14	0.20	0.17	0.04
RES	0.18	0.25	0.21	0.26	0.27	0.39	0.19	0.30	0.30	0.40	0.40	0.23	0.41	0.45	0.37	0.11	0.16	0.12	0.05
SEL	0.09	0.24	0.20	0.29	0.30	0.34	0.19	0.27	0.20	0.42	0.37	0.17	0.36	0.37	0.44	-0.02	0.09	0.15	0.02
SOC	0.23	0.21	0.05	-0.01	0.02	0.11	0.11	0.20	0.30	-0.04	0.03	0.24	-0.01	0.01	-0.06	0.65	0.17	0.04	0.19
STR	0.13	0.03	0.07	0.12	0.05	0.18	0.28	-0.07	0.18	0.04	-0.01	0.20	0.01	0.01	-0.01	0.11	0.44	-0.01	0.04
TOL	0.21	0.22	0.32	0.34	0.41	0.29	0.10	0.32	0.16	0.27	0.34	0.22	0.18	0.15	0.20	0.18	0.09	0.47	-0.01
TRU	0.00	0.13	-0.10	-0.06	-0.09	0.00	0.26	0.08	0.07	-0.02	-0.01	0.09	0.05	0.12	0.00	0.17	0.13	-0.06	0.40

Note: Correlations are corrected for attenuation.

Table 7.7. Correlations between parent (x-axis) and student (y-axis) reports of student skills for younger cohort

	ASS	COO	CRE	CRI	CUR	EFF	EMO	EMP	ENE	MET	MOT	OPT	PER	RES	SEL	SOC	STR	TOL	TRU
ASS	0.42	-0.04	0.10	0.16	0.06	0.14	-0.02	0.00	0.06	0.10	0.11	0.04	0.05	-0.04	0.01	0.07	0.06	0.07	0.01
COO	0.11	0.30	0.15	0.15	0.16	0.23	0.24	0.23	0.16	0.20	0.26	0.19	0.24	0.26	0.24	0.20	0.16	0.09	0.13
CRE	0.23	0.14	0.32	0.24	0.26	0.24	0.11	0.14	0.18	0.18	0.22	0.15	0.16	0.11	0.14	0.16	0.14	0.18	0.02
CRI	0.27	-0.02	0.24	0.40	0.15	0.21	0.08	0.03	0.07	0.15	0.13	0.00	0.14	0.09	0.07	0.03	0.17	0.06	0.01
CUR	0.21	0.24	0.32	0.28	0.45	0.34	0.14	0.20	0.26	0.31	0.39	0.24	0.28	0.24	0.27	0.20	0.12	0.26	0.00
EFF	0.27	0.19	0.22	0.22	0.24	0.33	0.17	0.16	0.23	0.24	0.31	0.25	0.27	0.24	0.23	0.18	0.20	0.13	0.07
EMO	0.04	0.13	0.05	0.11	0.02	0.16	0.32	0.12	0.10	0.14	0.16	0.12	0.19	0.21	0.17	0.06	0.22	0.02	0.13
EMP	0.20	0.21	0.18	0.20	0.13	0.17	0.17	0.29	0.15	0.15	0.20	0.10	0.16	0.19	0.11	0.18	0.15	0.15	0.19
ENE	0.24	0.18	0.17	0.17	0.21	0.26	0.17	0.16	0.37	0.19	0.27	0.22	0.20	0.19	0.15	0.28	0.19	0.11	0.09
MET	0.23	0.14	0.17	0.18	0.20	0.24	0.13	0.15	0.13	0.25	0.26	0.14	0.21	0.16	0.21	0.09	0.10	0.12	0.02
MOT	0.27	0.22	0.23	0.27	0.29	0.34	0.21	0.23	0.24	0.32	0.40	0.20	0.33	0.28	0.27	0.15	0.18	0.16	0.07
OPT	0.18	0.24	0.20	0.16	0.23	0.29	0.23	0.20	0.31	0.20	0.26	0.35	0.21	0.23	0.20	0.28	0.19	0.18	0.09
PER	0.17	0.17	0.17	0.23	0.21	0.29	0.21	0.17	0.22	0.23	0.31	0.15	0.33	0.33	0.24	0.10	0.20	0.07	0.09
RES	0.15	0.20	0.11	0.23	0.11	0.23	0.23	0.17	0.12	0.22	0.27	0.10	0.29	0.36	0.26	0.06	0.18	0.04	0.14
SEL	0.11	0.21	0.20	0.19	0.24	0.29	0.20	0.15	0.15	0.28	0.31	0.21	0.28	0.28	0.33	0.08	0.11	0.10	0.00
SOC	0.18	0.22	0.11	0.11	0.15	0.16	0.15	0.23	0.33	0.05	0.14	0.23	0.11	0.14	0.02	0.53	0.18	0.17	0.21
STR	0.09	0.06	0.05	0.13	0.04	0.12	0.20	0.03	0.12	0.06	0.08	0.12	0.09	0.13	0.06	0.10	0.29	0.05	0.12
TOL	0.20	0.18	0.23	0.20	0.27	0.24	0.12	0.15	0.17	0.20	0.24	0.17	0.18	0.12	0.15	0.18	0.10	0.26	0.03
TRU	0.08	0.15	0.03	0.07	0.00	0.06	0.14	0.12	0.09	0.03	0.07	0.07	0.08	0.11	0.03	0.18	0.12	0.03	0.28

Note: Correlations are corrected for attenuation.

Table 7.8. Correlations between teacher (x-axis) and student (y-axis) reports of student skills for the older cohort

	ASS	COO	CRE	CRI*	CUR	EFF	EMO	EMP	ENE	MET	MOT	OPT	PER	RES	SEL	SOC	STR	TOL	TRU
ASS	0.24	-0.07	0.10	0.05	0.04	0.11	-0.06	0.05	0.10	0.07	0.08	0.07	0.02	-0.04	-0.01	0.12	0.06	0.08	0.04
COO	0.07	0.27	0.12	0.12	0.19	0.19	0.18	0.21	0.10	0.21	0.20	0.17	0.18	0.22	0.23	0.20	0.07	0.11	0.16
CRE	0.15	0.08	0.21	0.09	0.19	0.20	0.02	0.12	0.12	0.18	0.16	0.13	0.10	0.06	0.11	0.09	0.05	0.15	0.04
CRI*	0.16	0.00	0.20	0.04	0.16	0.19	-0.04	0.10	0.12	0.17	0.17	0.05	0.11	0.09	0.10	-0.06	0.07	0.10	0.04
CUR	0.15	0.17	0.20	0.13	0.26	0.25	0.09	0.18	0.13	0.25	0.25	0.16	0.19	0.18	0.22	0.13	0.05	0.16	0.06
EFF	0.16	0.12	0.15	0.11	0.20	0.21	0.06	0.15	0.12	0.20	0.20	0.16	0.16	0.13	0.16	0.13	0.08	0.11	0.10
EMO	0.08	0.19	0.10	0.09	0.12	0.16	0.20	0.19	0.06	0.18	0.16	0.15	0.16	0.16	0.19	0.21	0.11	0.07	0.20
EMP	0.15	0.19	0.17	0.09	0.19	0.21	0.13	0.20	0.14	0.20	0.21	0.14	0.16	0.18	0.17	0.16	0.08	0.10	0.14
ENE	0.19	0.07	0.15	0.08	0.16	0.20	0.04	0.16	0.23	0.17	0.17	0.17	0.14	0.10	0.11	0.29	0.10	0.14	0.10
MET	0.12	0.10	0.11	0.08	0.12	0.14	0.05	0.13	0.09	0.19	0.15	0.10	0.11	0.08	0.14	0.07	0.04	0.07	0.08
MOT	0.22	0.19	0.21	0.14	0.25	0.29	0.14	0.20	0.17	0.28	0.30	0.20	0.25	0.20	0.25	0.20	0.10	0.14	0.14
OPT	0.11	0.15	0.12	0.09	0.16	0.17	0.12	0.17	0.15	0.17	0.16	0.19	0.14	0.13	0.17	0.24	0.08	0.11	0.13
PER	0.16	0.22	0.18	0.13	0.25	0.27	0.15	0.21	0.15	0.27	0.28	0.18	0.26	0.25	0.28	0.16	0.08	0.14	0.13
RES	0.12	0.24	0.15	0.13	0.21	0.22	0.16	0.22	0.09	0.26	0.25	0.16	0.24	0.26	0.28	0.13	0.06	0.11	0.18
SEL	0.06	0.21	0.10	0.09	0.17	0.18	0.15	0.18	0.05	0.19	0.17	0.13	0.16	0.17	0.25	0.11	0.06	0.09	0.09
SOC	0.17	0.09	0.12	0.04	0.16	0.16	0.08	0.14	0.23	0.15	0.15	0.16	0.12	0.08	0.08	0.42	0.11	0.13	0.09
STR	0.10	0.05	0.08	0.04	0.06	0.10	0.06	0.09	0.09	0.08	0.07	0.10	0.08	0.07	0.06	0.14	0.13	0.06	0.11
TOL	0.16	0.16	0.17	0.11	0.19	0.22	0.09	0.16	0.12	0.21	0.19	0.14	0.14	0.11	0.17	0.16	0.03	0.14	0.09
TRU	0.05	0.15	0.09	0.06	0.10	0.10	0.12	0.12	0.09	0.13	0.14	0.10	0.10	0.12	0.11	0.20	0.07	0.07	0.18

Note: Correlations are corrected for attenuation.

*Correction for attenuation was not implemented for critical thinking scale due to its very low reliability in teachers' reports.

Table 7.9. Correlations between the teacher (x-axis) and student (y-axis) reports of student skills for the younger cohort

	ASS	COO	CRE	CRI*	CUR	EFF	EMO	EMP	ENE	MET	MOT	OPT	PER	RES	SEL	SOC	STR	TOL	TRU
ASS	0.24	-0.07	0.10	0.05	0.04	0.11	-0.06	0.05	0.10	0.07	0.08	0.07	0.02	-0.04	-0.01	0.12	0.06	0.08	0.04
COO	0.07	0.27	0.12	0.12	0.19	0.19	0.18	0.21	0.10	0.21	0.20	0.17	0.18	0.22	0.23	0.20	0.07	0.11	0.16
CRE	0.15	0.08	0.21	0.09	0.19	0.20	0.02	0.12	0.12	0.18	0.16	0.13	0.10	0.06	0.11	0.09	0.05	0.15	0.04
CRI	0.16	0.00	0.20	0.04	0.16	0.19	-0.04	0.10	0.12	0.17	0.17	0.05	0.11	0.09	0.10	-0.06	0.07	0.10	0.04
CUR	0.15	0.17	0.20	0.13	0.26	0.25	0.09	0.18	0.13	0.25	0.25	0.16	0.19	0.18	0.22	0.13	0.05	0.16	0.06
EFF	0.16	0.12	0.15	0.11	0.20	0.21	0.06	0.15	0.12	0.20	0.20	0.16	0.16	0.13	0.16	0.13	0.08	0.11	0.10
EMO	0.08	0.19	0.10	0.09	0.12	0.16	0.20	0.19	0.06	0.18	0.16	0.15	0.16	0.16	0.19	0.21	0.11	0.07	0.20
EMP	0.15	0.19	0.17	0.09	0.19	0.21	0.13	0.20	0.14	0.20	0.21	0.14	0.16	0.18	0.17	0.16	0.08	0.10	0.14
ENE	0.19	0.07	0.15	0.08	0.16	0.20	0.04	0.16	0.23	0.17	0.17	0.17	0.14	0.10	0.11	0.29	0.10	0.14	0.10
MET	0.12	0.10	0.11	0.08	0.12	0.14	0.05	0.13	0.09	0.19	0.15	0.10	0.11	0.08	0.14	0.07	0.04	0.07	0.08
MOT	0.22	0.19	0.21	0.14	0.25	0.29	0.14	0.20	0.17	0.28	0.30	0.20	0.25	0.20	0.25	0.20	0.10	0.14	0.14
OPT	0.11	0.15	0.12	0.09	0.16	0.17	0.12	0.17	0.15	0.17	0.16	0.19	0.14	0.13	0.17	0.24	0.08	0.11	0.13
PER	0.16	0.22	0.18	0.13	0.25	0.27	0.15	0.21	0.15	0.27	0.28	0.18	0.26	0.25	0.28	0.16	0.08	0.14	0.13
RES	0.12	0.24	0.15	0.13	0.21	0.22	0.16	0.22	0.09	0.26	0.25	0.16	0.24	0.26	0.28	0.13	0.06	0.11	0.18
SEL	0.06	0.21	0.10	0.09	0.17	0.18	0.15	0.18	0.05	0.19	0.17	0.13	0.16	0.17	0.25	0.11	0.06	0.09	0.09
SOC	0.17	0.09	0.12	0.04	0.16	0.16	0.08	0.14	0.23	0.15	0.15	0.16	0.12	0.08	0.08	0.42	0.11	0.13	0.09
STR	0.10	0.05	0.08	0.04	0.06	0.10	0.06	0.09	0.09	0.08	0.07	0.10	0.08	0.07	0.06	0.14	0.13	0.06	0.11
TOL	0.16	0.16	0.17	0.11	0.19	0.22	0.09	0.16	0.12	0.21	0.19	0.14	0.14	0.11	0.17	0.16	0.03	0.14	0.09
TRU	0.05	0.15	0.09	0.06	0.10	0.10	0.12	0.12	0.09	0.13	0.14	0.10	0.10	0.12	0.11	0.20	0.07	0.07	0.18

Note: Correlations are corrected for attenuation.

*Correction for attenuation was not implemented for critical thinking scale due to its very low reliability in teachers' reports.

Table 7.10. Correlations between parent (x-axis) and teacher (y-axis) reports of student skills for the older cohort

	ASS	COO	CRE	CRI*	CUR	EFF	EMO	EMP	ENE	MET	MOT	OPT	PER	RES	SEL	SOC	STR	TOL	TRU
ASS	0.31	-0.01	0.18	0.09	0.11	0.18	-0.10	0.14	0.18	0.09	0.10	0.15	0.08	0.03	0.01	0.27	0.00	0.11	0.04
COO	0.10	0.23	0.07	0.05	0.16	0.13	0.12	0.16	0.11	0.11	0.13	0.16	0.14	0.13	0.15	0.17	0.02	0.05	0.06
CRE	0.12	0.07	0.16	0.07	0.16	0.15	-0.02	0.06	0.07	0.12	0.11	0.11	0.06	0.06	0.09	0.08	-0.06	0.09	-0.02
CRI*	0.14	0.05	0.18	0.06	0.26	0.22	-0.03	0.15	0.06	0.15	0.11	0.12	0.10	0.12	0.15	0.04	0.02	0.15	0.00
CUR	0.16	0.23	0.24	0.13	0.36	0.33	0.09	0.15	0.14	0.28	0.27	0.19	0.20	0.22	0.27	0.14	-0.02	0.21	0.01
EFF	0.22	0.19	0.23	0.12	0.31	0.34	0.08	0.21	0.19	0.26	0.23	0.24	0.23	0.22	0.26	0.24	0.06	0.16	0.06
EMO	0.06	0.25	0.08	0.05	0.16	0.17	0.31	0.12	0.09	0.14	0.17	0.15	0.17	0.19	0.20	0.12	0.13	0.08	0.13
EMP	0.13	0.21	0.10	0.06	0.19	0.17	0.06	0.23	0.14	0.14	0.15	0.16	0.16	0.18	0.17	0.20	-0.06	0.10	0.03
ENE	0.18	0.11	0.14	0.07	0.20	0.22	0.04	0.11	0.25	0.14	0.18	0.20	0.13	0.13	0.10	0.26	0.01	0.10	0.05
MET	0.16	0.22	0.16	0.11	0.30	0.30	0.07	0.20	0.10	0.27	0.24	0.18	0.23	0.27	0.29	0.13	0.01	0.15	0.10
MOT	0.23	0.26	0.23	0.17	0.40	0.38	0.10	0.24	0.18	0.38	0.37	0.22	0.33	0.29	0.38	0.19	-0.01	0.21	0.11
OPT	0.13	0.16	0.11	0.07	0.15	0.18	0.08	0.11	0.14	0.11	0.10	0.21	0.10	0.08	0.11	0.21	0.09	0.04	0.02
PER	0.15	0.23	0.16	0.14	0.32	0.33	0.14	0.21	0.14	0.32	0.32	0.17	0.31	0.30	0.36	0.11	0.02	0.17	0.13
RES	0.11	0.32	0.13	0.13	0.35	0.31	0.25	0.26	0.14	0.33	0.35	0.17	0.36	0.40	0.40	0.13	0.04	0.19	0.16
SEL	0.12	0.26	0.11	0.12	0.27	0.25	0.14	0.19	0.10	0.27	0.25	0.16	0.24	0.28	0.33	0.09	0.02	0.13	0.09
SOC	0.24	0.00	0.10	0.01	0.04	0.07	-0.05	0.09	0.26	-0.02	0.01	0.17	-0.02	-0.08	-0.10	0.42	0.04	0.00	0.04
STR	0.09	0.06	0.07	0.02	0.07	0.12	0.12	0.06	0.09	0.02	0.07	0.13	0.05	0.06	0.06	0.16	0.22	0.03	0.03
TOL	0.09	0.16	0.14	0.07	0.20	0.18	0.05	0.07	0.08	0.15	0.14	0.12	0.10	0.10	0.12	0.07	-0.05	0.15	0.00
TRU	0.06	0.10	0.02	0.01	0.09	0.08	0.12	0.12	0.09	0.06	0.07	0.11	0.07	0.08	0.06	0.19	0.10	0.01	0.19

Note: Correlations are corrected for attenuation.

*Correction for attenuation was not implemented for critical thinking scale due to its very low reliability in teachers' reports.

Table 7.11. Correlations between parent (x-axis) and teacher (y-axis) reports of student skills for the younger cohort

	ASS	COO	CRE	CRI*	CUR	EFF	EMO	EMP	ENE	MET	MOT	OPT	PER	RES	SEL	SOC	STR	TOL	TRU
ASS	0.36	-0.02	0.21	0.10	0.13	0.22	-0.07	0.11	0.21	0.15	0.15	0.17	0.06	0.00	0.02	0.30	0.10	0.12	0.04
COO	0.04	0.28	0.09	0.09	0.19	0.19	0.23	0.15	0.11	0.18	0.18	0.19	0.17	0.17	0.19	0.24	0.11	0.04	0.07
CRE	0.15	0.09	0.23	0.08	0.19	0.21	0.04	0.11	0.14	0.15	0.17	0.12	0.09	0.06	0.14	0.11	0.07	0.13	-0.01
CRI	0.17	0.02	0.16	0.05	0.16	0.19	0.04	0.09	0.11	0.13	0.19	0.09	0.13	0.11	0.14	0.04	0.13	0.05	-0.04
CUR	0.13	0.14	0.23	0.09	0.31	0.28	0.06	0.12	0.21	0.20	0.25	0.16	0.20	0.20	0.21	0.17	0.10	0.17	-0.04
EFF	0.22	0.21	0.24	0.14	0.29	0.37	0.19	0.23	0.23	0.30	0.30	0.24	0.27	0.22	0.27	0.29	0.19	0.13	0.02
EMO	0.06	0.33	0.12	0.10	0.21	0.24	0.35	0.21	0.10	0.23	0.24	0.21	0.24	0.23	0.27	0.28	0.21	0.06	0.16
EMP	0.10	0.21	0.14	0.09	0.22	0.21	0.21	0.18	0.16	0.15	0.20	0.20	0.16	0.19	0.17	0.30	0.10	0.11	0.06
ENE	0.25	0.14	0.17	0.09	0.25	0.28	0.11	0.17	0.34	0.20	0.21	0.25	0.21	0.15	0.13	0.46	0.12	0.14	0.07
MET	0.12	0.15	0.13	0.10	0.17	0.21	0.11	0.15	0.09	0.23	0.21	0.15	0.17	0.14	0.21	0.13	0.14	0.08	-0.04
MOT	0.20	0.29	0.24	0.16	0.33	0.36	0.20	0.24	0.21	0.34	0.38	0.23	0.32	0.27	0.32	0.25	0.14	0.16	0.06
OPT	0.14	0.15	0.13	0.09	0.20	0.20	0.16	0.16	0.15	0.16	0.16	0.24	0.14	0.09	0.15	0.30	0.15	0.10	0.05
PER	0.15	0.27	0.19	0.16	0.31	0.32	0.24	0.23	0.17	0.34	0.35	0.20	0.37	0.33	0.36	0.21	0.16	0.12	0.10
RES	0.09	0.35	0.16	0.13	0.33	0.34	0.29	0.23	0.16	0.30	0.36	0.21	0.39	0.43	0.39	0.20	0.16	0.10	0.13
SEL	0.06	0.27	0.12	0.15	0.21	0.22	0.24	0.19	0.03	0.26	0.24	0.15	0.25	0.24	0.34	0.10	0.11	0.07	0.02
SOC	0.21	0.03	0.09	0.01	0.13	0.14	0.03	0.11	0.29	0.06	0.06	0.22	0.04	0.03	-0.06	0.55	0.14	0.06	0.07
STR	0.14	0.16	0.12	0.05	0.18	0.21	0.22	0.16	0.16	0.16	0.19	0.18	0.19	0.17	0.14	0.26	0.25	0.05	0.16
TOL	0.14	0.09	0.18	0.07	0.19	0.21	0.06	0.08	0.14	0.12	0.16	0.12	0.10	0.09	0.10	0.19	0.04	0.16	0.01
TRU	0.02	0.10	0.04	-0.01	0.09	0.08	0.08	0.04	0.09	0.04	0.09	0.07	0.09	0.12	0.02	0.18	0.09	0.02	0.15

Note: Correlations are corrected for attenuation.

*Correction for attenuation was not implemented for critical thinking scale due to its very low reliability in teachers' reports.

Table 7.12. Correlations between skill factors in correlated MTMM models, by cohort

MTMM skill factors			Correlations	
			Older cohort	Younger cohort
Assertiveness	<->	Energy	0.55	0.52
Assertiveness	<->	Sociability	0.50	0.49
Sociability	<->	Energy	0.65	0.79
		Average	0.57	0.60
Emotional control	<->	Optimism	0.66	0.69
Emotional control	<->	Stress resistance	0.71	0.70
Stress resistance	<->	Optimism	0.64	0.67
		Average	0.67	0.69
Co-operation	<->	Empathy	0.82	0.80
Co-operation	<->	Trust	0.59	0.62
Trust	<->	Empathy	0.46	0.48
		Average	0.62	0.63
Achievement motivation	<->	Persistence	0.90	0.88
Achievement motivation	<->	Responsibility	0.85	0.78
Achievement motivation	<->	Self-control	0.79	0.71
Achievement motivation	<->	Self-efficacy	0.82	0.99
Persistence	<->	Responsibility	0.98	0.99
Persistence	<->	Self-control	0.83	0.87
Persistence	<->	Self-efficacy	0.77	0.85
Responsibility	<->	Self-control	0.87	0.90
Responsibility	<->	Self-efficacy	0.72	0.75
Self-control	<->	Self-efficacy	0.69	0.68
		Average	0.82	0.84
Creativity	<->	Critical thinking	0.72	0.74
Creativity	<->	Curiosity	0.69	0.75
Creativity	<->	Tolerance	0.56	0.65
Creativity	<->	Meta cognition	0.55	0.61
Critical thinking	<->	Curiosity	0.73	0.66
Critical thinking	<->	Tolerance	0.61	0.56
Critical thinking	<->	Meta cognition	0.94	0.97
Curiosity	<->	Tolerance	0.76	0.84
Curiosity	<->	Meta cognition	0.78	0.76
Tolerance	<->	Meta cognition	0.54	0.61
		Average	0.69	0.72
		Overall average	0.71	0.74

Figure 7.1. Relationship between persistence and school grades, by type of skill estimate and cohort

Figure 7.2. Relationship between curiosity and school grades, by type of skill estimate and cohort