Social and emotional development and school learning:

A measurement proposal in support of public policy

OECD
BETTER POLICIES FOR BETTER LIVES

INSTITUTO AYRTON SENNA

GOVERNO DO Rio de Janeiro
SOMANDO FORÇAS

SECRETARIA DE EDUCAÇÃO

Social and emotional development and school learning:

A measurement proposal in support of public policy

Preliminary results of the social and emotional skills measurement Project in Rio de Janeiro

Authors
Daniel Santos
Ricardo Primi

São Paulo, 2014

OECD BETTER POLICIES FOR BETTER LIVES
INSTITUTO AYRTON SENNA PARCERIA SOCIAL
GOVERNO DO Rio de Janeiro SOMANDO FORÇAS
SECRETARIA DE EDUCAÇÃO

Planning:
Instituto Ayrton Senna

Authors:
Daniel Santos
Ricardo Primi

Partners:
Rio de Janeiro State Education Department (SEEDUC)
Organization for Economic Cooperation and Development (OCDE)

Project Staff:
General Coordination: Tatiana Filgueiras
Technical Coordination: Laura di Pizzo
Communication Coordination: Daniela Arai
Contribution: Eduardo Silva

Authors

Daniel Santos
Having graduated with a Doctorate in Economics from the University of Chicago, Daniel Santos is currently a Professor of Economics at the University of São Paulo's Ribeirão Preto campus, where he coordinates the Laboratory for Study and Research into Social Economics (LEPES) and participates in the coordination of the Nucleus for Research Support in Neurodevelopment and Mental Health (NAP-N&SM) and the Nucleus of Excellence for Early Childhood (NEPI). His investigative interests are focused on the projection of demand and evaluation of the impact of policies concerning early childhood and the measurement of the impact of social and emotional characteristics that are influenced by the school environment.

Ricardo Primi
Ricardo Primi graduated in psychology from PUC University in Campinas, and received his Educational Psychology and Human Development Doctorate from the University of São Paulo. This was partly performed at Yale University (United States) under the guidance of Robert J. Sternberg. He is Coordinator of Psychological and Educational Evaluation Laboratory (LabAPE), receiving financing from the CNPq (research productivity), FAPESP AND CAPES. He is associate professor on the Psychology post-graduate program at the São Francisco University (Master's and Doctor's in Psychological Evaluation). He is past President of the Brazilian Institute of Psychological Evaluation (IBAP) and a member of the Consultative Commission in Psychological Evaluation of the Federal Board of Psychology and Commission in the Area of Psychology for the MEC/INEP. In 2009, he was visiting Professor at the University of Toledo (Ohio, United States) on a Fulbright/CAPES scholarship. He teaches courses in psychological, statistical and psychometric evaluation and develops research works on the evaluation of intelligence, personality and career development, and Item Response Theory.

Instituto Ayrton Senna

The **Ayrton Senna Institute** is a non-profit organization that researches, produces and applies a wide variety of knowledge for an improvement in the quality of education. Inspired by the ideas of the three-time Formula 1 World Champion Ayrton Senna, the Institute has been operating for 20 years, promoting the full development of the different skills of children and young people by designing and implementing innovative educational solutions in the areas of administration, evaluation and public policy. Financed by its own resources, donations and partnerships with the private sector, the Institute makes diagnostic and planning services freely available to public bodies, provides training for administrators and educators, and performs monitoring and analysis of indicators, whilst also making recommendations for public policies based upon successful studies and practices. Due to its widely recognized production and publication of information and solutions relating to human development, the Ayrton Senna Institute has formed part of the UNESCO 'Chairs' program since 2004. The Ayrton Senna Institute trains 75,000 educators every year, whilst its programs directly benefit around two million students in more than 1,200 municipalities in various regions throughout Brazil.

Presentation

Education is a basic human right and a condition for the exercising of other basic rights such as health, work, security and citizenship. However, in order to effectively fulfill the right to education, it is not enough to simply guarantee future generations' access to school and the ability to pursue their studies, it is also necessary to guarantee a meaningful education to confront the academic, professional and political challenges of the 21st century.

Faced with this important and immediate task, a number of questions arise: what skills are necessary to promote the full development of humankind along with the socioeconomic progress of nations? What sort of mechanisms should be employed to develop these skills and accelerate their progress? How can the effectiveness of schools be measured to promote the skills in order to offer support in the design of fairer and more effective pedagogic practices?

Even though a great deal of research and debate should be undertaken before a consensus can be reached on these issues, there already exists enough evidence to confirm that our education systems, in general, fail to largely take into account one specific set of skills that are essential for the success of children and young people both in and out of school: the so-called 'social and emotional skills'.

Well-founded studies performed by economists, psychologists and educators over recent decades have revealed that skills such as perseverance, autonomy and curiosity are just as important as the cognitive abilities(measured by means of achievement tests and IQ) in obtaining good results in different areas of individual and collective well-being, such as ongoing studies, income and health. Furthermore, the evidence suggests that these skills benefit results in adult life *via schooling*, or in other words, by means of their contribution to academic success.

Given the potential that the social and emotional approach has to improve the indexes of education in general, reduce inequality in education systems and promote social and economic prosperity, the Ayrton Senna Institute and its collaborators have joined together with the Centre for Educational Research and Innovation (CERI), part of the Organization for Economic Cooperation and Development (OECD) to develop a pioneering project in the measurement of social and emotional skills in the school context.

With the objective of supporting administrators and educators in the design and monitoring of public policies in this area, the Project has aimed to develop a reliable instrument for the gauging of socio-economic skills on a large-scale and validate it with concrete evidence through a pilot application implemented with a representative sample of students from the Rio de Janeiro state school system. The information collected from the implementation of the pilot provided material for a database that has been analyzed by researchers on the Project aiming to extend investigations into the relationship between the skills evaluated and educational performance, as well as research the channels along which specific aspects of school, the student and the family influence social and emotional development.

This Report describes the planning, execution and analysis phases of the Project, as well as the results of the study of the technical properties of the instrument developed, that attest to its quality for use as a diagnostic and monitoring instrument in public policy. The preliminary results of the analysis of the students' social and emotional profile are also presented, as is the relationship of this profile with academic performance and the social and economic characteristics.

Acknowledgements

This publication would not have been possible without the contribution of a number of different people. The authors and the staff of the Ayrton Senna Institute would like to thank the researchers Koji Yamamoto, Katarzyna Kubacka, Mirela de Carvalho, Izabela Moi, Rogério Lerner, Elaine Pazello, Luiz Guilherme Scorzafave, Alexandre Luiz de Oliveira Serpa, Everson Cristiano de Abreu Meireles and the entire staff at the Psychological and Educational Evaluation Laboratory (LabAPE-USF), and the Laboratory for Study and Research into Social Economics (LEPES-USP); the Rio de Janeiro State Secretary for Education, Wilson Risolia and his staff, especially Antônio Vieira Neto, Patrícia Tinoco, Daniela Ribeiro, Carina Moura, Éder Terrin, Vânia Machado and Eliane Dantas; the Rio de Janeiro Municipal Secretary for Education, Claudia Costin and her staff, especially Helena Bomeny, Ana Lucia Barros, Antonio Augusto Alves Mateus Filho and Jurema Holperin; the CAEd Staff, especially Waldirene Barbosa.

Contents

CONTENTS

List of Graphs, Tables and Annexes

Tables

Graphs

Annexes

Introduction

Preparing children and young people for the challenges of the 21st century means creating conditions for the development of all the necessary skills for academic, professional and personal success in an ever more demanding world. Amongst these skills are those that were recognized and measured by educational systems, such as those related to literacy and numeracy, but these are also factors that are not sufficiently captured by performance tests and which do not generally form part of schools' focused curriculums, even though they are just as important for the full development of a human being and for the social and economic progress of nations.

Studies performed by economists, psychologists and educators over recent decades have revealed that skills such as perseverance, responsibility and cooperation have a significant impact on individuals' performance both inside and outside of school, being just as important as the cognitive abilities in obtaining good results in different areas of individual and collective well-being, such as the final level of education achieved, employment and health. These studies also highlight that individuals with more highly developed social and emotional skills show a greater facility for absorbing academic content.

In fact, the idea that students who are more organized, focused and confident learn more is nothing new to educators, likewise that students who are more persistent and resilient tend to be able to commit themselves to long-term goals and dealing with frustrations and conflicts in a better way. It is well-known, for example, that the act of learning curricular content not only involves skills connected to the speed of reasoning and to memory, but also requires motivation and the ability to control anxiety and other emotions. Creativity, meanwhile, involves the ability to put traditional ways of thinking on hold and requires a certain amount of self-esteem and confidence.

Schools recognize the importance of this dimension, understanding that the development of students is multidimensional and that learning involves the mastering of "non-cognitive" skills of an emotional and behavioral nature. However, despite a dedicated understanding of the social and emotional area by parents and teachers, very little effort is made for its planned development and for an evaluation of the effectiveness of interventions designed to promote it.

Amongst the main causes of this limitation is a lack of understanding concerning the mechanisms by which these skills can be developed and measured in various contexts of learning. Part of this limitation is derived from the relative scarcity of studies that investigate the relationship between social and emotional characteristics and variables related to learning. This relative scarcity of studies is, in turn, largely the result of a lack of databases that contain both information of this nature and the variables of learning results.

Committed to the challenge of expanding the base of knowledge on the subject, the main objective of this Project on measuring social and emotional skills was the construction of an instrument to measure social and emotional attributes that was both economically viable for application on a large-scale (in such a way as to serve as a tool for the formation of public policies) and also showed itself to be scientifically robust enough to support academic research

in the area. The construction process behind the instrument is described in this report, as is the evaluation phase which was conducted by means of a pilot project applied to a sample of around 25,000 students from Rio de Janeiro state education network.

The starting point for this journey was a revision of the existing literature concerning those social and emotional skills that are important for the improvement of students' educational performance and the promotion of well-being in adult life. The first chapter of this report is, therefore, dedicated to a description of the main conclusions drawn from the scientific work performed on the ways in which social and emotional attributes influence individual and collective results, and how education is capable of affecting their development.

The second chapter describes the first stage in the construction of the instrument used for the measurement of social and emotional skills. The first section covers the first stage, including an analysis of a wide-reaching set of psychometric instruments that have been extensively accepted by international writings and which, in theory, were considered as candidates to contribute to the new measurement instrument. Following this, the stages conducted for the evaluation of these instruments and consultation of specialists and education administrators for analysis of the selection are described. Next, the report describes the stages of translation, adaptation and revision of the selected instruments, as well as the studies that formed a base for the selection and composition of items for the construction of a unique instrument for gauging ocial and emotional skills for large-scale application. Lastly, the report looks at the designing of the pilot scheme, the data collection process and analysis of technical properties of the instrument.

The third chapter presents preliminary results of the analysis of the data collected in the pilot application, seeking to better investigate the relationship between avaluated skills and academic performance, as well as research manners in which aspects of the school, the student and the family influence socio-emotional development.

Finally, the conclusion brings together the main findings of the preliminary analysis of the data drawn from the pilot application and presents the results of this stage of the Project: the social and emotional measurement program named *SENNA (Social and Emotional or Non-cognitive Nationwide Assessment)*, the first version of which may be found as an Attachment to this Report.

Chapter 1:

Social and emotional development and learning

1.1 The importance of social and emotional characteristics for learning

In this section, we will be discussing the two most important lines of academic investigation concerning the importance of social and emotional development for learning. The first approach is based upon the policy debate pursued between the various interested players in the area of education relating to which skills are important to individual and collective success in the 21st century. The second line is based upon the collection and summary of empirical evidence relating to which abilities most influence learning. As will become evident, each approach has both advantages and limitations, showing themselves to be complementary

In relation to the first approach, some of the most important theoretical frameworks oriented towards the transformation of the process of formal teaching in line with the demands of the contemporary world will be described. To a large extent, these frameworks are the results of extensive theoretical discussions and consultations with leading players directly involved in the organization of society to develop both public property and better institutions in terms of both goods and services. The benefit of this type of approach is typically **the extent of its reach**, since it arises from individual or collective analysis which does not demand empirical proof, or in other words, it is not necessarily restricted by need for or the feasibility of objective validation. Its weakness, on the other hand, is its greater susceptibility to misunderstanding or subjectivity, given the lack of empirical evidence to support it.

The second approach is based upon the collection and synthesis of empirical evidence concerning the relationship between social and emotional development and learning. In this case, the benefit lies in its **accuracy**. The weakness is obviously the limitation imposed by the availability of studies that have investigated the relationship between social and emotional characteristics and variables related to learning. The reason for this scarcity is that the set of social and emotional attributes is extensive and there is little consensus on how the attributes should be measured, whilst they only started to be measured at a relatively late date when compared to the other individual characteristics. The data bases that simultaneously contain information of this nature and the results of learning variables

are relatively scarce, and the majority of these databases are not in the public domain. This second approach, therefore, is restricted to those characteristics that have been investigated (and which have, therefore, a priori been considered of interest and are possibly able to be measured by a researcher), and possibly do not exhaust the set of potential social and emotional characteristics of interest.

The approaches are complementary in at least two areas. On the one hand, the empirical evidence, even though it is incomplete, at least partially confirms the opinion of the players worldwide concerning the importance of the new skills proposed in determining individual and collective success. On the other hand, the empirical evidence shows that the school, as it stands today, whilst it may be out of step in relation to the needs of these new times, is, even so, capable of contributing to the development of non-cognitive attributes associated with success. This evidence is, in some way, a portrait of both the past and the present, and tells a story of accidental success. In contrast, the thoughts of the global players point to the future, running the natural risks of those who try to anticipate the needs of a society undergoing change.

In this way, the two lines of investigation show themselves to be complementary, in that, together, they provide an accurate, but not limited, view of the skills that should be forming part of a quality education in the 21st century. On the one hand, this view encompasses those skills that have had their positive impact proven by scientific evidence and which, therefore, may not be neglected by education systems and those responsible for establishing public policies. On the other hand however, we cannot ignore those skills that have still not been studied and which could prove to be important.

1.1.1. Global initiatives: theoretical frameworks of skills for the 21st century

The current moment in time, replete as it is with far-reaching technological and institutional transformations, points towards the need for equally far-reaching transformations in learning systems. The Delor Report (UNESCO, 1996) is a watershed document in the shift in educational discourse in response to the new challenges, suggesting a learning system based upon four pillars: (i) Learning to Know; (ii) Learning to Do; (iii) Learning to be; and (iv) Learning to Live Together.

Since its publication, initiatives have sprung up around the world that have tried to define more strictly and more thoroughly exactly what those skills are that are necessary for the achievement of the four pillars mentioned above, as well as investigate whether there are any other important learning objectives in addition to these four. The conclusions of these initiatives are not identical, but they do suggest close points of contact. The intense flow of information in the contemporary world, allied with the advent of new technologies and institutions, allows the appearance of new forms of production and much more flexible interaction. Given this situation, students and teachers are no longer mere receivers and transmitters of existing information, but rather they need to develop the ability to construct part of the knowledge and adapt the existing knowledge to their needs. They should also be adaptable to deal with different situations and cultures. **Table 1**, taken from Lee (2013), summarizes some of the most important international frameworks and highlights their common directives[1]:

1. In **Table 1**, P21 is the *Partnership for the 21st Century Skills:* ATC21s refers to *Assessment & Teaching of 21st Century Skills*; OECD is the *Organization for Economic Cooperation and Development*; European Reference Framework forms part of the recommendation made by the European Parliament and Council to the European Commission and Member States concerning the *Education and Training 2010 Work Program*.
Source: Binkley et al. (2010); Gordon et al. (2009); OECD (2005); http://www.p21.org/overview.

TABLE 1: OVERVIEW OF KEY COMPETENCIES IDENTIFIED BY VARIOUS 21CC IMPLEMENTATION FRAMEWORKS

P21	ATC 21S	OECD	European Reference Framework
Learning and innovation skills 1) Creativity & Innovation 2) Critical Thinking & Problem Solving	*Ways of thinking* 1) Creattivity and innovation 2) Critical thinkiing, problem solving, decision making 3) Learning to learn, metacognition		1) Learning to learn
3) Communication & Collaboration	*Ways of working* 4) Communication 5) Collaboration (teamwork)	*Interacting in heteregoneus groups* 1) Relate well to others 2) Co-operate, work in teams 3) Manage and resolve conflicts	2) Communication in the mother language 3) Communication in foreign language
Information, media and technology skills 4) Information literacy 5) Media literacy 6) ICT literacy	*Tools for working* 6) Information literacy 7) Technological literacy (ICT)	*Using tools interactively* 4) Use language, symbols and texts interactively 5) Use knowledge and information interactively 6) Use technology interactively	4) Mathematical competence and basic competences in science and technology 5) Digital competence
Life and career skills 7) Flexibility & Adaptability 8) Initiative and self-direction 9) Social and cross cultural skills 10) Productivity & Accountability 11) Leadership & Responsability	*Living in the world* 8) Citizenship (local and global) 9) Life and career 10) Personal and social responsability (including cultural awareness and competence)	*Acting autonomously* 7) Act within the big picture 8) Form and conduct life plan and personal projects 9) Defend and assert rights, interests, limits and needs	6) Social and civic competences 7) Sense of initiative and enterpreneurship 8) Cultural awareness and expression

As Lee (2013) rightly notes, the skills and abilities listed in the initiatives above are closely connected with the so-called '*soft skills*' (Heffron, 1997; Heckman e Kaultz, 2012), as well as with the concept of 'Social Capital' (Putnam, 1995). These concepts cover talents that are flexible and sensitive to experiences and interaction with other individuals (thus the use of the term '*soft*', in contrast to the less flexible intelligence and knowledge such as gauges by achievement tests and IQ). A number of these '*soft skills*' are related to the learning that is empirically investigated by psychologists, economists and social scientists. The next sub-section summarizes these writings.

1.1.2 Scientific evidence: skills that promote learning

In this section, we will be separately analyzed the roles of different social and emotional characteristics on the lives of people, grouping the available evidence within the five areas defined by the so-called 'Big Five' personality traits ('**Big Five**')[2].

Over the last few decades, a consensus has appeared amongst psychologists that the most effective way of analyzing the human personality consists of observing it in five specific dimensions, known as the Big Five Traits: **Openness to New Experiences, Extraversion, Agreeableness, Conscientiousness and Emotional Stability**. The "Big Five" are latent constructs obtained by factor analysis performed on the answers to extensive questionnaires with varied questions about behaviors that are representative of all the personality characteristics that an individual could have. When applied to people from different cultures and at different points in time, these questionnaires show them to have the same latent factor structure[3], giving rise to the hypothesis that the personality traits of human beings would effectively group around five big areas.

A great many of the empirical works performed have used scales and tests that measure particular personality traits and fit them into at least one of the domains of the Big Five. We have adopted the method proposed by John and Srivastava (1999) and cited by Almlund et al (2011) to frame the traits collected by scales and tests within the 'Big Five' groups, as can be seen in **Table 2** below.

Openness to Experiences

2. The pioneering nature of the Big Five theory is attributed to Gordon Allport and his colleagues who developed it in the mid-1930s. Influenced by Francis Galton's lexical thesis, according to which the most important individual differences ought to be present in day-to-day language, Allport and his colleagues searched the dictionaries for all the adjectives that could describe personality attributes (such as, for example, "kind", aggressive", etc.). Greater refinement was proposed in the 1940s by Raymond Catell, who reduced the list of adjectives down to 171, and then grouped them together by likeness, via factorial analysis, into 35 clusters. The next step was to construct personality tests that allowed for the establishment of multiple dimensions of personality. From the 1960s on, with large samples coming from the application of various different personality tests and re-analyses of Cattell's studies, various authors found that five main factors explained the greater part of the variation that existed in the tests. Starting in the 1960s, the authors who most contributed to the model and who were, as such, considered to be its "fathers", were: Lewis Goldberg, Robert R. McCrae e Paul T. Costa, Jerry Wiggins and Oliver John.

3. A statistical method that allows for the identification (in a set of multiple measurements of individual results) of a smaller vector of factors that satisfactorily explains the variability existing in such results. The technique is fundamental for a relatively large set of demonstrations of behavior to be able to be explained by a smaller number of individual characteristics, even though these characteristics cannot be directly measured.

TABLE 2. PERSONALITY TRAITS AND THEIR DIFFERENT FACETS

Trait (Big Five)	Description in the AAP Dictionary*	Facets	Related Attributes	Character attributes (childhood)
Openness to Experience (including intellect)	The tendency to be open to new aesthetic, cultural, or intellectual experiences	Fantasy (imaginative) Aesthetic (artistic) Emotional (excitable) Actions (wide range of interests) Ideas (curious) Values (unconventional)	-	Pleasure taken in low-intensity activities Curious Emotional Sensitive
Conscientiousness	The tendency to be organized, responsible, and hardworking	Competence (efficient) Order (organized) Autonomy (doesn't expect help) Fights for objectives Discipline (not lazy) Deliberation (not impulsive)	Strength of character Perseverance Postpones rewards Controls impulses Plans and fights for objectives Ambition Ethics at work	Attention Concentration Commitment to controlling actions Controls impulses/ delays rewards Persistence Activity**
Extraversion	An orientation of one's interests and energies toward the outer world of people and things rather than the inner world of subjective experience; characterized by positive affect and sociability	Welcoming (friendly) Aggregator (sociable) Affirmation (self-confident) Activity (energetic) Seeks excitement (adventurer) Positive emotions (enthusiastic)	-	Social dominance Social vitality Shyness** Activity** Positive emotions Sociable/ affiliation Seeks sensations
Agreeableness	The tendency to act in a cooperative, unselfish manner	Confidence in others (tolerant) Objectivity (direct when talking to someone) Unselfish Obedient (not stubborn) Modest Docile (likeable)	Empathy Looks at problems from different angles Cooperation Competitive	Irritable** Aggressive Willing Available
Emotional Stability (Neuroticism)	Predictability and consistency in emotional reactions, with absence of rapid mood changes; a chronic level of emotional instability and proneness to psychological distress	Anxious (worried) Hostile (irritable) Depressed Introspective (shy) Impulsive Vulnerable and stressed (not self confident) Optimistic Psychopathology (mental disorders) including depression and anxiety	Locus of Control Self-esteem Self-efficacy	Fear (intimidation) / behavior inhibition Shyness** Irritability** Frustration Sadness Difficulty in calming down

*American Psychology Association. ** May be related to more than one of the 'Big Five'.

Openness to new experiences is defined as the tendency to be open to new aesthetic, cultural and intellectual experiences. The individual who is open to new experiences is characterized as being imaginative, artistic, excitable, curious and unconventional, whilst having a wide range of interests.

Since this trait is intrinsically associated with curiosity, imagination and questioning, it is no surprise that it appears in various studies as being strongly related to certain educational indicators such as the final level of education reached, school performance and the opting for more difficult courses. There is, however, great difference with respect to whether this evidence can be interpreted as causal. This is because *Openness* is a characteristic highly correlated to *Intelligence*. It is unclear, therefore, whether the best educational results obtained by young people who are more open is due to the fact that they have higher levels of this attribute, or if in fact it is a simple consequence of having better cognitive indicators. Various different studies document the fact that people who are more open to new experiences generally obtain a higher level of schooling, but this relationship lessens when people with the same level of intelligence are compared.

Almlund et al (2011) conclude that the increase of a standard deviation in the Openness to New Experiences construct is associated with a rise of up to 0.2 years of study. In a very methodologically rigorous study, Lounsbury et al (2004) established that those high school students who are more open to new experiences missed fewer classes and opted for more difficult mathematics courses when asked to choose, even though in the end they did not receive higher grades than the others.

Even though amongst the Big Five, Openness is the second most common trait to correlate with the final average of grades in school (after Conscientiousness), it can be said that this correlation is relatively modest, being around one third of that estimated between intelligence and grades[4].

Amongst the facets of Openness particular attention has been paid to Creativity[5], this having been listed in a number of the initiatives mentioned in the previous section as one of the most important attributes for the 21st century. In reviewing the writings on the matter, Gutman and Schoon (2013) find a moderate correlation between Creativity and school grades, but a higher correlation of this construct with the quality of work done for the conclusion of courses in university (r = 0.46).

4. According to Poropat's meta-analysis (2009).
5. Authors such as Matthews and Deary (1998) classify Creativity as a concept that is extremely close to Openness, but this is far from being a consensus. Sternberg (1999), for example, defines Creativity as part of Intelligence, whilst Kyllonen et al (2011) define it as a quasi-cognitive factor. The decision to present this evidence here was solely through the choice to present the evidence in five large groups, associated to the Big Five.

Conscientiousness

Conscientiousness is defined as the tendency to be organized, hard working and responsible. The conscientious individual is characterized as being efficient, organized, autonomous, disciplined, lacking impulse and guided towards his objectives (a fighter).

Of all the personality traits, *Conscientiousness* is, without doubt, that most associated with the different measures of success in learning. In fact, within this group are included characteristics such as perseverance, discipline, endeavor and responsibility, characteristics that are important in any activity that involves medium and long-term commitment, such as study or work. In what is referred to as staying in school, *Conscientiousness* is, together with *Openness to New Experiences*, the attribute most associated with the level of schooling which an individual ultimately reaches.

The magnitude of the impact of *Conscientiousness* on the years of study is close to that observed for intelligence (0.2 years of study for each increase of one standard deviation, being especially important for men). There is also evidence that behavior related to *Conscientiousness*, such as punctuality for classes and delivery of homework, is capable of predicting the final level of education reached as much as ten years in advance (Lleras, 2008). In terms of school performance, as measured by grades obtained during the educational cycle, *Conscientiousness* is the most important of the personality traits, rivaling measurements such as intelligence in the extent of its impact. However, contrary to *Intelligence* and *Openness*, the association of *Conscientiousness* with grades does not decline during the educational cycle[6].

Martin (1989) shows that measures of persistence and distractibility reported by parents on individuals whilst still in early childhood are already extremely correlated with grades in school and standard tests. In the same way, Mischel et al (1989) show that children with a greater capacity to delay rewards as measured in the Marshmallow Test[7] obtained higher grades in the standard SAT[8] tests used for university entry in the United States. The correlation between these variables is surprisingly high: 0.42 and 0.57 in language and mathematics exams respectively.

In another study, Duckworth and Seligman (2005) show that a proportion of the variance of the grades obtained by a group of students in the eighth grade that is explained by self-discipline measured at the beginning of the school year is more than twice as much as that explained by intelligence. Finally, the studies seem to consistently point towards the fact that *Conscientiousness* is more related to grades obtained in school than to the standard tests, suggesting that there are mechanisms that go beyond learning capacity that relate *Conscientiousness* to success in school[9].

6. As would be expected from a homogenization of the classes (due to the holding back of those students demonstrating the poorest performance). This fact suggests that conscientiousness may indeed be more important than intelligence if this fact were to be appropriately considered.
7. In the Marshmallow Test, the researcher offers a marshmallow to a child and says that if they resist the temptation to eat it while he leaves the room, they will get a second marshmallow. The reactions of the children are filmed and the time the child waits before eating the marshmallow is a measure of their ability to delay rewards.
8. The SAT test (Scholastic Assessment Test) is a standard test given to high school students in the United States, serving as the criteria for entry into North American universities.
9. For example, the self-control required for a test that the students do under pressure, or the discipline and perseverance to learn to deal with problems set by a teacher in a course exam, and which differ from a standard test in which the questions are developed by external examiners and that are possibly entirely new to the students.

From another perspective, Jacob (2002), studying the reasons for the increase of the difference in the percentage of girls and boys that decide to go to university (with more girls going than boys), discovered that differences of self-discipline between the two groups are found to be amongst the main determining factors of the phenomenon, being much more important, for example, than intelligence.

Extraversion

Extraversion is defined as the orientation of interests and energy towards the external world, people and things (instead of the internal world of subjective experience). The extravert individual is characterized as being friendly, sociable, self-confident, energetic, adventurous and enthusiastic.

Of all the Big Five traits, *Extraversion* is that which probably presents effects that are the most difficult to capture in statistical exercises, since the relationship of this trait to educational results does not appear to be monotonic, that is, not always having more (or less) of this trait suggests an advantage in the classroom. The associations of measurements of *Extraversion* with educational results and the work market are, in the majority of available articles, statistically zero, or at least very low. In an example that goes against the rule, Carneiro et al (2007) found that, even though *Extraversion* is not particularly important in determining grades obtained, or results in languages or mathematics measured amongst adults, it can be relevant in young people's decisions to remain in school longer.

Agreeableness

Agreeableness is defined as the tendency to act in a cooperative and unselfish manner. The agreeable or cooperative individual is characterized as being tolerant, altruistic, modest, likeable, flexible and objective (direct when dealing with people).

Characteristics such as aggressiveness, irritability[10], availability and kindness are associated with *Agreeableness*, which is why it is supposed to have a special impact on activities performed in a group. Duncan and Magnusson (2010) found that aggressiveness in childhood is an important predictor (negative) for the conclusion of high school, suggesting that this facet of *Agreeableness* may perform an important role in determining educational results.

In terms of school grades, a correlation can be seen between *Agreeableness* and grades similar to those obtained for *Conscientiousness* during elementary school, but contrary to the latter, this correlation disappears with the advancement of the educational cycle[11].

10. Also associated with *Neuroticism*.
11. Probably due to the homogenization of the classes.

Emotional Stability (Neuroticism)

Emotional Stability (or Neuroticism) is defined as the predictability and consistence of emotional reactions, without quick mood swings. The emotionally unstable individual is characterized as being worried, short-tempered, introspective, impulsive and lacking in self-confidence, with a tendency towards depression and anxiety disorders.

Various studies use measurements of behavioral and attention problems together with school performance. The scale of behavioral problems largely consists of a mixture of the facets of *Emotional Stability* (such as anxiety or depression) with those of *Agreeableness* (such as aggressiveness). Duncan and Magnusson (2010), and Fergusson and Horwood (1998) established that anti-social behavior in infancy is negatively associated with the chances of the young person completing high school. Cunha et al (2010) used the synthetic measurement of *Non-Cognitive Talents* that is mostly made up of behavioral problem indicators[12], and also showed that people with more non-cognitive skills in infancy find it easier to increase their cognitive indicators, thereby leading the important side effects of personality characteristics to have an effect on educational performance.

In an in-depth study connecting the Big Five personality traits to the final level of schooling reached in a representative sample from Germany, Almlund et al (2011) found that *Emotional Stability*, along with *Conscientiousness*, are the only two statistically significant predictors. It is interesting to note that for males, Conscientiousness was the individual characteristic most associated with the number of years of study, coming in ahead of the measurements of *Fluid* and *Crystallized Intelligence*[13]. Amongst women, however, *Emotional Stability* was the personality characteristic most associated with remaining in school, coming in below *Crystallized Intelligence* and slightly higher than *Conscientiousness*.

If, on the one hand, there is strong evidence to suggest that *Emotional Stability* and its facets are important predictors for remaining in school and for chances of concluding high school, on the other, it does not seem to be true that this attribute is associated with grades and results in standard tests.

12. Scales from the *Behavior Problems Index* (BPI).
13. Fluid intelligence is the capacity to interpret and resolve problems in new situations, not necessarily using knowledge that has been acquired previously. It involves originality and on-the-spot rationalizing, and includes deductive and inductive reasoning. Crystallized intelligence is the capacity to use acquired knowledge and experience to resolve problems. It includes the capacity to activate long-term memory and identify solutions to problems in similar situations that have previously been experienced or understood. It is generally believed that IQ tests measure a little of each of these types of intelligence, but there are specific instruments for measuring each of them.

Motivation and beliefs

Motivation and beliefs refer to the conative skills (relating to "desire") associated with the conscious, proactive effort to perform actions and behaviors (McGrew, 2007). Part of these skills are related to beliefs and orientations found within the person themselves: **Self-Concept**, **Self-Efficacy**, **Self-Esteem** and **Locus of Control** are measures of this type.

Self-Concept[14] is associated with the judgment that the individual has of him or herself, based upon their previous performance in various activities.

Self-Efficacy[15] is related to the expectation that the individual has in satisfactorily completing a task in the future.

Self-Esteem[16] represents the emotional evaluation we have of ourselves, incorporating the reflex of Self-Conception on the emotional state

Locus of Control reflects the measure to which individuals attribute current experiences either to decisions and attitudes they have taken in the past (internal Locus), or randomly, to luck or actions and decisions taken by others (external Locus).

Constructs related to Self-Efficacy, Self-Concept, Self-Esteem and Locus of Control, perform a central role in determining the success and well-being of individuals either by the direct effect they can have on individual decisions ("what" to do, and "how" to do activities), or by the indirect effect that these constructs have on the motivation and interests of the agents.

For some authors (e.g. Almlund et al, 2011), these constructs can be understood as facets of the Big Five Personality Traits, given the manner in which they relate statistically to measurements of these traits. For others (e.g. Kyllonen et al, 2013), these constructs belong to a category that is separate from attitudinal values, influenced as much by personality traits as by context and experience.

There is a great deal of evidence that Self-Concept and Self-Efficacy are closely correlated to academic performance (Denissen et al, 2007; Marsh and Craven, 1997). In meta-analysis, Hansford and Hattie (1982) show that the average correlation of global Self-Conception with performance measures in cognitive tests is 0.21, increasing to 0.42 when only academic self-conception measures are used. Marsh and Craven (2006) add that there is evidence of mutual causality between these two indicators during the run of the educational cycle.

In the case of Self-Efficacy, Multon et al (1991) and Richardson et al (2012) show that this construct is associated with the best grades and the best rates of continuing in school. In relation to higher education, Greene and Miller (1996) point out that people who believe themselves to have the power of self-efficacy tend to take decisions with the goal of actually learning rather than simply getting good grades, whilst also dedicating a large portion of their time to going after their goals (Bandura, 1986 and 1989).

14. Harter, 1982; Marsh e Shavelson, 1985; O'Mara et al, 2006
15. Bandura, 1977 e 2001
16. Branden, 1969

Two measures belonging to this group – *Rosenberg's Self-Esteem Scale and Rotter's Locus of Control*, underpin a series of studies conducted by Heckman and his co-authors, as well as other personality researchers. The outstanding feature of these studies is the care in controlling the problem of reverse causality and dynamic complementarities in the formation of different individual attributes. In the case of Heckman and his co-authors, these measures (or standard indexes based upon them) are used as summary statistics of the wide spectrum of non-cognitive attributes, in opposition to IQ measurements (taken as measurements of cognitive attributes).

The studies by Heckman and his coauthors show that an increased standard-deviation in *Rotter's Locus of Control* is associated to an increase of around 1.5 percentage points in the chances of completing high school – especially for males and individuals at the base of the distribution of this measurement.

1.2 The role of the school in social and emotional development

Analyses of the impact of policies and programs that seek to affect the rate of accumulation of human capital on individuals' personality attributes are relatively scarce. On the one hand, such impact analyses still fail to awaken sufficiently great interest amongst psychologists, whilst on the other, economists, who traditionally conduct this type of investigation more frequently, have only recently started to value the non-cognitive aspects of human development and attempt to understand and control the measuring instruments developed predominantly by psychologists.

Amongst the studies available, the majority focus on small-scale programs in which it is relatively easy to obtain groups of control and randomized treatment, or which are arguably similar in everything except for the fact that one of them takes part in the intervention. The problem of these studies is that the ability to replicate their conclusions in large-scale programs can be questioned, and so too the use of the analysis for the purposes of public policy.

Despite the relative scarcity of studies dedicated to investigating the impact of educational programs on social and emotional development (when compared to studies on cognitive development), scientific writings show a number of examples in very successful interventions in this field. Below, we highlight some of the studies that have undergone stringent evaluation.

Pre-school interventions

High Scope/ Perry Preschool Project

One of the programs that has been most studied through until today has been the **High Scope/Perry Pre-school Project**, implemented in the town of Ypsilanti, Michigan (USA) in 1962. The program involved the offering of high-quality children's education to a group of children considered to be at risk of delayed development. As criteria for eligibility, the children had to show IQ levels of lower than 85 at 3 years old, be Afro-American descent, and be from a low-income family.

Amongst principal characteristics of the program, the following can be highlighted: (i) an innovative curriculum based upon the children's interactivity with the studied objects; (ii) a pre-established and expected routine; (iii) the construction of an environment conducive to learning; (iv) shared control between the adults and children in the choice of activities, privileging the demonstration of the skills of the latter and encouraging their problem-solving abilities; (v) monitoring of the progress of the children through development indicators; and (vi) a specific approach, moving forward in well-defined stages, for the resolution of conflict situations.

In total, it has been estimated that for every dollar spent on the program, at least 16 additional dollars were generated, whilst of these, 7 came through channels that were not salary gains[17]. Since at 15 years old the control and treatment groups returned to similar IQ levels (suggesting the temporary nature of the cognitive benefits obtained), it is plausible that the majority of this incredible return of 16 to 1 was the result of improvements in the individuals' non-cognitive attributes[18].

Amongst the principal results of the program are:

- **One extra year of schooling achieved by 27 years old;**
- **A reduction of 1.3 years, on average, in the use of ongoing special education services (for mental, emotional, speech and auditory backwardness);**
- **A lower proportion of children without formally married parents (57% vs. 83%);**
- **A lower rate of teenage pregnancy (0.6 vs. 1.2 child/woman);**
- **A lower rate of possible imprisonment (28% vs. 52%);**
- **A lower proportion of arrests for violent crime (32% vs. 48%);**
- **Higher salaries (around 40% higher for those treated).**

Carolina Abecedarian

Various interventions similar to 'Perry' have since been evaluated, some of them reaching similar conclusions and others being inconclusive. **The Carolina Abecedarian Project** (Chapel Hill, 1972) extended the Perry Pre-school Project to children who were just a few months old and to up to 5 years old[19], obtaining similar results in the non-cognitive fields, but with the advantage of also having caused a permanent increase in the IQs of the participants (with an annual rate of return for society of 3.4%, with only 1/3 of this resulting from salary gains).

17. It could be said that the annual rate of return from the intervention was around 8%, which could be considered high (the average rate of return in the post-war US share market is 5.8% per year, and the return on Government Bonds is 1% per year).
18. In fact, indirect measurements of the personality characteristics were substantially affected by the program (for example – the incidence of theft, lies, absences and the use of swear words, as reported by primary education teachers)
19. An important modification for the abecedarian was the maintenance of extremely reduced-size classes, moving from 1:3 for newborns, to 1:5 for 5-.year-olds.

The STAR Project

The **STAR Project** in Tennessee (USA) allocated children randomly in different-sized pre-school classes. Just as in the Perry Project, the cognitive benefits were fleeting (they had disappeared by the 8th grade[20]), this being different to the significant salary gains obtained in adulthood and markedly better behavior according to the teachers of the 4th and 8th grades[21], who quantified the students in terms of effort, initiative, interest in classes and inconvenient behavior. These facts suggest that the channels by which the size of the classrooms could have affected salaries were in fact due to impacts on non-cognitive attributes.

Interventions designed to develop social and emotional skills

PATHS (Promoting Alternative Thinking Strategies)

Educational interventions designed to influence solely non-cognitive aspects are very rare. Amongst them, **PATHS** is a curriculum that has a number of rigorous evaluations.

The principle of the PATHS Program is that all feelings are essentially acceptable, but not every form of behavior is acceptable as a response to the feeling. It is included amongst the curriculums based on social and emotional learning. The results show significant improvements in aggression, pro-social behavior and engagement in academic activities, as well as helping to improve grades.

According to the PATHS pedagogy, students create their representations of possible feelings ("boxes of feelings") and outline possible replies, classifying them a priori from between "stop and calm down" (red), "go slowly" (yellow) and "go ahead with your plans" (green). After the activities, the students revisit their boxes and signs and reclassify the attitudes in accordance with the success they achieved.

Interventions in Adolescence

Empresários pela Inclusão Social (Entrepreneurs for Social Inclusion /EPIS)

Interventions in more advanced age groups are less common, but there is at least one example of a very successful program designed to develop personality traits in adolescence. In Portugal, the **Empresários pela Inclusão Social (EPIS)** ('Entrepreneurs for Social Inclusion'), being concerned with improving income and reducing school drop-out levels in the 7th and 8th grades[22], dedicates a great deal of effort to locating students at risk of dropping behind, and directs them towards sets of individualized activities according to their respective difficulties.

A large number of the modules focus on the promotion of non-cognitive attributes, and are conducted by mediators (specialized professionals) in small groups or one-on-one. Amongst the individual activities are motivational discussions and the application of self-control and problem resolution techniques. As a group, the students learn to study and work as a team, work on their social skills, and practice controlling their anxiety and excessive criticism of others[23].

20. In the American education system, primary education is divided into two cycles of four years (elementary and middle schools), and high school consists of four years. The 8th grade this case is the last grade of middle school.
21. Under the current Brazilian grades system – 5th and 9th years.
22. Under the current Brazilian grades system – 8th and 9th years.
23. From the start, the chosen participants and their mediators establish performance goals and discuss where they aim to get to. The meetings take place on a regular basis, but outside school time.

At the time of the program's evaluation (2010), 65 mediators financed by 90 large companies were taking care of between 50 and 100 students. A scientific committee contributed to the designing of the modules and offered suggestions to the supervisors and senior members of the administrative staff. **In total, 10% of the country's students (around 15,000) took part in the EPIS program.**

The results (Martins, 2010) show that there was a reduction of 10 percentage points in repetition rates (with a cumulative effect during the period of primary education of up to 30 percentage points on the chance of having at least one repetition), grades improved substantially, and the cost-benefit relationship was much more favorable.

National Guard Youth Challenge

In another example of effective intervention amongst young people, the **National Guard Youth Challenge** program, focused on adolescents who skip school, arranges 17 months of counseling and activities that emphasize motivation and discipline[24]. In an evaluation performed after 9 months, the participants already showed a greater probability of completing high school, of being in full-time employment and being less likely to be arrested, as well as achieving higher levels on the *self-efficacy* scale.

General Education

Heckman et al. (2006) ask whether formal education, from primary school onwards, significantly influences cognitive and non-cognitive developments, based upon the structural model incorporating American longitudinal data. The most important conclusions of the work are that there exists a significant impact of great magnitude in multi-dimensional intelligence tests[25], especially as of the end of high school; and that there is an even greater impact on self-esteem and on the internal locus of control. In terms of self-esteem, it is primary school that is of most importance, with little effect being found after it has finished. As for the locus of control, the impact is great and continues growing along with the level of education.

24. Starting with two weeks of preparatory meetings; followed by 20 weeks spent in residence, generally on a military base, and ending with one year of tutoring and counseling with a specialist.
25. Arithmetical Reasoning, Vocabulary, Reading Comprehension, Mathematical Knowledge and Speed of Programming.

1.3 Summary: what the evidence suggests

The aim of this Chapter has been to offer a summary of theoretical frameworks and the empirical evidence that exists on the importance of social and emotional skills for academic learning and the establishment of well-being for life.

As explained, the scientific evidence largely confirms the importance of the skills proposed by global education initiatives for the 21st century, and shows that the choice influences, to a large extent, the development of social and emotional traits associated with success.

In fact, research has shown that social and emotional characteristics contribute about the same as the cognitive ones in determining academic success, in terms of grades, probability of dropping out and the final level of schooling achieved. Furthermore, in the work market, social and emotional characteristics are compensated in the form of higher salaries and shorter periods of unemployment. In both cases, the attribute of Conscientiousness, which includes the facets of responsibility, discipline and perseverance, appears to be the most relevant.

In the following chapters, the Project's stages involving measurement of social and emotional skills are covered. These stages were intended to construct a tool for the measurement of these skills in the context of Brazilian schools and validate the tool by applying it to a representative sample of students with the intention of contributing to the investigation of the relationship between social and emotional development and academic learning in Brazil.

Chapter 2:

Project for the measurement of social and emotional skills in the school context

2.1 Construction of the instrument for measurement of social and emotional skills

As explained in the introduction of this report, one of the great challenges for this initiative was to construct an instrument to measure social and emotional attributes that can, at the same time as serving as an instrument to monitor and support the formulation of public policies, also serve as a reliable gauge of the non-cognitive development stage of individuals at the moment of application. In other words, we needed an instrument that was simple and robust enough to be applied to large numbers of students in order to provide a comprehensive picture of the distribution of social and emotional characteristics in Brazil, whilst at the same time being precise and interpretable enough to be able to be used scientifically in studies focusing on documenting the social and emotional development of the individuals over the cycle of their lives and investigate how interventions affect this trajectory. The aim of this chapter is, therefore, to describe the process of construction of this instrument.

The starting point in this process was for a team of researchers to carry out evaluations of a large set of psychometric instruments that have long been established in international literature[26] and which, in principle, were considered to be possible candidates for satisfying the objectives of the job. Each of the selected instruments was considered from four perspectives relevant for their use, after which eight were chosen to be tested in a school environment. It was recognized that all instruments showed themselves to have interesting properties, but none of them carried the broad spectrum of features that were required for the measurements, nor did it seem practicable to use all of them at the same time with all the children we were hoping to sample. We therefore decided to create a new instrument that grouped together the best items from the eight selected.

In the next section, we will be presenting the pre-established criteria for which a psychological instrument could be considered suitable for use in this initiative. Section 2.1.2 describes the results of the evaluation performed on a total of more than 70 pre-selected instruments.

26. The team of researchers responsible for evaluation of the instruments was made up of professionals with different technical skills, amongst which were psychologists, psychometrists, specialists in educational evaluation and specialists in impact evaluation.

One important limitation of this evaluation was that the most important criteria for the inclusion of psychological instruments in our pre-selection was the existence of empirical evidence that the constructs they measured were statistically associated to measurements of the success and well-being of the individuals. As such, older instruments and/or instruments that had been used within databases in the public domain naturally had a better chance of making it into our pre-selection rather than more recent instruments (which possibly resulted in creating improvements on the earlier ones).

The results of the evaluation of the pre-selected instruments was therefore extensively discussed with both specialists and education administrators, and from these conversations there emerged a new set of instruments that, similarly to the earlier ones, incorporated more recent innovations and improvements, at the same time as they aligned themselves with the interests and restrictions of the education system. The final sections are dedicated to a description of this discussion and the choice of instruments that served as a base for creating the final instrument.

2.1.1 Selection of instruments based upon scientific evidence: relevance and suitability criteria

As explained above, the great challenge at this stage was to find an instrument that was both useful for periodic application in the classroom and consequent use in the formation of educational policy, and also allowed for monitoring of the development of students' non-cognitive skills over time. To meet the first goal, the instrument needed to be simple and cheap so that it could be used on a large-scale with random samples representative of the registered students and children of school age as a whole. It also needed to be strong enough to withstand interferences in the context to which it was to be applied, and preferably be suitable for students of different ages and in different grades. Finally, the language and format would need to be in keeping with that used by the teachers in the classroom.

In relation to the second objective, the instrument needed to be wide reaching, consistent with the *a priori* 'blind' position concerning social and emotional aspects influencing and being influenced by the school environment. It should also include internationally recognized and validated aspects since its use for scientific purposes is potentially greater if it can easily be adapted to other cultures, thus allowing the monitoring of process of students' social and emotional development in other countries.

In summary, in order to achieve the aims of the Project, the final social and emotional skills measurement instrument needed to be:

- **Simple, suitable and economically viable for large-scale use in education systems, in a way that could support the monitoring and formation of public policies;**
- **Sufficiently precise and interpretable to be used scientifically in studies intending to document the social and emotional development of individuals throughout their lives and investigate how interventions affect this trajectory;**
- **Robust enough to not suffer interference in the context in which it is applied;**
- **Preferably suitable for students in different age groups and grades;**
- **Wide-reaching with respect to the characteristics measured, in such a way as to reflect an a priori 'blind' stance concerning which social and emotional aspects most influence and are most influenced by the school environment;**
- **Made up of internationally recognized and validated items, since its use for scientific purposes is potentially greater if it can be easily adapted to other cultures.**

In the first phase of the construction of the tool, a team of five professors from the departments of Economics and Psychology at the University of São Paulo and São Francisco University, assisted by six post-graduate fellowship students, scoured scientific magazines relating to different areas for articles discussing social and emotional constructs measured by self-report and hetero-report questionnaires with variables relating to individual and collective well-being.

The requirement of statistical evidence associating the construct with measures of well-being was adopted to avoid the selection of instruments being influenced by the idiosyncratic favoring of particular constructs that would perform a central role in certain psychological theories, whilst also allowing a larger number of constructs and instruments to be investigated and a theoretically neutral discourse to be constructed in defense of the adoption of those instruments that would possibly be chosen, thus avoiding prior judgment of the choices by adherents to other theories that had not been contemplated.

In total, 113 instruments were identified as having constructs statistically associated with some sort of measurement of results associated with well-being, and of these, 72 were effectively evaluated. In order to achieve an objective judgment, the researchers were asked to classify each instrument within four categories directly related to the instrument's degree of suitability to the aims of the initiative, as will be described below.

Predictive Power

Firstly, the existing empirical evidence was analyzed in relation to whether the constructs measured by the instrument in question were in fact causing changes in the individuals' well-being. Even though each one of the 72 pre-selected instruments had to present some sort of statistical relationship with measurements of well-being published in a scientific magazine, there were instances in which this statistical relationship just meant a simple statistically significant correlation in a cross-section, and only under very strong hypotheses they could be interpreted as a causal relationship. Meanwhile, in other cases, it could be seen that much greater attention had been paid to this association, which could be much more convincingly interpreted as an effective cause-effect relationship.

In the analysis of this section, called the instrument's 'Predictive Power', the following criteria were used:

i. range: an instrument that is associated with a wider set of measurements of well-being is preferable to another that is associated with a more restrictive set;

ii. extent: the greater the extent of the estimated statistical association, the more important it is to monitor the evolution of the construct measured by a certain instrument, since slight changes in this trajectory can cause enormous alterations in the levels of future well-being;

iii. internal validity: this measures how convincingly the estimated statistical relationship (from amongst the many well-being instruments and measurements) can be interpreted as causal [27];

iv. (external validity: studies that used representative samples of larger populations took precedence over studies that used small groups of individuals that could possibly present isolated characteristics. If the predictive power of a certain skill had been established by studies with small groups, its conclusions could not necessarily be extended to larger communities, meaning they received a lower grade.

27. t is difficult to randomize numbers of non-cognitive characteristics amongst individuals and estimate their impact on the measurement of well-being. The highest marks in this sub-category were given to instruments where there already existed studies in which average scores in younger age groups were associated with averages of well-being collected from the same individuals when they were older. Estimates of impact based upon regression strategies with discontinuity and instrumental variables received good evaluations when the respective discontinuities and instruments were convincing. Estimates based upon multiple linear regression analysis by least squares in cross sectional analysis came next, only ahead of simple correlation analyses in cross section.

Feasibility

To meet the proposals of this Project, we searched for simple, resistant measures that could integrate broad, longitudinal databases (possibly in different cultures) and thus allow scientific investigation into how the education system and its characteristics influence this trajectory over time, as well as to what extent students from different nationalities present particular characteristics along their paths of development.

To meet all these objectives, any instrument chosen would have to offer low application costs, not require the involvement of specialized teachers, be sparing in the number of items included, and be easily accepted by teachers, education administrators and members of the community. It is important to note that only instruments in the form of questionnaires were investigated, since it would be difficult to regularly implement other forms of measuring (such as tasks[28]) on a large-scale whilst meeting a single application protocol.

In analyzing the feasibility of the instrument, the evaluators considered:
i. The cost of implementation: the existing instruments could only possibly be used upon purchase of a license[29], whilst the cost of the license may or may not be proportional to the world in which it is to be applied, and licenses may or may not be regularly renewable – expensive instruments are, ceteris paribus, less attractive;
ii. Ease of application: long questionnaires, with elaborate language or which require other materials in order to be applied (for the performance of tasks) received lower grades;
iii. Need for an applicator: instruments sometimes need to be applied by third parties[30], Self-applicable instruments (that is, self-report questionnaires) were subsidized[31];
iv. availability of a validated version in Portuguese: instruments that have undergone the most rigorous processes of validation and adaptation to the Brazilian context should be prioritized;
v. acceptability by society: this aspect is subjective, but that does not make it any less important. If a certain skill is important, but its measurement involves taboos or questions that could be possibly embarrassing, the instrument should therefore be penalized.

Malleability and transformative capacity in school

As the most important objectives of this initiative are those that document the social and emotional development of children and young people of school age, and identify the role of the school's characteristics in this process, it is preferable that the chosen instruments measure constructs that are flexible within the ages covered by the education cycle and that can be modified by the characteristics of the education system.

In order to analyze this dimension, the researchers once again turned to scientific literature in search of empirical evidence that would support the malleability and ability of the school to transform the construct.

28. It is possible to obtain measurements of socio-emotional characteristics through both questionnaires (in which the individual, or someone who knows them well, describes their behavior) and asking the person to complete a task. One example of a task is the 'Mischel Task' (or 'Marshmallow Test'), in which a candy is placed in front of a child and the researcher tells them that he is going to leave the room for a while and if, when he returns, the child has not eaten the candy, he'll be given a second one. During the researcher's absence, the child is filmed and the time it takes before he succumbs to temptation and eats the candy is timed. The results of this task measure the ability to postpone rewards, which is seen by many as a facet of Conscientiousness.
29. The cost of licenses may or may not be proportional to the world in which it is applied, and licenses may or may not be renewable from time to time.
30. Maybe because the age group makes it difficult to be filled in by the young person that is being gauged, or maybe because there are institutional restrictions (such as the requirement that the applicator be a psychology graduate).
31. There are three reasons for this. Firstly, it is difficult to foresee all the uses that the instruments will have for public administrators. If the volume of resources destined for a school is possibly associated with the results of the evaluations, an instrument filled out by the teacher may be biased, since this individual may have reason to manipulate the results in favor of the institution. Secondly, the teacher's access to the individual results of the students could stigmatize the students and make it difficult for the instrument to be accepted by professional boards and ethics committees. Finally, if it is necessary for the instrument to be applied by a professional from outside the education system, the application costs could rise and in some regions there could be a scarcity of individuals to apply it

In this case, the criteria which guided the evaluation were:
I. internal validity: the most convincingly methodical estimates that the construct is in fact modifying itself to the target ages and/or that it is being modified by the school, received a better evaluation than less convincing (or non-existent) estimates;
II. external validity: estimates obtained from more wide-reaching and diverse groups tended to be more convincing than the social and emotional development process described and are valid for the world of students as a whole rather than just one specific group and, as a result, are valued more highly in this area of evaluation;
III. age range: a skill that remains flexible throughout the entire education cycle is relatively more important for our proposal than another that is flexible for just a short period of time;
IV. Transforming capacity of the school: studies that simply state that the school affects certain skills are relatively less informative than those that actually implement the changes.

Psychometric Properties

The fourth and final dimension investigated involved the psychometric properties desirable in the selected instruments. In this analysis, two measurements available for virtually all the pre-selected instruments were compared, and instruments with higher values were better evaluated.

The gauges for evaluation of the psychometric properties were:
i. Cronbach's Alpha, an indicator of the internal consistency of the instrument. Psychometric instruments generally try to measure latent constructs based upon observed behaviors. In one questionnaire, the behaviors are the responses to constant items and the instrument is internally consistent if differences in the pattern of replies obtained amongst individuals can be rationalized (only) by the fact that these have distinct levels of latent constructs supposedly measured by the instrument.
ii. Test-retest correlation. A robust instrument should not be significantly affected by the environmental conditions in which it is applied. High correlations between the results of a single test applied close together in time suggest that the individual's replies are not greatly influenced by the prevailing environmental conditions at the time of the application.

2.1.2 Selection of instruments based upon scientific evidence: results of the evaluations

The 72 pre-selected instruments were grouped into six categories, depending upon the constructs they intended measuring. It was believed that this way of comparing the instruments would be easier, making the analysis more accurate. In four of these categories, one or more instruments were found with characteristics that were very close to those sought. The categories used and the respective evaluation results were:

Personality inventories

Personality inventories: wide-ranging questionnaires that, in principle, should cover the entire range of personality traits.

The instruments evaluated were:

a. Eysenck Personality Inventory (EPI)
b. Escala fatorial de extroversão (Extroversion Factorial Scale) (EFS)
c. Escala Fatorial de Neuroticismo (Factorial Scale of Neuroticism) (EFN)
d. Adolescent Personal Style Inventory (APSI)
e. Factorial Battery of Personality (BFP)
f. Big Five Inventory (BFI)
g. Revised NEO Personality Inventory (NEO)
h. California Psychological Inventory (CPI)
i. Omnibus Personality Inventory (OPI)
j. Scale of Personality Traits for Children (BFC)

Of this group, the Big Five Inventory (BFI) was classified as being appropriate for the intentions of the initiative. We found a reasonable number of good studies which demonstrated its predictive power on positive affect, happiness and personal satisfaction; the instrument is short (44 items), free and there is a version in Portuguese; the psychometric properties are good and well documented in different cultures; the majority of the constructs measured are especially flexible for the target ages.

Social Skills

Social Skills: questionnaires with a fairly wide reach, focused on the measurement of constructs related to interpersonal or intrapersonal talents that have direct consequences on the ability to socialize. In many cases they were especially developed for use in the classroom, as a support instrument for the teacher and pedagogic coordinator.

In this category, the following instruments were evaluated:

k. Battelle Developmental Inventory (BDI)
l. Behavior Problems Index (BPI)
m. Behavioral and Emotional Rating Scale Second Edition(BERS-2)
n. Devereux Early Childhood Assessment (DECA)
o. Devereux Student Strengths Assessment (DESSA)
p. Grasmick et al.'s Low Self-Control Scale (GRASMICK)
q. Home & Community Social Behavior Scales (HCSBS)
r. Inventário de Habilidades Sociais (Social Skills Inventory) (SSI)
s. Inventário de habilidades sociais para adolescentes (Social Skills Inventory for Adolescents) (SSIA)
t. Matson Evaluation of Social Skills with Youngsters (MESSY)
u. Prosocial Behavior Questionnaire (PBQ)
v. Pupil Evaluation Inventory (PEI)
w. Revised Class Play (RCP)
x. Selection, Optimization, and Compensation Questionnaire (SOC)
y. School Social Behavior Scales, Second Edition (SSBS2)
z. Sistema Multimídia de Habilidades Sociais de Crianças (Children's Social Skills Multimedia System) (SMSH)
aa. Social Skills Improvement System (SSIS) Rating Scales
bb. Social Skills Rating System (SSRS)
cc. Socio-metric Ratings and Nominations
dd. Student-Teacher Relationship Scale (STRS)
ee. Vineland Adaptive Behavior Scales, Second Edition (Vineland-II)

From this group, the Social Skills Improvement System (SSIS) was elected as that best suited to the intentions of the initiative. Amongst its qualities, it should be highlighted that the results of the instrument's measurements of pre-school children strongly predict the future educational performance and the chances of having a cognitive deficit. The instrument has numerous versions, some involving self-reporting and others hetero-reporting, and it can be used with younger age groups, whilst it is also relatively short and partially validated[32] in Portuguese.

Externalizing and internalizing characteristics and characteristics of mood

Instruments that predominantly measure **externalizing and internalizing characteristics**, as well as characteristics of mood, and which may also be used as a support instrument in the formation of diagnostics by psychologists. One of their frequent characteristics is to measure more precisely the extreme values of social and emotional characteristics in question.

The following were included in this set:

a. Child Behavior Checklist (CBCL)
b. Jesness Inventory-Revised
c. Minnesota Multiphasic Personality Inventory-2 (MMPI-2)
d. Behavior Assessment System for Children, Second Edition (BASC-2)
e. The Woodlawn Program Survey - TOCA-R e POCA-R
f. Depression Self-Rating Scale for Children
g. Strengths & Difficulties Questionnaires (SDQ)
h. Communities That Care (CTC) Youth Survey
i. Social Anxiety and Phobia Inventory for Children (SPAI-C)
j. Rutter Child Scales
k. Preschool Behavior Questionnaire (PreBQ)
l. Eyberg Child Behavior Inventory (ECBI)
m. Conners Rating Scales-Revised (CRS-R)
n. Inventário de Expressão de Raiva como Estado e Traço (State-Trait Anger Expression Inventory) / STAXI-2)
o. Caregiver-Teacher Report Form
p. Escala de Transtorno do Déficit de Atenção / Hiperatividade (Inconvenience of Attention Deficit Scale / Hyperactivity) (TDAH)
q. Center for Epidemiological Studies (CES) – Depression Scale
r. Child Behavioral Questionnaire (CBQ)
s. The Early Adolescent Temperament Questionnaire Revised (EATQR)
t. Indicadores Clínicos de Risco para o Desenvolvimento Infantil (Clinical Indicators of Infant Development Risk
u. Early Development Instrument (EDI)
v. Ages and Stages Questionnaires (ASQ)

In this group, the 'Strengths & Difficulties Questionnaires' (SDQ) showed itself to have desirable features, predicting, amongst other aspects, individual performance in cognitive tests. Furthermore, the instrument is free, there is a version in Portuguese, it is self-reported, and has already been used in the school environment.

32. The Social Skills Rating System (SSRS) is part of the SSIS, of which there is a version in Portuguese.

Motivation and Beliefs

Motivation and beliefs: in this category, there are scales that seek to measure the Locus of Control, Self-concept, Self-efficacy and Self-esteem.

The group is made up of:
w. Rotter's Locus of Control
x. Pearlin Mastery Scale
y. Belief in Personal Control Scale (BPCS)
z. Rosenberg Scale
aa. Self-Perception Profile for Children (SPPC)
bb. Escala de Autoconceito Infanto-Juvenil (Infant-Juvenile Self-concept Scale) (EAC-IJ)

Within this group, Rotter's Locus of Control and the Rosenberg Self-esteem Scale received good evaluation reports. As far as the first is concerned, the extensive evidence of predictive power in longitudinal studies on salaries, schooling and professional success can be highlighted, as well as the fact that it is a short instrument that is easy to apply and has been validated in Portuguese (whilst a great number of international comparisons have also been made). As for the second, there is equally extensive evidence of its predictive power on salaries and entry into more highly regarded occupations, as well as on educational performance. The scale is small and has been validated in various different languages, including Portuguese.

Professional Interests and 'miscellaneous'

Professional interests: professional interests can be understood as the readiness to involve oneself in certain labor/occupational activities, being capable of influencing the subjects' decision-taking in relation to career options and occupational activities.

Included here were:
cc. Holland's self directed search (questionnaire)(SDS)
dd. Vocational Preference Inventory (VPI)
ee. Avaliação dos Tipos Profissionais de Holland
(Holland's Evaluation of Professional Types) (ATPH)
ff. Escala de Aconselhamento Profissional (Professional Counseling Scale)(EAP)
gg. Escala de autoeficácia para escolha profissional
(Self-efficacy Scale for professional choice) (EAP-EP)

None of the instruments in this group received an evaluation sufficiently satisfactory to later be considered as suitable for the intentions of this initiative.

Miscellaneous: the final group of instruments is essentially made up of inventories and scales that do not fit into the other categories. Just as in the previous case, none of these instruments were considered satisfactory for this initiative.

hh. Classroom and School Community Inventory (CSCI)
ii. Life Orientation Test (LOT)
jj. Escala de Maturidade para a Escolha Profissional (Scale fo Maturity for Professional Choices) (EMEP)
kk. Inventário de Percepção de Suporte Familiar (Perception of Family Support Inventory)
ll. Alcohol, Smoking and Substance Involvement Screening Test (ASSIST)
mm. Torrance Creativity Test
nn. Child Health Questionnaire (CHQ)
oo. Positive and Negative Affect Schedule (PANAS)

As in the previous case, none of these instruments were considered satisfactory for the needs of this initiative.

2.1.3 Discussion of the selected instruments: consultation with specialists and those responsible for the education systems

Limiting the instrument selection strategy based upon a revision of the scientific literature resulted from the fact that the various existing psychological instruments compete under unequal conditions in the criteria that involve publication of the results in scientific magazines. Firstly, the older instruments were more likely to have been used in a greater number of studies, therefore having more chance of being empirically associated with measurements of well-being and having their malleability and ability to be transformed by the activities of the school demonstrated. This means that more modern instruments may have been unfairly left out, which could be an important setback given that this area of knowledge has been undergoing unprecedented transformations recently.

Secondly, instruments that have been included in databases found in the public domain were studied more frequently, thus once again meaning an advantage over the others with respect to the possibility that there are studies documenting their statistical relationship with variables of well-being, malleability of school ages and sensitivity to the actions of the school. As the criteria for such instruments to appear in the public domain databases was often the very feasibility of including it, the evaluation of Predictive Power performed in the previous stage is a part that was placed on the evaluation of feasibility, possibly prejudicing instruments that were a little less feasible, but which contained various other advantages.

Due to the suspicion that important instruments could have been left out of the list produced during the first stage, it was decided that leading international specialists producing material in the area of Personality Psychology should be consulted, with the intention of increasing the list of instruments selected. The most important conclusions drawn from these consultations were:

• The instruments in the self-perception category, Rotter and Rosenberg, although being established and well-respected, do not represent the group of instruments existing in this field, where new instruments have recently been created with very promising characteristics. It was proposed that the Rotter's Locus of Control instrument should be replaced with that of Norwick-Strickland. Also added in were the 'Self-Efficacy Questionnaire for Children' (SEQ-C) and the self-valorization scale 'Core Self Evaluations' (CORE).

• Inclusion of the 'Grit Scale' was suggested, as it measures the individual's ability to set him or herself a goal and persevere until it is reached. The 'Grit Scale' has been especially used in schools and has been widely accepted by educators in the United States.

• It was suggested that the Hierarchical Personality Inventory for Children (HiPIC) should be included. There were two reasons for this. Firstly, it was recognized that the BFI instrument contained items that could possibly be difficult for small children to understand. Additionally, it was necessary for there to be an instrument that could be used right from the first years of the education cycle, in order that the students' development could be monitored in these dimensions from the very beginning. As the instruments selected in the first stage strongly prioritized self-report instruments in which the young people responded to questionnaires focusing on their behavior, it was recognized that children under 10 years old would not be able to provide precise information on their social and emotional characteristics. The possible inclusion of the HiPIC, of which there are auto-report and hetero-report versions and which covers a wide range of personality characteristics, could allow for the measurement of the first years of the education cycle. This would be at the same time in which, being applied to both auto and hetero report versions and together with other instruments with higher age groups, it could allow the latent constructs relating to social and emotional characteristics to be obtained, even amongst younger children (thus allowing the idea of having longitudinal monitoring of non-cognitive development throughout the entire cycle). As well as the HiPIC, the 'Big Five for Children' (BFC) instrument was suggested as an alternative to the BFI.

2.1.4 Final list of instruments that were candidates for the social and emotional skills measurement instrument

Following consultation of specialists and education administrators[33], the list of instruments that would possibly contribute to the final social and emotional skills measurement instrument was as follows[34]:

- **Core Self Evaluations**
- **Grit Scale**
- **Hierarquical Personality Inventory for Children (HiPIC)**
- **Norwick-Strickland Locus of Control Scale**
- **Rosenberg Self-Esteem Scale**
- **Social Skills Improvement System (SSIS)**
- **Big Five for Children (BFC)**
- **Big Five Inventory (BFI)**
- **Self-Efficacy Questionnaire for Children**
- **Strengths and Difficulties Questionnaire (SDQ)**

33. Professionals from the Rio de Janeiro Municipal Education Department and the Rio de Janeiro State Education Department.
34. To check the references of these instruments, please consult the bibliography attached to this report.

In the following section, the stages of translation, adaptation and revision of these instruments will be described, as well as the studies upon which we based the selection and composition of items for the construction of our unique instrument for their measurement of social and emotional skills for application on a large-scale.

2.2 Studies that preceded large-scale application of the evaluation of social and emotional skills

As was explained in the previous section, following consultation with specialists for a revision of the list of selected instruments, we arrived at a group of ten instruments that would be applicable on a large-scale and which brought together all the important constructs for this Project.

Based upon this list, a series of preliminary studies was conducted aimed at:
a) Translating and analyzing the instruments in order to identify areas of weakness such as language and/or content that may possibly be unsuitable for the target public;
b) Performing a qualitative analysis of the comprehension of the items to establish the students' degree of understanding of the questions as well as their relevance to the respondents' socio-cultural realities;
c) Checking the time required to perform the tests with the students;
d) Performing the psychometric analysis and selecting those items for inclusion in the final instrument.

These studies preceded the principal research evaluating social and emotional skills on a large-scale. Their main goal was to create conditions for the development of a unique instrument with psychometric properties verified beforehand for use in the pilot project.

2.2.1 Translation, adaptation and content review of selected instruments

Going beyond a simple transference into another language, the process of translating psychological instruments requires their content to be adapted to the target culture in a manner that maintains the planned relationship between indicator and construct. To do so, it is often necessary to incorporate alterations or adaptations that require bilingual professionals and individuals who understand the constructs being measured (Borsa, Samásio & Bandeira, 2012).

The **instruments** are a generally standardized means of obtaining behavioral samples/indicators that reveal individual differences in relation to constructs, latent traits or underlying mental processes. When giving their answers to the instruments, the children say, for instance, that they "find it easy to make friends" or "like to chat" - proposals that describe expressions of a broad underlying inclination called **extraversion**, defined as the "orientation of interests and energies towards the external world, people and things (instead of the internal world of subjective experience)". It is presumed, therefore, that the latent traits are the causal variables of behavior that are manifested in the test situation. As such, the process of measurement consists of an indirect route that, by means of observation of the indicators, makes it possible to infer something about the construct one wishes to evaluate.

The process of translation/adaptation of the instruments listed was performed in five steps:
I. *Translation:* researchers from the group made between two and four independent translations of the instruments. When there was already a version of the instrument in question available in Portuguese, this version was included in the analyses as an additional translation. Therefore, even those instruments that already had Portuguese versions available in related publications were translated by the group's researchers.

II. *Examination by a focus group of administrators:* two groups of administrators from the Rio de Janeiro state and municipal education systems examined the items and the available translations with the intention of checking their suitability for the target public and recognizing possible negative reactions on the part of the teachers, directors and parents in relation to the questions asked of the children. These focal groups' guiding questions were:

- Are the questions appropriate for the child's age?
- Can children with reading difficulties answer the questions?
- Are the questions appropriate to the school context?
- Are they appropriate to the children's socio-economic reality?
- Are they useful to know the planning of activities for children?
- Could any questions cause embarrassment?
- Could any questions create a risk of stigmatization?
- Could any questions generate problems with parents, teachers or unions?
- Are there any other problems?

The comments from these sessions were noted and compiled together with the translations done in step 1.

III. *Consensus translation:* a final translation proposal was selected and a summarized version of the instruments was created including suggestions obtained in the previous phases.

IV. *Sample study for a better insight into the questions:* a pilot study was undertaken with a group of students aimed at checking whether the children understood what was being asked and if they knew how to grade their answers according to *Likert* five-point scale[35]. It was also checked that what was being asked was suitable for the age group and socio-cultural reality. In this phase, the items[36] that presented a more abstract vocabulary were selected and these were used to create four notebooks with 20 or 30 items in each one. These notebooks were applied to a sample group of 48 children from 5th to 9th grades divided into four groups with four or five children in each one. In some sessions, the students read the items with the help of the researchers. In others, they just responded to the tests and, at the end, explained to the researchers what they had understood about each item. The inputs gained from this study served to review the items in the instruments.

V. *Finalization and back translation*[37]*:* With the revisions suggested during these phases having been implemented, a final version of each test was established. Later, all the tests were translated back into English. These translations were sent to the authors of the instruments for authorization.

35. Subjective numeric points scale that operationalizes the degree of agreement with assertions about personal characteristics in self-report questionnaires used in research in different areas (psychology, education, medicine and others). The name is taken from its inventor, the psychologist Rensis Likert.
36. N=154 items
37. A retro-translation process; that is, the translation of a text back into its original language after it has been translated into another language. This is a process frequently used to evaluate the faithfulness of translations and which consists of the following process: a professional translates a text from language A to language B; another professional then translates it back from language B to language A. The two versions in the original language are analyzed to check their similarity.

In the study performed by the focal group, the administrators considered a large part of the items from the Social Skills Improvement System (SSIS) to be unsuitable. As there was a certain amount of intersection between the constructs underlying the SSIS and the subscales of other instruments, this was removed from the list. The Hierarchical Personality Inventory for Children (HiPIC) was very long and not yet studied in Brazilian samples; as there were already two instruments measuring the Big Five with studies in Brazil, this was set aside for future studies.

The final product of this phase was the version translated and adapted for use in the school context of the instruments listed[38]. After these studies, a set of eight tests was defined to be used in subsequent studies: the Big Five Inventory (BFI); the Big Five Inventory for Children (BFC); the Strengths and Difficulties Questionnaire (SDQ); the Self-Efficacy Questionnaire for Children (Grit) Scale; the Rosenberg Self-Esteem Scale; the Core Self-Evaluations, and; the Norwick-Strickland Locus of Control Scale.

2.2.2 Average time required to answer the instruments

The aim of this preliminary study was to define the maximum number of items that could be answered in a 40-minute session (the length of time defined for the main, large-scale study[39]), especially by younger children and those with greater reading difficulties. The sample for the study was made up of 228 children drawn from two groups from the 5th grade (N=60), two from the 6th grade (N=42) and four from the 10th grade (N=126).

> ***Methodology:*** *The eight tests were randomly distributed to the children in each class and they were told to answer to the items in the order in which they appeared. After this first test, the students received another, until the time of the session was completed. The start and end times of the tests were recorded. In parallel, while the tests were taking place, the researchers gave a sign that the students should change pens every ten minutes, meaning they would be using different colors. This strategy allowed the researchers to record the number of items answered in each ten minutes .*

The results are presented in Table A in the annexes to this report. On average, the students answered 4.1 items per minute, meaning they take longer in the Rosenberg Test (1.6 items per minute) and less time in the BFI and BFC (5.5 and 5.1 items per minute). The 5th grade students answered 2.4 items per minute and the subgroup of slower students answered 1.5 items per minute.

Adopting a conservative criterion, it was concluded that the test for 5th grade students should contain 60 items at the most and, for students from the 10th to 12th grades, this could rise to 135 items. Finally, it was established that the final test should contain between 60 and 90 items in order to offer conditions that are favorable for the students to obtain more consistent answers, even amongst those with greater difficulties.

38. It is important to stress that all these instruments were freely available for use in academic research.
39. The total class time is 50 minutes, with 10 minutes being reserved for instructions on the test and how to fill out the socio-economic questionnaire.

2.2.3. The Underlying Latent Structure of Tests for Non-Cognitive Constructs in Education

As shown in the first chapter of this report, the set of instruments selected for the pilot project contained measurements of **personality traits, self-concepts, self-esteem, motivation, attitudes and belief in control of events** (attributing control internally or externally) and **social and emotional adaptation issues.**

For some authors, these constructs can be understood as being facets of the Big Five personality areas, given the way measurements of these domains relate statistically. **Kyllonen, Lipnevich, Burrus & Roberts (2008) suggest that organization of the relevant psycho-social skills in the educational context included in the concept of "non-cognitive constructs" can be based upon the Big Five personality traits model.**

This idea suggests that, even though there are 21 sub-scales in the set of instruments (see **Table 3** below), these may be reflections, or in other words, effects of the latent traits relating to broad factors of personality.

In this work, the authors suggest an understanding of scales that are measured in the selected instruments as based upon the five latent variables of the Big Five: extroversion (E), conscientiousness (C), openness to experience (O), agreeableness (A) and neuroticism (N). In Table 4.2, we present the details of these instruments that make up a set of 209 questions organized into 21 subscales.

According to this understanding, the organization of the 21 scales in the Big Five model can simplify the number of scales and, at the same time, maintain the representative nature of the relevant constructs, whilst employing a unifying model. In line with this proposal, the objectives of this study have been to:

a) Study the statistical structure underlying the scales and the items, checking how many latent dimensions would be needed to explain the correlations between the scales. In other words - study how many groupings would be needed to summarize the items and the scales.

b) Check whether it would be possible to use the Big Five model in order to understand this dimension, and;

c) To study the basic psychometric properties of the scales for the target sample in the main study.

The **statistical analysis** of items allows us to empirically infer how many dimensions are necessary to explain the grouping of similar items (with the understanding being that similarity between items is a synonym for the high-correlation between them). If one group of items is the effect of the same latent cause, then the scores on the items are correlated. As such, by means of a statistical analysis that groups together the correlated items in factors, the number of existing dimensions that explain the correlations can be inferred.

TABLE 3 - DETAILS OF THE GROUP BASE (NUMBER OF ITEMS AND SUB-SCALES)

Test	Items	Number of Subscales	Subscales/Latent Dimensions
Big Five Inventory (BFI)	44	5	Extroversion, , Conscientiousness, Openness to experience, Agreeableness, Neuroticism
Big Five for Children (BFC)	65	5	Extroversion, , Conscientiousness, Openness to experience, Agreeableness, Neuroticism
Strengths and Difficulties Questionnaire (SDQ)	25	4	Emotional Symptoms, Behavioral problems, Hyperactivity, Peer problems, Pro-social
Norwick-Strickland Locus of Control Scale	21	1	Locus
Self-Efficacy Questionnaire for Children	24	3	Academic, Social, Emotional
Grit Scale	8	1	Grit
Core Self Evaluations	12	1	Self Esteem
Rosenberg Self-Esteem Scale	10	1	Self Esteem
Total	**209**	**21**	**16**

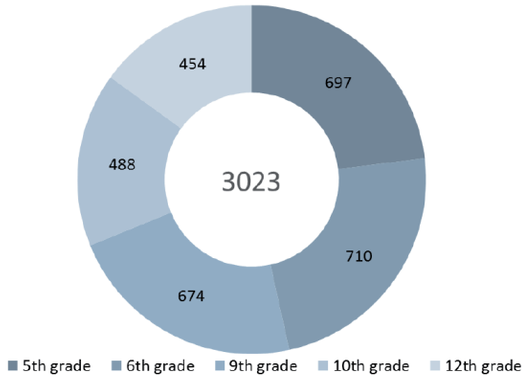
In order to conduct a statistical analysis of these 209 items, it is necessary to calculate the "inter-tem" correlation matrix, or in other words, the correlation of each item with the other 208 items. To do so, it would be necessary that, ideally, all the students answer all the items. However, in practice, this is not possible. In order to allow such information to be obtained without the students having to answer all 209 items, the 'Balanced Incomplete Block' designs were employed (BIB, Sailer, 2005).

Methodology. *Applying this design, notebooks with two tests (20 notebooks for the 5th and 6th grades) and three tests (seven notebooks for the 9th, 10th and 12th grades) were created, combining tests systematically and in a balanced manner to guarantee that all the combinations of these items would be present in at least one notebook and that the tests would be equally distributed to the students. The products of these combinations were notebooks with an average of 67 items. At the time the data was collected in the classrooms, these notebooks were distributed in ascending order, always returning to the 1st notebook after the twentieth student (in the 5th and 6th years) or the seventh student (in the 9th, 10th and 12th grades), thus producing equivalent random samples of respondents for each notebook. By doing so, random samples were produced for each pair of items allowing the calculation of the complete correlation matrix between the items, even though different segments of students were answering each item. Considering the statistical power of 0.80 of detection of correlations of a magnitude*

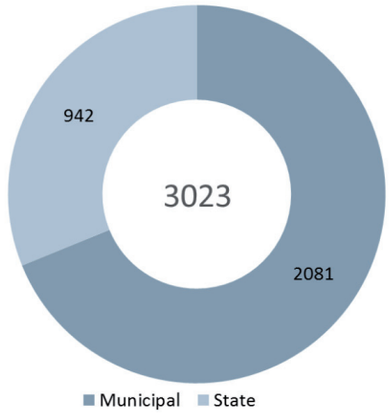
of 0.30 (that is, an 80% chance of detecting a correlation of a magnitude of 0.30 or more being statistically significant when, in fact, it is not the product of chance), it was established that samples of 85 students for each pair of items would be necessary (Cohen, 1992).

The sample was made up of 3,023 students from 16 schools with average performance in standard tests, from which 86 classes took part. The number of respondent students per test notebook varied from 93 to 252. The graphs below (**Graphs 1 to 3**) **show** the distribution of students per test, series and education system:

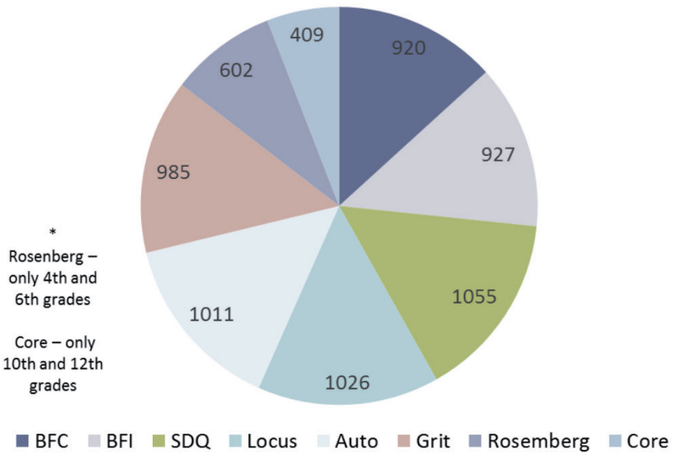
GRAPH 1A: NUMBER OF STUDENTS PER GRADE PARTICIPATING IN THE PRE-STUDY



GRAPH 1B: NUMBER OF STUDENTS PER EDUCATION SYSTEM PARTICIPATING IN THE PRE-STUDY



GRAPH 2: NUMBER OF RESPONDING STUDENTS PER TEST PARTICIPATING IN THE PRE-STUDY



Two general statistical analyses were performed:

1. Exploratory Factor Analysis of the Scales (EFA): this analysis investigated the underlying dimension of the tests used, verifying whether the Big Five dimensions framework would be able to explain the underlying structure among the subscales.

Methodology. *An exploratory factorial analysis was conducted by the minimum residual estimation method (or an 'Unweighted Least Squares – ULS - method performed by the Psych R package, Revelle, 2010). A parallel analysis was also performed to investigate the number of factors with eigenvalues higher than random chance. The analysis was performed twice - once with all the samples and the other with 5th and 6th grades. The results of these two analyses were compared.*

2. Exploratory Factor Analysis of the Items: this analysis also verified the structure but this time at the level of the items. The basic question was "Which items were the best ones to measure the intended constructs?"

Methodology. *An EFA was performed under the Unweighted Least Squares (ULS) method and geomin rotation, assuming that the responses to items were (non-continuous) categorical variables, a method that circumvents the problem of non-normality of the distribution of the items (this method is similar to an analysis of a polychoric inter-correlation matrix). This was performed using the MPLUS statistical analysis program package (Muthen & Muthen, 2012). Ten factors were taken from two rounds and compared as in the previous analysis (one with the complete sample and the other with only 5th and 6th grade students). In these analyses, we tried to discover whether latent factors similar to those found in the previous scale-level analyses could be found.*

Table 4 shows the results of the factor analyses of the scales. The parallel analysis suggested that, at most, six factors could be considered to be different random fluctuations. An analysis of the solutions of five or six factors was performed, all being conceptually similar. As such, the most parsimonious solution from the five factors was chosen.

TABLE 4 - RESULTS OF THE EXPLORATORY STATISTICAL ANALYSIS OF THE SCALES

Test-Subscale	F1	F2	F3	F4	F5
Self-Efficacy: Academic SC	**0.83**				
GRIT	**0.80**				
BFI: Conscientiousness	**0.76**				
BFC: Conscientiousness	**0.63**				0.21
BFC: Openness/Intellect	**0.53**	**0.39**			
SDQ: Hyperactivity	**-0.47**	0.24	0.27		
BFC: Extraversion		**0.74**			
BFI: Extraversion		**0.72**			**-0.35**
Self-Efficacy: Social SC		**0.65**	-0.20		
BFI: Openness/Intellect		**0.42**			
BFI: Emotional Stability/Neuroticism			**-0.93**		
BFC: Neuroticism		0.22	**0.68**		
Self-Efficacy: Emotional SC		**0.40**	**-0.44**		
SDQ: Conduct Problems		0.23	**0.33**	**-0.33**	
BFI: Agreeableness				**0.80**	
BFC: Agreeableness				**0.71**	
SDQ: Pro-social behavior				**0.70**	
Norwick-Strickland Locus of Control Scale:					**0.53**
SDQ: Peer Problems				-0.26	**0.52**
Rosenberg/Core Self Ev.: Self esteem	0.27	0.21			**-0.50**
SDQ: Emotional Symptoms			**0.35**		**0.43**
SS loadings	3.31	2.36	2.23	2.21	1.44
Proportion Var	0.16	0.11	0.11	0.11	0.07
Cumulative Var	0.16	0.27	0.38	0.48	0.55
Factor Correlations					
F1					
F2	0.18				
F3	-0.24	0.09			
F4	**0.48**	**0.30**	-0.20		
F5	-0.12	-0.13	**0.40**	-0.05	

In **summary**, it can be observed that:

1. The dimensions can be clearly interpreted by the Big Five model: F1 (Conscientiousness) self-motivation/determination skill; F2 (Extraversion) ability to start social interactions; F3 (Emotional Stability/Neuroticism/Internalization) the ability to deal with negative emotions, especially anxiety and depression; F4 (Agreeableness) social abilities to maintain positive interactions, and; F5 (Locus of Control or Emotional Stability/Neuroticism/Externalization) associated with more positive attitudes of seeing oneself as responsible and exercising control over what occurs in your life. This factor can also suggest the ability to deal with negative emotions, especially anger. The opposite pole of this factor – the external locus – can be associated with low levels of kindness and with antagonism.

2. The tests traditionally used in educational contexts (self-concept, self-efficacy, self-esteem, motivation, and attitudes and beliefs of control over events and adaptability), which are often considered to be measuring different constructs from traditional personality traits, can be understood in a broad context of the Big Five, having loadings consistent with the meaning of the global factors. Some examples: social self-concept with E+[40] academic self-concept as C+[5] (self-conscientiousness) and emotional self-concept and pro-social behavior as A+ (high agreeableness).

3. The correlations between the five factors are moderate and reflect a possible solution at a higher level, bringing together Conscientiousness, Extraversion and Agreeableness as well as the two factors tied to Neuroticism.

4. The Openness Factor was the only one that did not appear in isolation. The subscales were sometimes grouped with conscientiousness and sometimes with extraversion. This could have occurred as a result of the non-existence of this domain in the other scales, resulting in the little covariance associated with this domain in item indicators as compared with the other big five domains that are more wider represented in the indicators.

2.3 Large-scale social and emotional skills measurement Project

At the first meeting of specialists in instruments for the evaluation of skills and abilities for the *Education and Social Progress - Longitudinal Study of Skill Dynamics*[41], project, the proposal that the Big Five model could be used as a unifying model for the orientation/selection/construction of instruments for evaluation of socio-emotional or non-cognitive constructs in large-scale studies was supported. Also approved was the suggestion that certain parts of the existing instruments could be selected in order to create a new instrument that meets the set of criteria discussed in Chapter 2 of this report.

The empirical study of the latent structure of the selected tests (a statistical analysis of the scales) empirically reinforces the recommendation for the use of the Big Five as a nuclear organizing model for the constructs studied under the classification "non-cognitive" and "socio-emotional" in the educational context.

40. The factor loading is equivalent to the correlation of answer to the scales with the latent variable. In this text, E+ means a positive pole (high scores) in factor E that is, extraversion. E- would mean scores from the negative pole (low scores) of extraversion – that is, introversion. Therefore, high scores in social self-concept are associated with high scores in extraversion.

41. The OCDE meeting took place in Washington, United States, in June, 2013.

As such, the results of this study represent an important finding since they provide a scientific, empirical base for that which was recommended for the Project and justify the idea that it is possible to arrive at a common language to approach to these constructs, allowing communication and research on the issue. Furthermore: the organization of the socio-emotional constructs around the Big Five connects the 21st century skills assessment, a relatively new area, to a rich knowledge base developed in the sphere of studies on areas of personality and other correlated variables – such as interests and cognitive skills – contributing to the understanding of classic constructs in the psychology of education, such as self-esteem, locus of control and self-efficacy.

It is possible to understand these classic constructs (such as **self-esteem**, **locus of control** and **self-efficacy**) as descriptors of more specific behaviors a process connected to the educational context, but from a broad perspective their basis lies in the personality traits such as dispositional characteristics that, upon interacting with the context, create different standards in the ways they express themselves (in terms of self-esteem and self-efficacy, for example). As such, a readiness to experience negative emotions (neuroticism) may express itself as negative thoughts about oneself (low self-esteem) or desperation and attribution of one's victories attributed to uncontrollable external events (external locus of control).

2.3.1 Composition of an instrument optimized for the evaluation of social and emotional skills in schools on a large-scale

In line with the suggestion of creating a single optimized instrument based upon the selection of items found in instruments that already exist, the employment of the exploratory statistical analysis of the items was then pursued, in order to identify the grouping of items that were consistent with the factors uncovered in the previous analysis, and select items that could be included in a **new instrument, with optimized properties in terms of internal consistency (fidelity) of the scales and representation of the relevant domains (bandwidth).** In addition, we aimed to make sure that this instrument satisfied the condition of extension proposed in the time-length study (60 to 90 items).

Within this task, 10 factors in two analyses were examined, one with the complete sample and another with just students from 5th and 6th grades. The results were similar and contained the broad factors identified in the exploratory statistical analysis of the scales. This analysis, at the level of the items, allowed a more detailed understanding of the nuances of the items/facets of the different instruments that were grouped together. At the same time, it was an optimized solution in terms of the internal consistency of the factors, since the items were grouped due to their correlations with all the other items and not only with their scale of origin, allowing them to possibly be grouped with items from other scales. The ten factors found were: **F1: neuroticism; F2: conscientiousness; F3: openness (items from the BFC mostly representing the intellectual skills facet); F4: agreeableness; F5: extraversion; F6: openness to experiences (items from the BFI mostly representing the creativity and artistic interests facet); F7: introversion; F8: locus of control and behavior problems (neuroticism with an emphasis on externalization); F9: self-esteem with a focus on sad feelings, and; F10: self-esteem with a focus on negative self-image.**

Based upon this analysis, four researchers from the IAS team independently reviewed item-factor loadings that contained the items ordered by their loadings and their descriptive statistics and made their selection of items[42] based upon six constraints:

(a) The items should represent the five factors, and an additional sixth factor of externalizing neuroticism (locus of control);
(b) The items should have high loadings and variance;
(c) The items should represent the positive and negative poles of the Big Five dimensions;
(d) The scales should combine items from different tests whenever possible;
(e) The maximum number of items in the final instruments should be kept within the range of 60 to 90; and
(f) The items should have passed the qualitative pilot analysis and the focus group study (without presenting any problems).

The selections were contrasted and reviewed by the group and a final consensus version of 92 items was decided upon (for 10th to 12th grades). A subset of 62 items from the original 92 was selected for the 5th grade[43].

Based upon the complete inter-item correlation matrix that it was possible to calculate using the Balanced Incomplete Blocks (BIB) design, internal consistency coefficients were calculated for the new instrument subscales that are presented in **Table 5**. As can be seen, the internal consistency coefficients of **the new scales were higher than 0.70, reaching a benchmark higher than is usually found in scales with this age group.**

TABLE 5 - INTERNAL CONSISTENCY OF THE SENNA SCALES.

Scale	Alfa 92 itens	# of Items	Alfa 62 Items	# of Items
F1 Conscientiousness (C)	.88	17	.83	10
F2 Extraversion (E)	.79	14	.74	11
F3 Emotional Stability (N1)	.82	14	.72	9
F4 Locus of Control (N2)	.77	15	.72	12
F5 Agreeableness (A)	.76	15	.70	12
F6 Openness (O)	.80	17	.73	8

42. In the case of items for which authorization for use was not received in time, these serve simply as parameters for the creation of new items that measure the same constructs.
43. The instruments developed can be consulted in the Annexes section of this report.

2.3.2 Large-scale evaluation study of social and emotional skills in the State of Rio de Janeiro

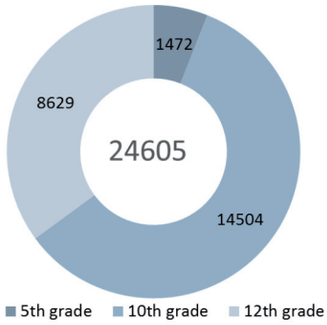
Once the instrument developed in previous phases was available, a large-scale evaluative pilot study of the social and emotional skills in Rio de Janeiro state system was performed.

The objectives of this study were:
a) to evaluate in more detail the psychometric properties of the SENNA instrument created in the previous study;
b) to assess the social-emotional skills of the students and describe their distribution with respect to important variables related to the school system, such as school year, sex, socio-economic level, types of school organization;
c) to investigate the associations of social and emotional skills with academic achievement and standardized performance tests.

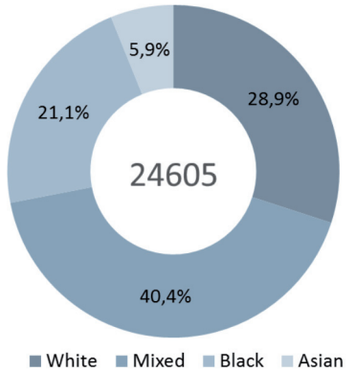
The participants in this study were a representative sample of students from the state of Rio de Janeiro. The total number of students was N= 24,605[44]. They came from 14 regions, 79 cities, 431 schools and 1,062 classes. The number of students per class varied between 1 and 50, with the average being 'A'=23 and the standard deviation being 'SD'=7.6. The graphs below (**Graphs 3a, 3b and 3c**) describes the distribution of the respondents according to sex, school year and race. The distribution by age can also be seen in **Table 6.**

44. Approximately 9,475 students did not attend school on the day of the test. This figure includes the average number of absences recorded on normal school days added to the absences due to the teachers' strike that took place during the period in question.

GRAPH 3A: NUMBER OF STUDENTS PER GRADE PARTICIPATING IN THE LARGE-SCALE STUDY



GRAPH 3B: ETHNIC ORIGIN OF THE STUDENTS PARTICIPATING IN THE LARGE-SCALE STUDY



GRAPH 3C: SEX OF THE STUDENTS PARTICIPATING IN THE LARGE-SCALE STUDY

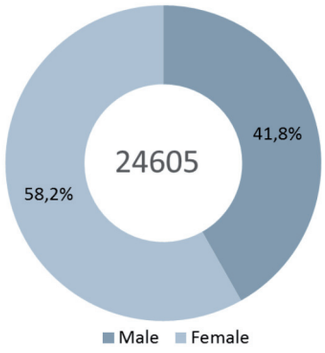


TABLE 6: AVERAGE AGE AND STANDARD DEVIATION PER SCHOOL YEAR

Average per age	Average Age	Standard deviation
5th grade	11.9	1.46
10th grade	16.45	1.05
12th grade	18.21	1.02

2.3.3 Psychometric Properties of the Instrument

Three analyses were performed to verify the Instrument's psychometric properties:

a. Firstly, we performed a item-level confirmatory factor analysis, verifying whether it was possible to recover the structure of the six factors as defined in the previous study;
b. Secondly, we performed a multiple group confirmatory factor analysis in the samples from the 5th, 10th and 12th grades, testing the invariance of the measurement model parameters on these three groups. The basic issue was to understand whether the Instrument created measures the same factors, in the same way, in different school grades;
c. Finally, Item Response Theory[45] was applied to calibrate item parameters and equate student's latent trait scores in the two forms of the test.

a) Confirmatory Statistical Analysis of the Items

Methodology. *The first part consisted of an item-factor analysis based upon 'MPLUS' (Muthen & Muthen, 2012) employing Exploratory Structural Equation Modeling (ESEM) with the WLSMV estimator (weighted least square parameter estimates). This was intended to test whether we can recover the factor structure of six latent factors with the same indicators. An ESEM approach is an intermediate way between the fully exploratory approaches to the fully independent cluster confirmatory approach and imposes the only restriction on the number of factors leaving their loadings free to be estimated in all of the factors extracted. This method has proved to be more suitable particularly in item-level factor analysis, as is being done here (Marsh et all 2010). A parallel analysis was performed to investigate the number of factors using simulation of matrices of random polychoric correlations[46] with the Psych R package (Revelle, 2010). This analysis suggested up to 13 non-random factors. Item indicators were assumed to be categorical variables that accommodate non-normality in their distributions. The extraction of six correlated factors using the geomin oblique rotation was specified.*

Table 7 presents the results from the factorial 'ESEM' analysis. In the upper part of the table (line four) classical internal consistency is presented. As can be seen this was very high (above .75). In the middle part of the table the fit adjusted indexes from the ESEM are presented along with a traditional independent cluster confirmatory factor analysis (CFA) for comparison purposes. As can be seen, the six-factor ESEM model shows acceptable levels of model adjusted indexesfit considering the often used benchmarks (CFI and TLI >0.90, RMSEA < .05).

In the lower half of **Table 7**, a cross-classification is shown of the expected results with the ESEM results. In the six columns, the classification of the factors can be seen, in which it was expected that the items would have a high factorial loading. The numbers presented in the body of the table show the number of items classified in each expected versus obtained combination in the ESEM. In the ESEM analysis, it was understood that an item would be classified in any of the six factor if it had a factor loading greater or equal to 0.29. As can

45. Item Response Theory (IRT) is a statistics model that represents the testing situation seeking to predict what a person's reply to a question will be, considering the quantitive parameters relating to the person's level in the latent trait measured by the question and, at the same time, the properties of the questions, such as, for example, what level of the trait the question measures most suitably.
46. Polychoric correlations were calculated because they are more adequate for this type of categorical variables where the observed distributions usually deviates from normality but because these variables have downgraded a continuous normally distributed variable – assumed to exist at the latent unobserved level - to a categorical variable - at observed level.

be seen, the item-factor organization is particularly well recovered in the ESEM analysis. Some crossed loadings (items that have high loadings in more than one factor), particularly between extraversion and agreeableness, as well as between conscientiousness - the negative pole of which is impulsiveness - and locus of control - neuroticism/externalization- can be seen. These results were expected considering the correlations between the factors found in the previous study.

In summary, the results show that that the underlying structure of the 92 items of the Instrument in six factors was clearly recovered in a new sample considerably bigger that the previous one, under much different testing conditions (since the BIB technique was not used) and independent of the sample from the first study (when the instrument was first developed). As such, this study is an example of a very successful replication of the factor structure of the items in this Instrument.

TABLE 7 - RESULTS OF ITEM FACTOR ANALYSIS OF THE SENNA INSTRUMENT

Internal consistency results

Factor	f1	f2	f3	f4	f5	f6
Name	Cons	Extr	Nint	Nloc	Agre	Opns
Alfa	0.91	0.83	0.89	0.75	0.81	0.78
Nr. of items	17	14	14	15	15	17

Adjustment Indexes

	Chi-Square	df	n.FreeP	CFI	TLI	RMSEA
CFA	418,777	4079	475	0.700	0.692	0.064
ESEM	120,588	3649	905	0.915	0.903	0.036

Distribution of loadings greather than 0.29

Expected factors	f1	f2	f3	f4	f5	f6	
	Cons.	Extr.	Nint.	Nloc.	Agre;	Opns.	
Fatores ESEM	**0.94**	**0.64**	**0.93**	**0.73**	**0.67**	**0.71**	
f1	16	0	0	0	1	0	Cons
f2	0	0	13	0	3	0	Nint
f3	5	1	0	11	1	3	Nloc / Cons (impulsiv.)
f4	0	7	1	1	10	5	Agre. / Extr.
f5	0	0	0	1	1	12	Opns.
f6	0	9	1	2	0	0	Extr.

b) Invariance test of the model for subgroups from the sample

The second question under investigation was whether we could assume measurement invariance of the instrument for students of different ages/grades. In other words, **is it possible to know for certain that the instrument will measure the same factors in the same way with different school years?** For this analysis we used the short '62-item' version, since this included items answered in common across all the grades.

Methodology. *The analysis of invariance was performed on the MPLUS using the Exploratory Structural Equation MNodeling (ESEM) form Multiple Groups. This analysis tests three increasingly more restrictive models: (a) the configural invariance model that tested whether the basic six-factor structure model is suitable to represent the underlying structure of the 62-item covariance matrix when calculated separately, based upon samples from different school years. In this model, all the parameters (item factor loadings and item intercepts – item indicator averages) are freely estimated. Only the general structure (number of factors) is held constant between the groups; (b) the metric invariance model that tested whether the factor loadings of the items, in the factors, varied across grades or not when calculated separately. The additional restrictions imposed here are that factori loadings of the items in the factors are constant across the subgroups defined by the school years. This is equivalent to assuming that item doesn't show Differential Item Functioning (DIF) in the discrimination parameter "a" of the Item Response Theory (IRT); (c) scalar invariance model that tested whether the intercepts (averages in the items) were equal when calculated separately across grades (controlling for possible differences in the latent factor among groups). The additional restriction imposed on this model is that, in addition to assuming the six factors and equal item loadings, the averages of the items are kept constant across the three groups, but the differences that could possibly exist due to the different averages in the latent constructs are controlled. This model tests whether students with the same average latent factor, in different years, showed different averages in the items. As a result, this is equivalent to the analysis of the 'Differential Item Functioning' under the difficulty parameter "b" (in this case, the ease of agreeing with the assertions) in Item Response Theory ('IRT').*

TABLE 8 - ADJUSTMENT INDEXES OF THE EXPLORATORY MODELING ANALYSIS WITH STRUCTURAL EQUATIONS IN MULTIPLE GROUPS

Models	Chi-Square	df	n.FreeP	CFI	TLI	RMSEA	SRMR
Configural	41,413	4602	1443	0.901	0.878	0.031	0.022
Metric	43,884	5274	771	0.896	0.889	0.030	0.026
Scale	46,219	5386	659	0.890	0.885	0.030	0.026

The results in **Table 8** show that the restrictions imposed by the more demanding models containing the invariance assumption, do not greatly reduce the adjustment achieved by the base model. The RMSEA and SRMR indexes remain practically unaltered. The CFI and TLI indexes drop a little in the metric model and stay the same with the additional restrictions of the scalar model. This suggests that there might be some changes in item loadings for some items across the groups that need to be further explored. But the most important aspect is that there are no huge differences in item intercepts (scalar model). **This therefore suggests that a partially invariant model could be assumed for the Instrument.** The question of DIF will be further explored below when item parameters were calibrated via Item Response Theory.

c) Gauging of items in line with the Item Response Theory

The third and final stage of analysis was the calibration of item parameters and estimates of the latent trait scores for students using the Item Response Theory (IRT) approach for polytomous items with the use of the Rasch-Andrich Rating Scale Model (RSM, Wright, & Masters, 1982). At this stage the objective was to investigate the psychometric properties of the scales; analyze *likert* rating scale category structure used in the items; produce test information functions to better analyze the reliability of the scales; fully equate the scores of the 92-item test with the short-form 62-item test (using the common item equating method); analyze item maps (construct maps) to improve the understanding of the underlying construct of the items and provide additional information on scale interpretations (item-referenced meaning (Embretson & Reise, 2000) and, finally; test the Differential Item Functioning (DIF) with respect to grades. The analysis was repeated six times, one for each factor, as each one can be assumed to be one-dimensional as the previous analysis has evidenced.

Methodology: *The rating scale model estimates item threshold parameters for transition from points 1 to 2, 2 to 3, 3 to 4 and 4 to 5 (depending upon the difficulty of the item). The item and student parameters (in long and short forms) were calibrated simultaneously. The 62-item test was treated as different form from full-length test but with all the common items in the full version. The metric of the scale was identified by fixing the average of the theta sample as M=0 and the standard deviation as SD=1. The calibrations were done using WINSTEPS software (Linacre, 2013).*

The results of these analyses showed a good fit to the model. None of the items approached the cutoff point that would recommend their elimination from the scale. The *likert* rating scale structure that assesses whether each one of the five points was providing information about different levels on the latent scale was very good for five of the six factors. Factor 4 (Neuroticism/Locus of Control) showed a structure that could be reduced to three points instead of five without losing any psychometric information. This could be due to the fact that on the original locus of control scale, the reduction of the items was dichotomously formulated for 'yes' or 'no' answers, being scored with 0 or 1.

The Differential Item Functioning (DIF) analysis showed that only 11 of the 92 items showed DIF at a level that could suggest a potential problem with the item. These items concentrated on Factors F4 (Neuroticism/Locus of Control), F5 (agreeableness) and F6 (openness). But most importantly, there were approximately balanced DIF levels favoring, on the one hand, the 5th grade (examples: "I usually do as I am told", "I am able to create new games and ways of playing"), and, on the other, favoring 10th and 12th grades (examples: "Starts fights with others", "Do you feel that most of the time parents listen to what their children have to say?") within the same scale which usually produces a situation whereby the DIF cancels itself out in the total score of the test.

In summary, together with the previous analysis, these results favor the idea that the measurement invariance of the instrument is psychometrically equivalent across different school grades.

2.4 Summary: construction and validation of the instrument designed to measure social and emotional skills

To meet the aim of the Project tof measuring social and emotional skills, we sought to construct an instrument that was simple and robust enough to be applied on a large scale, in such a way that it would provide an extensive picture of the distribution of social and emotional characteristics in Brazil. At the same time, it needed to be precise and easy to interpret in order for it to be used scientifically in studies focused on documenting the social and emotional development of individuals throughout the cycle of life and investigate how interventions affect this path.

During the first phase of the construction of this tool, a team of researchers evaluated a wide range of internationally recognized psychology tools that were considered as being candidates for satisfying the objectives of this Project. The set of instruments selected contained measurements of personality traits, self-conception, self-efficacy, self-esteem, motivation, attitudes and beliefs in control of events (internal and external attribution) and issues involving social and emotional adaptation.

Each one the selected instruments was analyzed from the perspective of four different dimensions considered relevant to their use and, in the end, eight of them were chosen to be tested in a school environment. Due to suspicions that certain important instruments could have been left off the list produced during the first stage, it was decided that well-known specialists in the area of Personality Psychology should be consulted in order to increase the list of selected instruments. Following consultation with the specialists, a set of instruments was established, each of which were applicable on a large-scale and could bring together all of the constructs that were important for this Project.

The conclusions demonstrated that all instruments presented robust properties, but none of them individually covered the wide spectrum of characteristics that needed to be measured. The alternative was to construct a new instrument that grouped together the selected instruments. Based upon this list, a series of initial studies was conducted with the aim of translating and analyzing the instruments, performing a qualitative analysis of the students' understanding of the questions, checking the time required to perform the tests and performing a psychometric analysis and selection of questions for the final instrument.

In line with the suggestion of creating a unique instrument with a design based upon the selection of items from already existing instruments, the authors proposed an understanding of the measured approached in the instruments selected based upon the five latent variables of the 'Big Five': extraversion, conscientiousness, openness to new experiences, agreeableness and emotional stability. Based upon this understanding and backed up by international writings on the subject[47], we sought to identify ways of grouping the items that were consistent with these factors and select questions that could be included in the new instrument, with properties that were optimized in terms of consistency in the scales and representation of the important skills.

Once in possession of the instrument developed in the preliminary stages, a large-scale, pilot study was conducted to evaluate the socio-emotional skills present in Rio de Janeiro state education system. The objectives of the pilot study were: to evaluate the psychometric properties of the Instrument created in the previous stage in more detail evaluate the socio-emotional skills of the students and describe their distribution as a result of important variables related to the education system, including school year, sex, socio-economic level and

47. Ver Kyllonen, Lipnevich, Burrus & Roberts (2008).

forms of school structure, in order to investigate the association of the socio-emotional skills and academic performance in standard performance tests.

In Chapter 2, only the conclusions related to the first objective were touched upon. In summary it was empirically identified that the dimensions considered by the instruments selected could be clearly interpreted by the Big Five model, corroborating the theory that the tests traditionally used in educational contexts (self-conception, self-efficacy, self-esteem, motivation, and attitudes and beliefs relating to the control of events and adaptability) can be understood in a broad context of the Big Five factors. This result, as has been suggested, represents an important finding since it empirically confirms the idea that it is possible to arrive at a common language for the treatment of social and emotional characteristics in education, thus facilitating communication and research on the subject. **In relation to the results of the psychometric analyses, these suggest that the instrument has a high degree of internal consistency and is psychometrically equivalent for the different school years sampled.**

In the next chapter, the preliminary results of the analysis of the data collected in the pilot project will be presented, with the intention of better investigating the relationship between the skills evaluated and educational performance, as well as study the ways in which aspects of school, the student and the family influence social and emotional development.

Chapter 3:

Empirical analysis of the association between social and emotional traits and individual and family characteristics

As explained in earlier chapters, the aim of this initiative is to create an instrument capable of measuring the most important social and emotional student's characteristics- those that affect their learning performance and their adult well-being – and to use this instrument to diagnose, monitor, evaluate and design better public policies.

The objective of this chapter is to conduct a preliminary investigation into whether the constructs measured by the instrument developed in the Project are amongst the potentials that determine learning performance, and discuss how characteristics such as those relating to the individual and family can alter students' social and emotional characteristics.

In the next section, we would like to present the database used for this chapter. In section 9, we will be looking at whether social and emotional characteristics affect the learning performance of students in Rio de Janeiro in areas such as Mathematics and Portuguese. The next section investigates whether the characteristics of the student, his or her family, and their classmates determine the levels of their social and emotional characteristics. The last section is dedicated to the conclusions of this stage of the research.

3.1 Presentation of the database

The main source of information used in this study is the micro-database drawn from the field studies performed in Rio de Janeiro state school system in order to validate the instrument developed as one of the objectives of the Project.

The data was collected at the beginning of October 2013, from 24,605 students

in the 5th grade (6% of the sample), and the 10th and 12th grades (59% and 35% of the sample, respectively). In addition of including the results of the instrument, which gauged six social and emotional constructs, the base also included answers from the students to questions relating to their individual characteristics (sex, age and race), the family environment in which they live, and their parents' behavior. The results of a test in Portuguese and Mathematics completed by the same students in the second semester of 2013 (SAERJINHO[48]), as part of Rio de Janeiro state education evaluation system, was also included in the set of available information.

Table 9 provides a summary of the variables used:

3.2 Social and emotional traits and academic performance

TABLE 9: LIST OF VARIABLES USED TO THE EMPIRICAL ANALYSIS OF THE ASSOCIATION BETWEEN SOCIAL AND EMOTIONAL TRAITS AND INDIVIDUAL AND FAMILY CHARACTERISTICS

Variable	Source/ Description
Social and emotional traits	**Micro-data from the pilot evaluation in Rio de Janeiro**
Conscientiousness Extraversion Emotional Stability Locus of Control Agreeableness Openness to New Experiences	Factorial Analysis/TRI[49] based upon the social and emotional questionnaire (62 items for the 5th grade and 92 items for the 10th and 12th grades)
Family environment	**Micro-data from the pilot evaluation in Rio de Janeiro**
Mother's education	5 educational levels
Father's literacy	Dummy[50]
Socioeconomic Status (SES)	Linear index combining 12 characteristics of dwellings and ownership of durable goods
Number of Books	4 categories (none to more than 200)
Does the mother live at home?	Dummy
Economically disadvantaged	Dummy (first fifth of SES)
Individual characteristics	**Micro-data from the pilot evaluation in Rio de Janeiro**
Sex	Dummy
Age	Integer
Race	Dummy (1 = white, Asian)
Attitudes of parents and children	**Micro-data from the pilot evaluation in Rio de Janeiro**
Reading frequency	4 categories (never to frequently)
Parents' reading frequency	4 categories (never to frequently)
Do parents encourage the children to read	Dummy
Class characteristics	**Micro-data from the pilot evaluation in Rio de Janeiro**
Size	Number of enrolled students
Averages of social and emotional traits, individual characteristics and family environment, as well as family attitudes	
Learning performance	**SAERJINHO**
Grade in Portuguese	Standardized Grade (m = 0, v = 1)
Grade in Mathematics	Standardized Grade (m = 0, v = 1)

48. 'Saerjinho' is a bimonthly evaluation system focused on the teaching and learning process in the Rio de Janeiro state education system. At the end of every two-month period, students in the regular education state schools' 5th and 9th grades and the three grades of High School take tests in a number of different subjects. For this stage of the study, only those associations with performance in Portuguese and Mathematics were investigated.

49. Item Response Theory (IRT) is a statistics model that represents the testing situation seeking to predict what a person's reply to a question will be, considering the quantitive parameters relating to the person's level in the latent trait measured by the question and, at the same time, the properties of the questions, such as, for example, what level of the trait the question measures most suitably.

50. A 'dummy' or 'indicative' variable takes on the value of '1' if an event tales place, and '0' if it does not. It is especially used for qualitative variables (e.g. sex) in a quantitative exercise. In the case of the 'economically disadvantaged' variable, for example, the variable is 1 if the individual belongs to the 20% with a lower socio-economic status index (SES), and 0 if not.

3.2.1 Description of the social and emotional profile of the students according to the database

Before starting a more in-depth investigation of the relationship between students' social and emotional development and academic performance, it would be useful to examine some basic statistics concerning the profile of these students in relation to the social and emotional traits evaluated.

Firstly, it is interesting to note that our instrument reproduces a number of important factors established in international writings in relation to the behavior of the constructs measured in this age group. **Graph A** in the Annexes, shows the averages of each one of the social and emotional factors by sex (Conscientiousness, Extraversion, Emotional Stability, Locus of Control, Agreeableness and Openness to New Experiences).

Girls tend to be more conscientious, extravert and agreeable but, on the other hand, they have less Emotional Stability. These results are in line with the meta-analysis performed by Soto et al (2011), which was based upon results obtained from different parts of the world by instruments that measure the Big Five personality traits. The differences are generally relatively small, except in the case of Emotional Stability.

Another Project's results which corroborated with the conclusions of this meta-analysis is the non-linear pattern of the averages of Conscientiousness by age, which initially decrease and then go up amongst individuals between 10 and 20 years old (Graph A2). The confirmation of these patterns reinforces the idea that the instrument in fact measures constructs that are similar and comparable to those used in studies of the Big Five.

In terms of the relationship of the social and emotional gauges with different variables from the family context, it is interesting to note that the averages of all the six constructs measured vary very little with the mother's educational level, typically interpreted as an time-invariant measurement of the quality of the family environment. In contrast, social and emotional characteristics appear to vary a great deal with indicators of the attitudes and habits of the student and his or her parents, such as the frequency of reading and the students' awareness that they are receiving encouragement from their parents to study (Graphs A3-A5)[51].

3.2.2 The importance of personality characteristics for learning in comparison to other variables

The starting point for our analysis was to quantify the importance of the socio-emotional traits vis-à-vis other determinants of the academic performance. In particular, we would like to understand **how much of the differences in performance observed among the students is associated with the different sets of individual and family characteristics, that is, which of these characteristics most influence academic performance.**

Following the division presented in **Table 9**, we grouped the possible determining

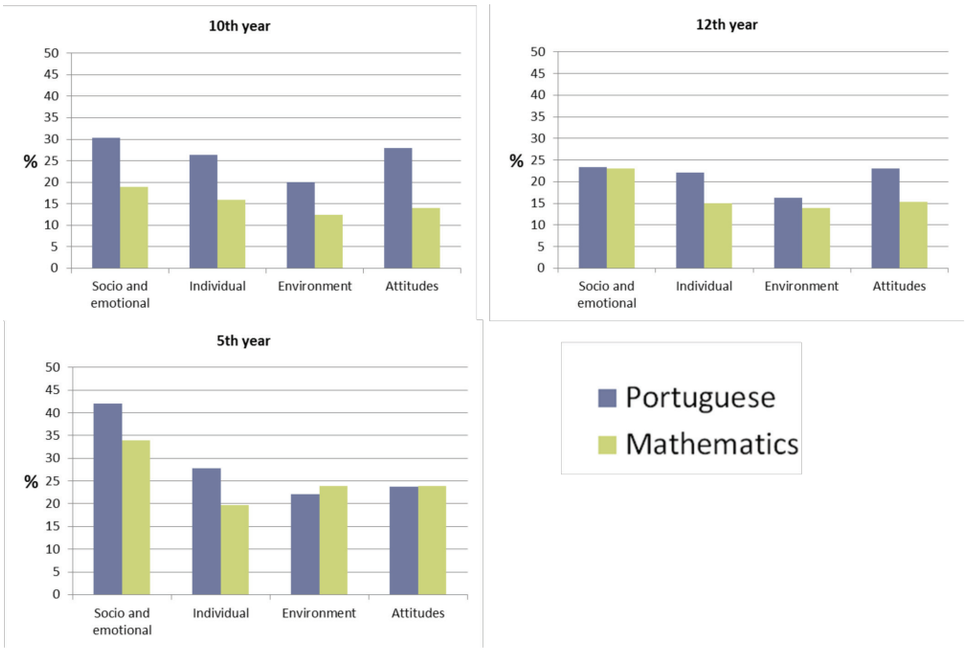
51. This is a surprising result, to the extent that some researchers believe that personality attributes are stable, durable individual traits, and should be mainly influenced by permanent household characteristics, such as the mother education. The fact that they correlate more with habits and attitudes may suggest that they are relatively malleable at the ages we are investigating in this study.

factors for school performance into four groups **(social and emotional characteristics, other individual characteristics [sex, race and age], family environment characteristics and the attitudes and habits of parents and children)** and sought to estimate how much the differences of the performances between the individuals can be associated with the differences observed in each of these sets of variables.

Methodology - *Our strategy consisted of successively projecting each of the performance measurements (grades in Portuguese and Mathematics) onto each group of determining factors, in order to isolate the part of the performance variation between the students that is systematically associated with the set of determinants in question. The projection of Math grades onto socio-emotional characteristics, for instance, represents the part of the mathematics grades that is associated with the differences that the individuals have in traits of this nature. Next, we calculated the ratio between the dispersion of the projection and the total dispersion of the grades or, in other words, the ratio between the part of the diversity in performance associated with a specific group of determinants and the total dispersion of performance. The ratio between the dispersion of the projection and the total projection is a means of quantifying the fraction of the total dispersion of grades that could be related to the differences that individuals have in the characteristics of that group.*

Graph 4 above, summarizes the results of this exercise.

GRAPH 4: PROPORTION OF PERFORMANCE IN PORTUGUESE AND MATHEMATICS ASSOCIATED WITH THE DIFFERENCES IN THE LEVELS OF INDIVIDUAL AND FAMILY CHARACTERISTICS



Amongst the individual and family characteristics, the set of determinants most associated with the observed differences in performance is that of socio-emotional traits and this relationship is even stronger than that between academic performance and the characteristics of the family environment (that include mother's education and family's socio-economic status)

In terms of differences between educational levels, socio-emotional traits seem to be significantly more associated with performance in lower grades, unlike other individual and family characteristics, amongst which the statistical relationship with performance appears to vary less in relation with respect to the grade.

3.2.3 The effects of social and emotional development on academic performance: the impact of an improvement in social and emotional characteristics on grades

In this section, we estimated the impact that a variation in the socio-emotional characteristics would have on performance in Portuguese and Mathematics, keeping family (environment and attitudes) and school characteristics constant[52]. Our measurements of the impact were based upon the "within-school" variation, that is, they associate the difference between the student performance and the respective school average, with the difference between the student socio-emotional trait and the respective school average.

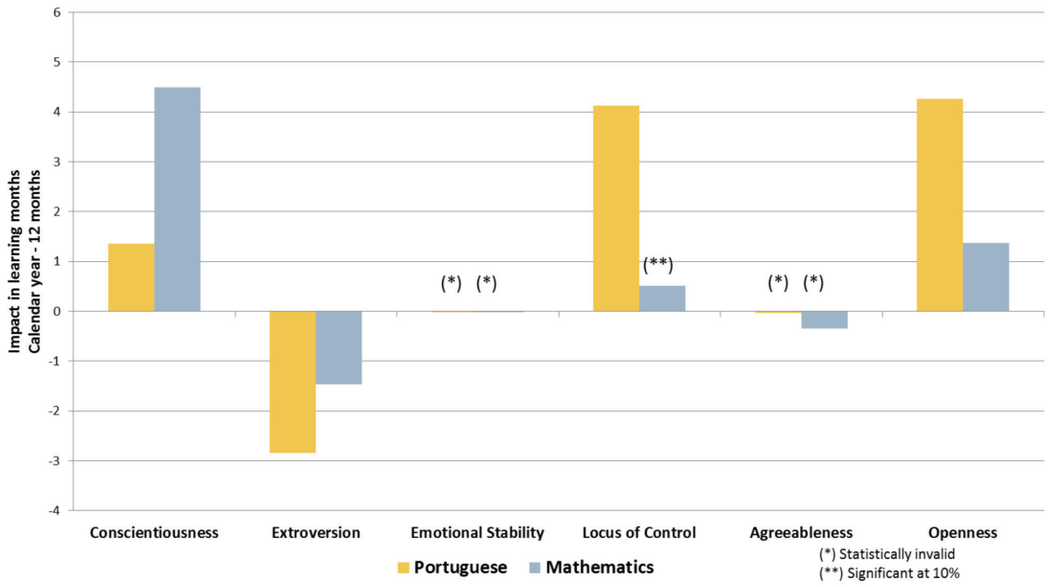
In theory, it would be possible for us to imagine that the better quality schools help their students achieve better grades, at the same time as they, in general, attract students who have undergone a more favorable socio-emotional development (be it because they perform some form of selection process for student entry, or because they have a pedagogic approach that attracts parents who are more engaged in the education of their children), causing a positive relationship between socio-emotional traits and grades that cannot be interpreted as causal. The strategy of within-school variation is used precisely to avoid this 'contamination'.

In the illustrations included in this chapter, we show the extent to which grades in Portuguese and Mathematics alter if a student in the 25th lowest percentile of a given socio-emotional skill were moved to the 75th percentile. Moreover, previous research with standardized tests in Brazil show that students increase, on average, 0.33 standard deviation in their Language and Math scores per additional year of education. We use this fact to compute the impacts in terms of 'months of learning'[53].

To illustrate this, we can see in Graph 5 that the impact of Conscientiousness on Mathematics grades is 4.5. This means that, if we raise the Conscientiousness of a student who is currently amongst those with a lower level of this skill (that is, a student in the '25' percentile) until he is found to be amongst those with a higher level of this skill (someone in the '75' percentile), this student's grades improve as though he or she had received 4.5 months of learning.

52. The statistical exercise that supports our affirmations is the estimation of a linear equation relating grades (in Portuguese and Mathematics) to individual characteristics, family environment and the attitudes and habits of parents and children (described in Table 9), including school variable indicators (dummies).
53. One month of learning in this case is equivalent to 0,33/12 = 0,025 standard deviation in standardized test scores.

GRAPH 5: IMPACT ON PERFORMANCE OF MOVING A CHILD FROM THE 25TH TO THE 75TH SOCIAL AND EMOTIONAL PERCENTILE



In the case of Portuguese grades, it is the variations in the Locus of Control and Openness to New Experiences that have a greater effect, whilst Mathematics is very much affected by alterations in Conscientiousness.

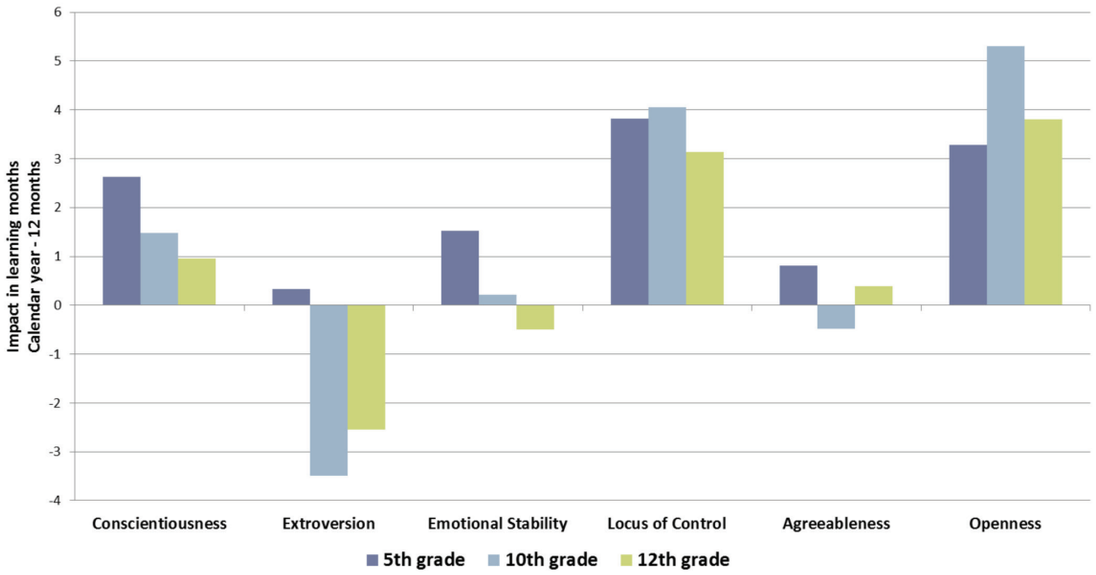
It is interesting to note that Extraversion appears to have a negative effect on grades, and this deserves further investigation. In fact (and this is supported by international writings), of all the Big Five traits, Extraversion is the one that probably presents effects that are most difficult to collect in statistics exercises, since the relationship of this trait with learning results does not appear to be entirely monotonic, or in other words, having more (or less) of this trait does not necessarily mean an advantage in performance. Finally, the levels of Agreeableness and Emotional Stability do not appear to be statistically associated with learning performance.

Making an in-depth investigation of the relationship between social and emotional traits and performance in Portuguese (Graph 8), it is possible to note that, for the children from the 5th grade, Conscientiousness also plays an important role and Extraversion does not have an adverse effect, as can be seen in the 10th and 12th grades.

It is important to mention that, in the field study, a much higher number of students enrolled in high school did not attend class on the day that the study was applied than those enrolled in primary school (a scenario that is no different to the education system's regular pattern), and this is a factor which could confuse the comparisons between estimates made in relation to distinct years. Along the same lines, the samples from the 10th and 12th grades may be considered as representative of the total number of those enrolled in these grades in the State of Rio de Janeiro, whilst the 5th grade sample students are all enrolled in the municipality of São Gonçalo[54].

54. As the Rio de Janeiro municipal education system did not take part in the final application, due to a teachers' strike, the state system made 5th grade students from 'Region II' available for validation of the small-scale instrument (62 questions).

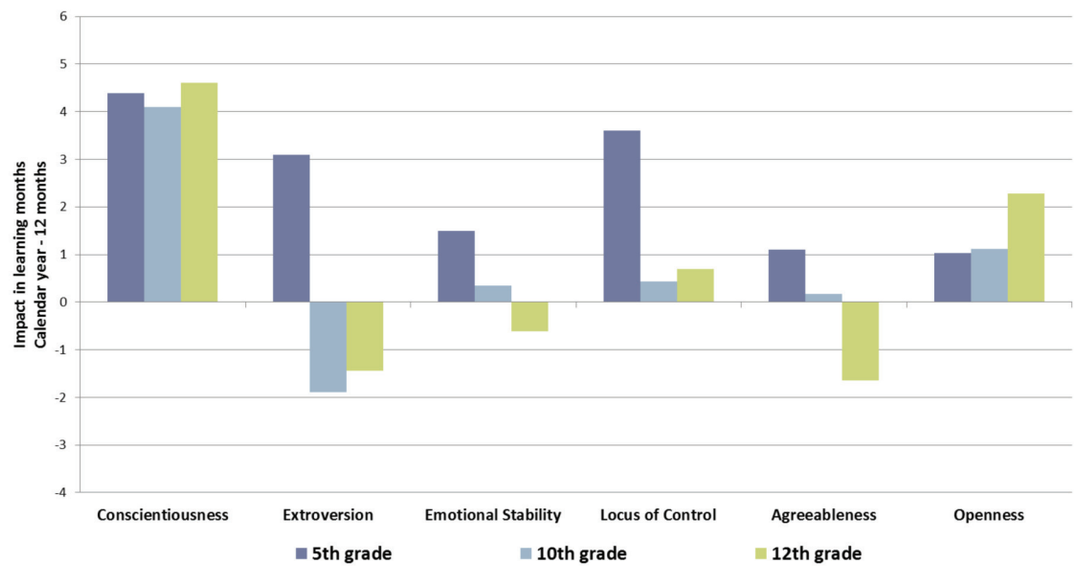
GRAPH 6: IMPACT OF DIFFERENT SOCIAL AND EMOTIONAL ATTRIBUTES ON PERFORMANCE IN PORTUGUESE LANGUAGE – 75TH TO 25TH PERCENTILE



In **Graph 7**, we investigate the effects on Mathematics grades in each year.

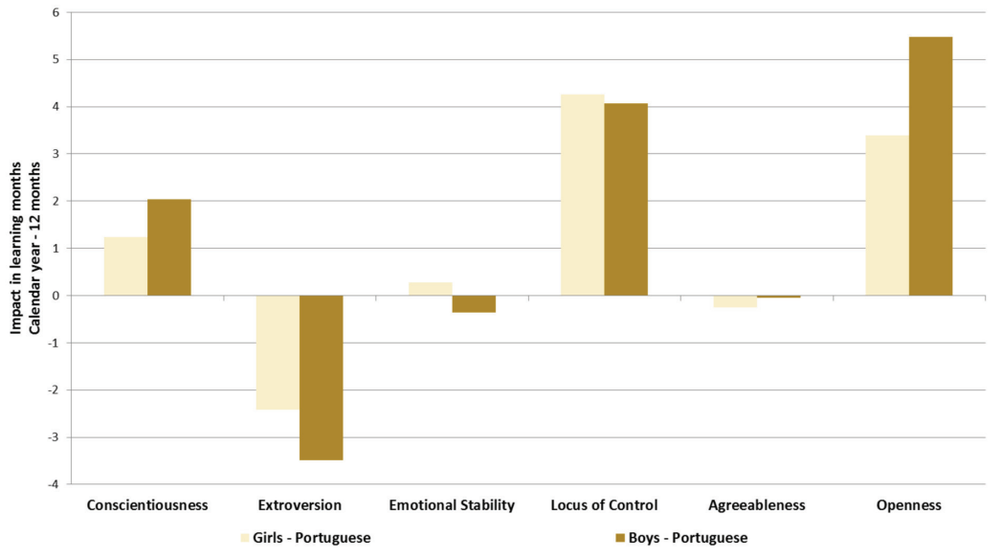
The first important thing to note is that the impact of Conscientiousness is more or less constant between the grades and is always greatest amongst the non-cognitive effects. For students in the 5th grade, Locus of Control and Extraversion stand out as important characteristics for performance, while Openness to New Experiences seems to become more important in the establishment of this result as time goes by.

GRAPH 7: IMPACT OF DIFFERENT SOCIAL AND EMOTIONAL ATTRIBUTES ON PERFORMANCE IN MATHEMATICS – 75TH TO 25TH PERCENTILE

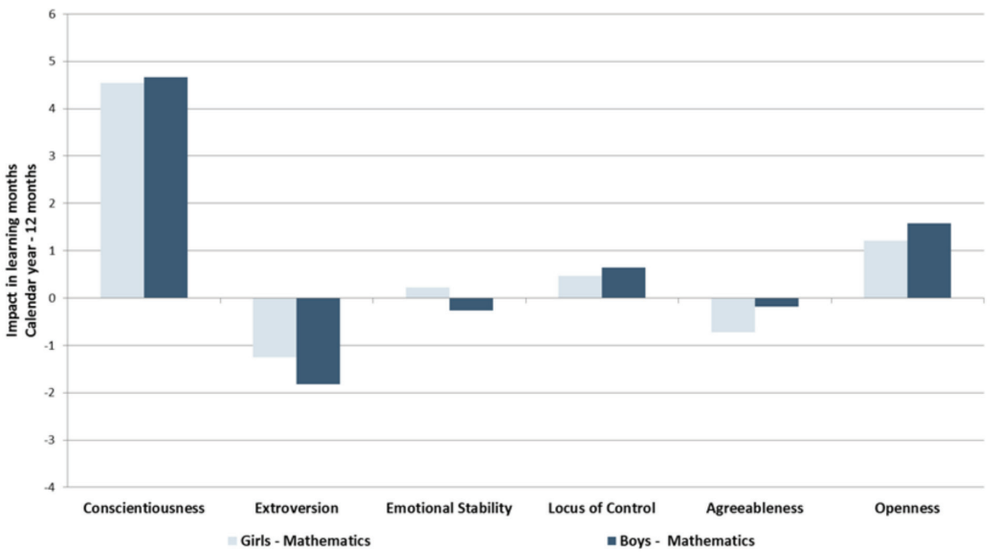


In terms of the sex differences of socio-emotional effects (Graphs 8a and 8b), we can see that, in general, the gains tend to be greater for boys. In terms of the gains associated with an increase in the Openness to New Experiences, a boy who is amongst the 25% least open would experience a rise of 0.15 standard deviation (this being equivalent to 5.5 months of learning) in his Portuguese grades, if he could manage to increase his levels in this trait up to the 75th percentile, which represents almost double the effect that a similar change would have on a girl. The negative effect of Extraversion on Portuguese grades is also markedly greater for boys.

GRAPH 8A: IMPACT OF SOCIAL AND EMOTIONAL TRAITS ON PERFORMANCE IN PORTUGUESE LANGUAGE, BY SEX

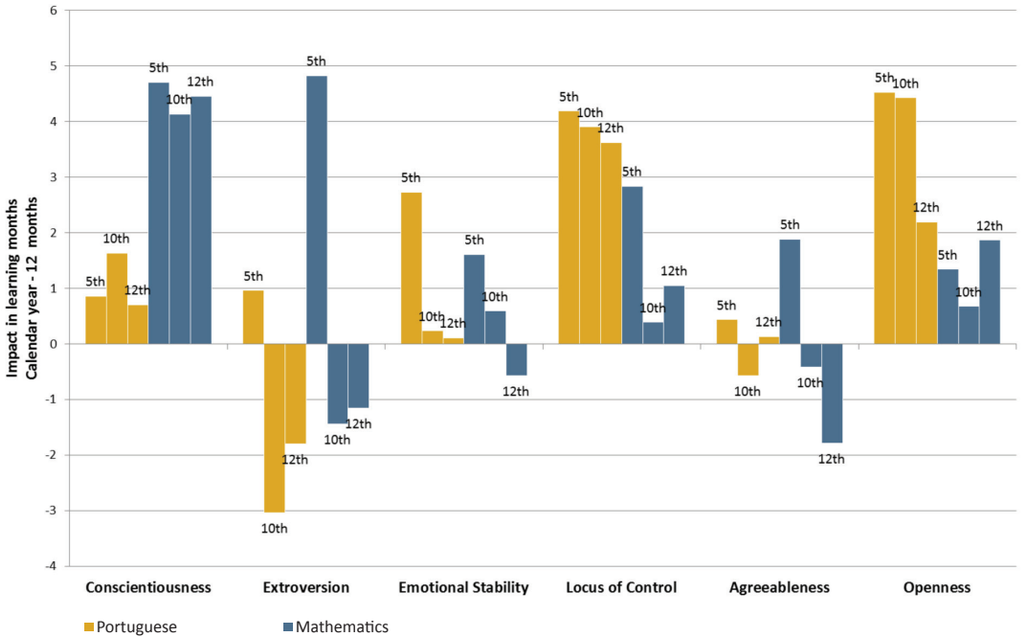


GRAPH 8B: IMPACT OF SOCIAL AND EMOTIONAL TRAITS ON PERFORMANCE IN MATHEMATICS, BY SEX

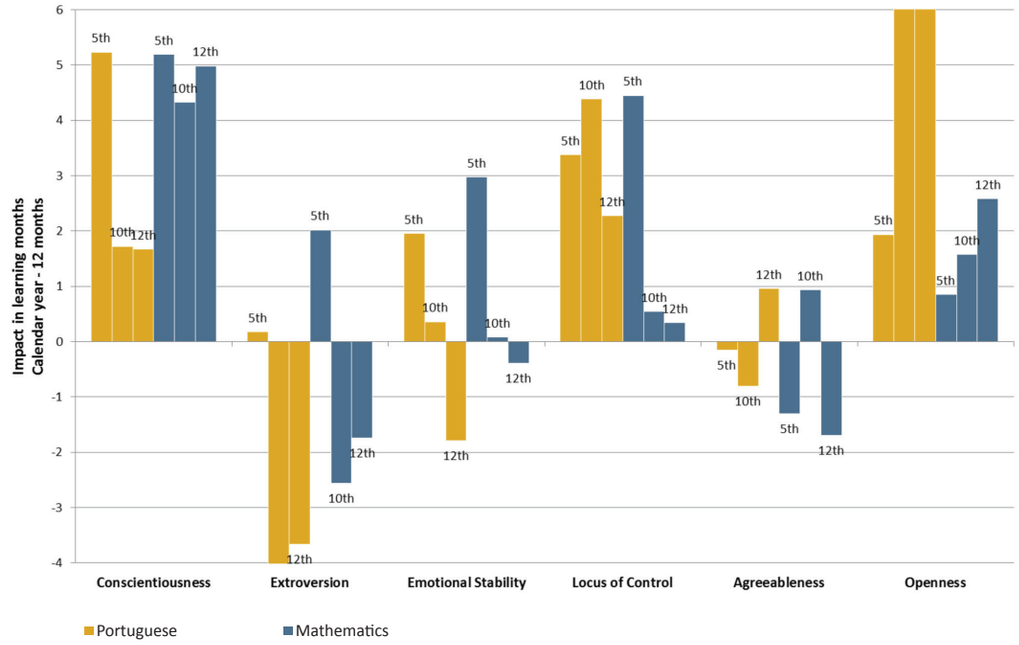


The sex differences by grade are most evident (Graphs 9a and 9b). In the 5th grade, Extraversion is positively associated with girls' performance in Mathematics. For the boys, Conscientiousness also provides benefits for Portuguese grades and Locus of Control for Mathematics. The relationship between Openness and Portuguese grades is strongest in younger age groups for girls, and in older age groups for boys.

GRAPH 9A: DIFFERENCES BETWEEN PORTUGUESE AND MATHEMATICS AMONGST GIRLS, BY GRADE

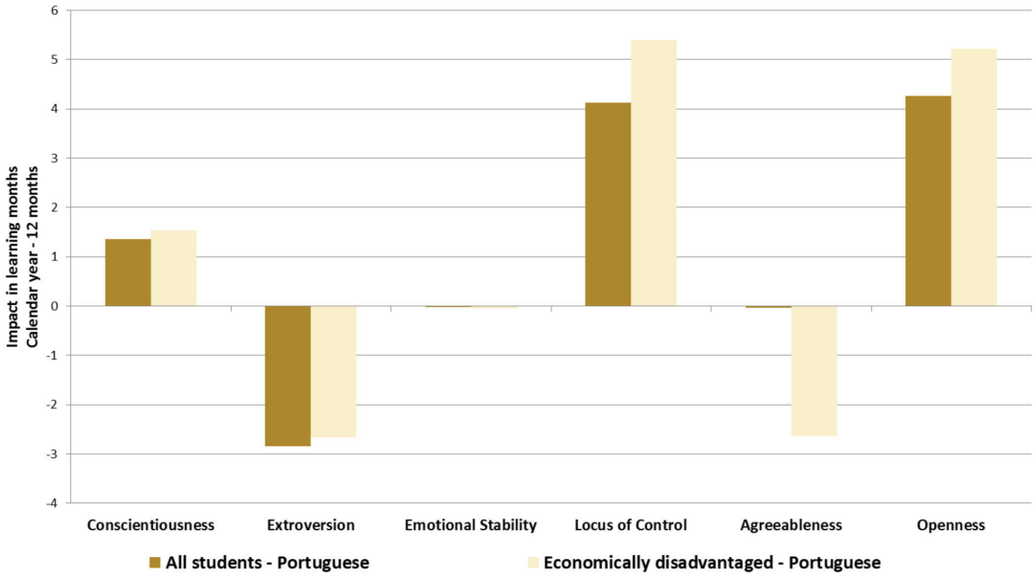


GRAPH 9B: DIFFERENCES BETWEEN PORTUGUESE AND MATHEMATICS AMONGST BOYS, BY GRADE

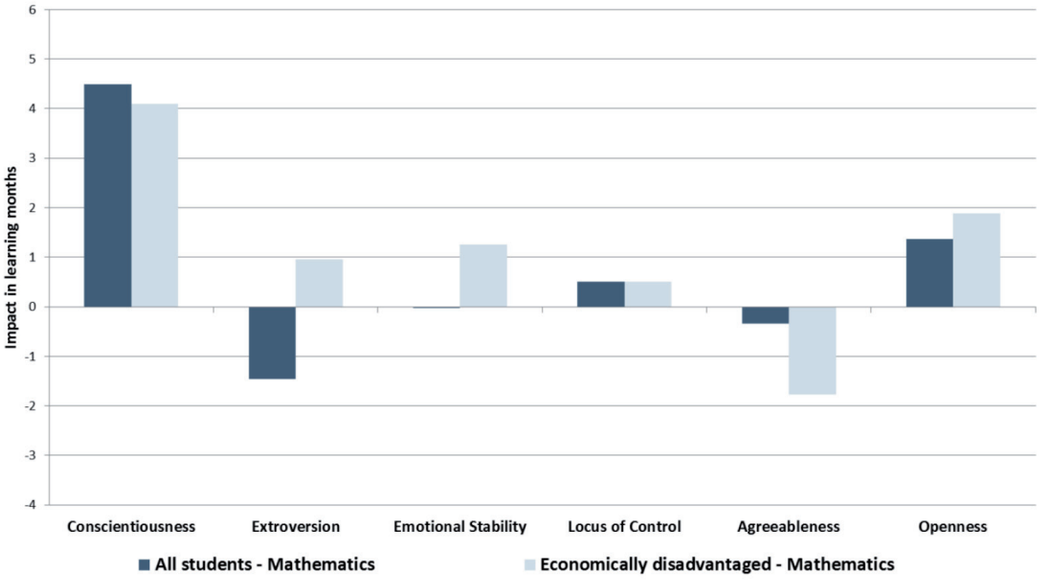


In the following section, we investigate whether the relationship between socio-emotional traits and school performance is different for economically disadvantaged students (Graph 10), here understood as those who are amongst the 20% at the bottom of the socio-economic status distribution[55].

GRAPH 10A: IMPACT OF SOCIAL AND EMOTIONAL SKILLS ON PERFORMANCE IN PORTUGUESE AMONGST THE ECONOMICALLY DISADVANTAGED



GRAPH 10B: IMPACT OF SOCIAL AND EMOTIONAL SKILLS ON PERFORMANCE IN MATHEMATICS AMONGST THE ECONOMICALLY DISADVANTAGED



55. The questionnaire applied to the students contained 11 questions concerning their home lifestyles and ownership of durable goods, all starting with a statement that went something like "Where you live, is there…" (e.g. '…paved roads' in relation to living conditions; or '…a DVD player' in relation to ownership of durable goods). Based upon the replies, binary variables were constructed that assumed the value of '1' if the student's home had the mentioned item or living condition. The variable of 'socio-economic status' is the sum of these binary indicators. Robustness checks with synthetic indices built from factor analysis and other dimension-reduction techniques were made and confirmed the qualitative conclusions of this study.

Overall, the relationship between socio-emotional characteristics and performance shows a similar pattern between the economically disadvantaged students and those who are economically advantaged. Locus of Control and Openness have a relatively greater effect amongst the economically disadvantaged, even though they are important to the whole sample as well.

3.3 Variables that influence social and emotional levels of students

1. Determinants for the social and emotional levels of the students

The aim of this section is to investigate which individual, family and classroom characteristics are amongst those that directly influence social and emotional levels of the students from the State of Rio de Janeiro.

Once again, the estimated model is linear in the explanatory variables and includes the fixed effect of the school. Table 10 shows the most important effects found, highlighting that the exercises also included the averages of socio-emotional skills and family characteristics of their classmates, as a *proxy*[56] for the possible peer-effects[57] that can affect the individual levels of the socio-emotional traits.

The impacts are all measured in standard deviations from the socio-emotional characteristics, which are equal to a little less than the difference between the 75th and 25th percentiles used earlier. Thus, for example, **if we find that students whose parents encourage them to study have an extra standard deviation of Conscientiousness, we could rephrase this fact saying that 23%**[58] **of the difference between students with high and low Conscientiousness would be eliminated if the parents of the students with low conscientiousness started to encourage them to study.**

Throughout the text, we have organized the results by groups of variables, starting with individual characteristics (sex, age, and an indicator of being at the right age for the grade), moving on through family characteristics (mother's education, father's literacy, socio-economic status, presence of the mother at home, number of books, and an indicator that parents encourage study), concluding with classroom characteristics (proportion of boys).

56. Sometimes we want to estimate the causal relationship between an outcome variable, Y, and its determinants, X, but not observe X. We instead observe another variable, Z, that is closely related to X and which theoretically should not have a direct effect on Y. In these cases, Z is called in as a Proxy for X. When Z is included in a statistical exercise that seeks to estimate the determinants of Y, we interpret its role as if it were in fact capturing the effect of X on Y.
57. The peer-effects are the effects that classmates' characteristics can have on the student's development.
58. 0.26/1.13 =23%

TABLE 10: IMPACT OF THE CHARACTERISTICS OF THE STUDENT AND HIS/HER ENVIRONMENT ON NON-COGNITIVE SKILLS

Characteristic	Conscientiousness	Extraversion	Internalizing Behavior	Locus of Control	Agreeableness	Openness to New Experiences
Proportion of boys in the class	**-0.16**	**-0.23**	**0.41**	0.04	*-0.14*	-0.02
Age	0.00	-0.01	**0.03**	**0.04**	0.01	-0.01
Repeated years	0.02	**-0.18**	**-0.1**	**-0.16**	**-0.11**	**-0.09**
Size	0.00	**-0.01**	*0.00*	0.00	**0.00**	**0.00**
Mother's education						
Finished Primary School	0.00	0.08	-0.02	0.05	-0.01	0.03
Did not graduate High School	-0.05	**0.11**	0.03	**0.10**	-0.04	*0.05*
Graduated High School	**-0.09**	**0.11**	-0.04	**0.08**	**-0.08**	0.01
College	**-0.13**	**0.1**	-0.06	0.06	-0.06	*0.07*
Doesn't know	**-0.09**	-0.05	-0.03	0.00	**-0.11**	**-0.1**
Literate father	**0.06**	0.01	0.03	**0.09**	*0.04*	-0.02
Socio-economic status	**-0.05**	**0.04**	**-0.02**	**0.03**	*-0.02*	**-0.02**
Number of books at home						
1 shelf	**0.05**	**0.1**	**0.04**	**0.05**	**0.11**	**0.19**
1 bookcase	**0.07**	**0.13**	0.02	0.00	**0.13**	**0.35**
More than one bookcase	**0.1**	**0.17**	0.02	-0.04	**0.12**	**0.45**
Parents encourage studies	**0.26**	**0.1**	**0.09**	**0.21**	**0.17**	**0.12**
Lives with the mother	*0.03*	**-0.07**	**0.07**	**0.09**	0.02	**-0.06**
Boy	**-0.12**	**-0.13**	**0.59**	**-0.04**	**-0.12**	**0.12**
Difference between the 75th and 25th percentiles	1.13	1.17	1.19	1.22	1.29	1.17

3.3.1 Individual characteristics

Our results confirm evidence obtained internationally (see, for example, Soto et al, 2011) that girls are kinder and more extrovert than boys, as well as having more conscientiousness. In our sample, the boys showed themselves to be more Open to New Experiences, but with a more external Locus of Control, reflecting a tendency to attribute commonly experienced situations to chance, luck or actions and decisions taken by others. The big difference, however, was in Emotional Stability, that was much higher for boys.

Age was a statistically significant variable in determining Emotional Stability and Locus of Control, the older students having a more internal Locus (or in other words, they have a greater tendency to attribute experienced situations to those decisions and actions they themselves take) and greater Emotional Stability.

One factor related to age, but which includes other phenomena of socio-emotional development that needs more study, is that students who are held back from advancing in school years tend to be less extravert and kind, less autonomous (or have an external Locus), less open to new experiences, and more likely to internalize their emotions to a greater degree. In all these cases, the effects were statistically significant and the importance relatively great[59].

3.3.2 Family characteristics

Regarding the family characteristics, one unexpected fact is that, controlling attitudes and habits, Conscientiousness declines with mother's education and with the socio-economic status of the family. The relationship between the education levels of mothers and wealth, on the one hand, and Conscientiousness, on the other, is complex and ambiguous. It is generally believed that richer and better educated families encourage

59. In all these cases, the effects were statistically significant and the magnitudes relatively large.

their children to study more and have a home environment that is more favorable to study (with books, silent areas in which to study, etc.), amongst other factors that, in theory, would stimulate greater Conscientiousness. On the other hand, it is possible to suggest that the children in these families need to make less effort to achieve what they want, meaning that the family's wealth may have a negative effect on the children's Conscientiousness. As in our exercise we compared only children who received the same encouragement from their parents and had a similar family environment, it is possible that the remaining relationship between the education of the mothers and wealth on the one hand, and Conscientiousness on the other, would be a negative factor.

Put another way, when we analyze children receiving the same encouragement from their parents, and the same standard of quality in the school environment, we can see that those whose mothers with a higher level of education seem to be less conscientious, or in other words, they showed less of a tendency to be organized, hardworking and responsible. **This fact highlights the potential for a social and emotional approach to boost the learning performance of more economically disadvantaged students,** since those parents who received less schooling and have fewer economic resources do not appear to be at a disadvantage compared to those parents who are more economically advantaged; on the contrary, they appear to be more able to assist their children in developing these characteristics.

Amongst the effects associated with family characteristics, the enormous effect that parents' encouragement in their children's studies has on Conscientiousness, the internalization of the Locus of Control, Agreeableness and Openness to New Experiences, is of particular note. In all these cases, this variable was amongst the two main determining factors of the construct in question.

Another relationship that clashes is the effect that the existing number of books in the household (which could possibly suggest the value that the parents give to education and knowledge) has on the Openness to New Experiences, and to a lesser degree on Extraversion and Agreeableness. Around 40% of the existing difference between young people with a high level of Openness and those with a small amount of this trait could be eliminated if the quality of the family environment (as measured by the number of books) moved from the minimum level to the maximum.

3.3.3 Learning environment characteristics

Although our database is not especially rich in information on the classroom and the learning environment[60], some results suggest that the characteristics of the class could have an important effect on the students. Amongst these effects we can highlight the negative impact that the size of the class has on Extraversion, Agreeableness and Openness to New Experiences, maybe reflecting the smaller amount of space that is available for the affirmation of individuality.

The proportion of boys in the group, meanwhile, proved to be of significance and had a great effect in the determination of Conscientiousness and Extraversion (negative impact), and Emotional Stability. In this case, this parameter is possibly noting the effect that characteristics of the peers have on the social and emotional development of the individual.

60. We would ideally have liked to have had information on the quality and intensity of the relations between teachers and students, the actual length of lessons, acoustic isolation and brightness of the classroom, indicators of violence in the schools' neighborhood, and indicators of student and teacher violence and criminal activity, amongst other factors that could directly influence the formation of social and emotional characteristics of the young people. Without these facts, our indicators ended up gathering not only the direct effects, but also part of the effects of these omitted variables, to the extent that they are correlated with the observed variables (and included in the exercise).

3.4 Summary of the Analyses: Implications for Public Policies

The data collected from Rio de Janeiro public school system shows that the students' social and emotional characteristics are amongst the main determinants of their learning performance measured by grades achieved in standard Portuguese and Mathematics tests.

Amongst the social and emotional traits, Conscientiousness (the tendency to be organized, hardworking and responsible) was shown to be that which is most closely connected with Mathematics results, and Locus of Control (a construct that reflects the extent to which individuals attribute situations currently being experienced to decisions and attitudes they themselves took in the past, or to chance, luck or decisions taken by others) and Openness to New Experiences as the most relevant in determining Portuguese grades.

These results are fully in line with the extensive review of the literature available performed by Almlund et al (2011). Separate analyses performed on different grades and sexs show that the role of social and emotional traits is greater in lower grades (especially that of Locus of Control, that stands out as an important characteristic for performance in the 5th grade) and amongst boys (since, as we recognized, an increase in Openness to New Experiences represents almost twice the impact for a boy that a similar change would have in a girl, in much the same way that the negative effect of Extraversion on Portuguese grades is also notably greater for boys).

In investigating the factors that explain individual differences in social and emotional characteristics, the great impact made by parents when they encourage their children to study is very noticeable. The number of books there are in the home shows itself to be highly associated with the development of social and emotional traits in a way that is favorable to learning, especially Openness to New Experiences. One possible interpretation is that this variable represents the value that the family attaches to education and learning. We also found evidence that the characteristics of classmates can significantly affect individual social and emotional development.

Conclusion

The importance of a quality education for individual and collective prosperity is universally recognized. However, there is no consensus as to what the definition of a quality education for the 21th century should be, since many of the current curriculums and pedagogic practices are unable to respond to the challenges of our time. It is understood that, despite being absolutely fundamental, the skills related to literacy, numeracy and the acquiring of knowledge relating to traditional curricular content, are not enough to guarantee academic, professional and personal success in the world today. We must ask ourselves what is missing.

What everything points to, and which the results of the Project back up, is that the social and emotional approach is one of the fundamental elements in finding an answer to this issue. As is clearly outlined in this Report, the studies in the areas of economics and psychology demonstrate that this dimension, generally neglected by public policies and evaluation systems, is especially relevant in the advancement of learning in school and providing benefits in the different areas of well-being that are important to adult life.

In response to the need for a deeper understanding of these skills and the mechanisms that can be used to develop them, the Ayrton Senna Institute and the Organization for Economic Cooperation and Development, through the Centre for Educational Research and Innovation, set themselves the challenge of developing a tool for the large-scale measurement of the skills in the school context. They further challenged themselves to establish a database of proportions never before seen in Brazil, which could be used to extend research in this area and offer raw material to support the development of public policies.

As presented in the report, this task was successfully performed considering that the measuring instrument developed showed that it could meet all the requirements necessary to effectively and feasibly evaluate the social and emotional skills described in scientific literature as being the most important. The psychometric exercises described in chapter 2 show that the instrument has a very high level of internal consistency and is able to reproduce the patterns that exist in international literature (see Soto et al, 2011) in terms of the different sexes' and ages' differences in social and emotional characteristics. On the other hand, the conclusions of chapter 3 show that the social and emotional characteristics particularly associated with learning in Brazil are the same as those reported by other authors such as Almlund et al (2011). These results lead us to believe that we are reliably gauging the most important characteristics.

As explained in Chapter 3, the impact of Conscientiousness on academic results is the greatest of the social and emotional skills and is more or less constant across the grades analyzed (5th, 10th and 12th grades). For students in the 5th grade, the Locus of Control[61] as well as Extroversion, can also be highlighted as important characteristics for performance, whilst Openness to New Experiences seems to take on more importance in the determining of this result as time goes by.

61. **The Locus of Control** reflects the degree to which individuals attribute commonly experienced situations to decisions and attitudes they themselves have taken (internal Locus), or to chance, luck or actions and decisions taken by others (external Locus). The locus of control is defined by a continuum between two opposing spectrums: the external locus and the internal locus.

As for the differences of impact between the subjects, it could be seen that grades in Mathematics across the board benefitted greatly from an increase in their levels of Conscientiousness (the tendency to be organized, responsible and hardworking), whilst grades in Portuguese were more effected by variations in the Locus of Control and Openness to New Experiences.

In investigating how family characteristics influence the social and emotional development of students, attention was drawn to the effect that parents' encouragement of their children to study had on Conscientiousness, internalization of the Locus of Control, Kindness and Openness to New Experiences. With reference to the relationship of our measurements with different family context variables, it is interesting to note that the averages of all the six constructs measured vary from the mother's level of education very little, typically seen as one of the main determining factors for school performance. In contrast, the social and emotional characteristics appear to vary a great deal as indicators of attitudes and habits of the student and his or her parents, as does the frequency of reading and the students' understanding that they are receiving parental encouragement to study.

Social and Emotional or Non-cognitive Nationwide Assessment

As the result of this first stage of analysis, the Project team is pleased to be able to present the measuring instrument developed and validated over the course of the research and which, together with its support materials (still under development)[62], makes up the social and emotional evaluation program that has been named *SENNA (Social and Emotional or Non-cognitive Nationwide Assessment)*. The first version of the social and emotional questionnaire may be found as an Attachment to this Report, as can the other documents that make up the set of collected data used for the pilot assessment developed in Rio de Janeiro.

The construction of the SENNA is an ongoing task, since each application generates important information contributing to the fine tuning of the instrument. New results of analyses of the information collected in the pilot application will be published shortly. At the same time, the final results of the analysis of the information collected during the pilot scheme will also be made available. The findings from both the SENNA program and the databases will be made freely available for administrators, educators and researchers to make use of them for improvements in the quality of education.

The Ayrton Senna Institute feels that the evaluation of learning, in its many varied forms, is crucial to the expansion of the global focus on what constitutes a quality education. In presenting a gauge for the evaluation of skills that have been proven to be important to the future of children and young people, but which are still not included in the established evaluation systems, our researchers intend to take the first step towards the construction of a monitoring system that is as fully developed as the education we intend to develop.

62. User manuals and application for public policies.

Bibliography

Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. D. (2011). Personality psychology and economics. *In Handbook of the Economics of Education*, Vol. 4, E. Hanushek, S. Machin, and L. Woessman, eds. Amsterdam: Elsevier. pp. 1-181.

Bandura, A. (1977). Self-efficacy: toward a unifying theory of behavioral change. *Psychological review* 84 (2), 191-215.

Bandura, A. (1986). Fearful expectations and avoidant actions as coeffects of perceived self-inefficacy. *American Psychologist* 41, 1389-1391

Bandura, A. (1989). Regulation of cognitive processes through perceived self-efficacy. *Developmental psychology* 25 (5), p. 729.

Bandura, A. (2001). Social cognitive theory: An agentic perspective*. Annual review of psychology* 52, 1-26.

Bodrova, E., & Leong, D. J. (2007). *Tools of the Mind*. Pearson.

Branden, N. (1969). The psychology of self-esteem. *New York: Bantam.*

Carneiro, P., Crawford, C., & Goodman, A. (s.d.). The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes. CEE Discussion Papers 0092, *Centre for the Economics of Education.*

Cohen, J. (1992). A Power Primer *Psychological bulletin*, 112(1,155-159).

Cunha , F., J. Heckman e S. Schennach (2010) "Estimating the Technology of Cognitive and Noncognitive Skill Formation,". Econometrica, 78(3), 883–931

Delors, J., Chung, F., Geremek, B., Gorham, W., Kornhauser, A., Manley, M., et al. (1999). Relatório para a UNESCO da Comissão Internacional sobre Educação para o século XXI. In: *Educação: um tesouro a descobrir*. São Paulo: UNESCO.

Denissen, J. J., Zarrett, N. R., & Eccles, J. S. (2007). I like to do it, I'm able, and I know I am: Longitudinal couplings between domain-specific achievement, self-concept, and interest. *Child Development* 78 (2), pp. 430-447.

Duckworth, A. L., & Seligman, M. E. (2005). Self-discipline outdoes IQ in predicting academic performance of adolescents. *Psychological science* 16 (12), pp. 939-944.

Duncan, G. J., & Magnuson, K. (2010). The Nature and Impact of Early AchievementSkills, Attention Skills, and Behavior Problems. In G. Duncan & R. Murnane (eds). Social Inequality and Educational Disadvantage. NY: Russell Sage Foundation.

Embretson, S. E.& Reise, S. P. (2000). *Item Response Theory for Psychologists.* New Jersey: Lawrence Erlbaum.

Evaluation of the National Guard Youth ChalleNGe Program . (s.d.). Acesso em 2014, disponível em MDRC: http://www.mdrc.org/project/evaluation-national-guard-youth-challenge-program#featured_content

BIBLIOGRAPHY

Fergusson DM, *Horwood* LJ. Early Conduct Problems and Later Life Opportunities. Journal of Child Psychology and Psychiatry, 1998; 39(8): 1097-1108.

Greene, B. A., & Miller, R. B. (1996). Influences on achievement: Goals, perceived ability, and cognitive engagement. *Contemporary Educational Psychology* 21 (2), pp. 181-192.

Gutman, L., & Schoon, I. (2013). *The impact of non-cognitive skills on outcomes for young people: literature review.* Fonte: Education Endowment Foundation: http://educationendowmentfoundation.org.uk/uploads/pdf/Non-cognitive_skills_literature_review.pdf

Hansford, B. C., & Hattie, J. A. (1982). The relationship between self and achievement/performance measures. *Review of Educational Research* 52 (1), pp. 123–142.

Harter, S. (1982). The perceived competence scale for children. *Child developmentI* 53 (1), 87-97.

Heckman, J. J., & Kautz, T. (2012). Hard evidence on soft skills. In: *Labour Economics* 19 (4), 451-464. Elsevier.

Heffron, J. M. (1997). Values in Education: Social Capital Formation in Asia and the Pacific. *Defining values*, 3–27.

HighScope Perry Preschool Study. (s.d.). Fonte: HighScope: http://www.highscope.org/content.asp?contentid=219

Jacobs, J. E., Lanza, S., Osgood, C. W., Eccles, J. S., & Wigfield, A. (2002). Changes in children's self-competence and values: Gender and domain differences across grades one through twelve. *Child development* 73 (2), pp. 509-527.

John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. *Handbook of personality: Theory and research*, pp. 102-138.

Kusché, C. A., & Greenberg, M. T. (1994). The PATHS *curriculum*.

Kyllonen, P. C., Walters, A. M., & Kaufman, J. C. (2011). *The Role of Noncognitive Constructs and Other Background Variables in Graduate Education*. Educational Testing Service (ETS). Princeton, New Jersey.

Kyllonen, P. K., Lipnevich, A. A., Burrus, J. & Roberts, R. D. (2008). *Personality, Motivation, and College Readiness: A Prospectus for Assessment and Development*. Educational Testing Service (ETS). Princeton, New Jersey.

Lee, W. O. (2013). *Education and 21st Century Competencies*. Keynote paper presented at the Education and 21st Century Competencies,hosted by the Ministry of Education, Oman, 22-24 September 2013

Linacre, J. M. (2013). *Winsteps® Rasch measurement computer program.* Beaverton, Oregon: Winsteps.com

Lleras, C. (2008). Do skills and behaviors in high school matter? The contribution of noncognitive factors in explaining differences in educational attainment and earnings. Social Science Research 37, pp. 888-902.

Lócus de Controle de Rotter. (s.d.). Fonte: http://www.mccc.edu/~jenningh/Courses/documents/Rotter-locusofcontrolhandout.pdf

Marsh, H. W., & Craven, H. (1997). Academic self-concept: Beyond the dustbowl. In: G. D. Phye, *Handbook of classroom assessment: Learning, achievement and adjustmen* (pp. 131-198). Orlando: Academic Press.

Marsh, H. W., & Craven, R. G. (2006). Reciprocal effects of self-concept and performance from a multidimensional perspective: Beyond seductive pleasure and unidimensional perspectives. *Perspectives on Psychological Science 1* (2), pp. 133-163.

Marsh, H. W., & Shavelson, R. (1985). Self-concept: Its multifaceted, hierarchical structure. *Educational Psychologist 20* (3), 107-123.

Marsh, H. W., Lüdtke, O., Muthén, B., Asparouhov, T., Morin, A. J., Trautwein, U., & Nagengast, B. (2010). A new look at the big five factor structure through exploratory structural equation modeling. *Psychological Assessment,* 22(3), 471.

Martin, R. P. (1989). Activity Level, Distractibility, and Persistence: Critical Characteristicsin Early Schooling. In: G. A. Kohnstamm, J. E. Bates, & M. K. Rothbart. Chichester: John Wiley and Sons.

Martins, P. S. (2010). Can Targeted, Non-Cognitive Skills Programs Improve. *IZA Discussion Paper* No. 5266.

Matthews, G., & Deary, I. J. (1998). Personality traits. Cambridge University Press.

McGrew, K.S.(2007).Beyond IQ: A Model of Academic Competence and Motivation (Institute for Applied Psychometrics). Retrieved October 9,2009, from http://www.iapsych.com/acmcewok/map.htm

Mischel, W., Shoda, Y., & Rodriguez, M. L. (1989). Delay of gratification in children. Science 244 (4907), 933-938.

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of counseling psychology* 38 (1), p. 30.

Muthén, L. K. & Muthén, B. O. (2012). *Mplus User's Guide*. Seventh Edition. Los Angeles, CA: Muthén & Muthén.

O'Mara, A. J., Marsh, H. W., Craven, R. G., & Debus, R. L. (2006). Do self-concept interventions make a difference? A synergistic blend of construct validation and meta-analysis. *Educational Psychologist* 41 (3), 181-206.

Poropat, A. E. (2009). A meta-analysis of the five-factor model of personality and academic performance. *Psychological bulletin* 135 (2), p. 322.

Putnam, R. D. (1995). Bowling alone: America's declining social capital. *The Journal of Democracy* 6 (1), pp. 65–78.

Ramey, C. T. (1974). Carolina Abecedarian. *The Carolina Abecedarian Project: A Longitudinal and Multidisciplinary Approach to the Prevention of Developmental Retardation*. Chapel Hill, North Carolina.

Revelle, W. (2009/2010). psych: *Procedures for Personality and Psychological Research.* R package version 1.0-91. http://personality-project.org/r, http://personality-project.org/r/psych.manual.pdf.

Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: a systematic review and meta-analysis. *Psychological bulletin* 138 (2), p. 353.

Ridgell, S. D., & Lounsbury, J. W. (2004). Predicting Academic Success: General Intelligence,"Big Five" Personality Traits, and Work Drive. *College Student Journal* 38 (4), 607.

Sailer, O. (2005): crossdes: A package for design and randomization in crossover studies. Rnews 5/2, 24-27.

Soto, C., John, O., Gosling, S., & Potter, J. (2011). Age differences in personality traits from 10 to 65: Big Five domains and facets in a large cross-sectional sample. J*ournal of personality and social psychology* 100 (2), p. 330.

Sternberg, R. J. (1999). *Handbook of creativity*. Cambridge University Press.

Wright, B.D. & Masters, G.N. (1982). *Rating scale analysis*. Chicago: MESA.

Instrument references

Big Five Inventory (BFI)
John, O. P., Naumann, L. P., & Soto, C. J. (2008). Paradigm Shift to the Integrative Big-Five Trait Taxonomy: History, Measurement, and Conceptual Issues. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), Handbook of personality: Theory and research (pp. 114-158). New York, NY: Guilford Press.
John, O. P., Donahue, E. M., & Kentle, R. L. (1991). The Big Five Inventory--Versions 4a and 54. Berkeley, CA: University of California,Berkeley, Institute of Personality and Social Research.

Big Five for Children (BFC)
Barbaranelli, C., Caprara, G., Rabasca, A., & Pastorelli, C. (2003). A questionnaire for measuring the Big Five in late childhood. Personality and Individual Differences, 34 (4), 645-664.

Core Self Evaluations
Judge, T. A., Erez, A., Bono, J. E., & Thoresen, C. J. (2003). The core self-evaluations scale: Development of a measure. *Personnel Psychology,* 56(2), 303-331. doi: 10.1111/j.1744-6570.2003.tb00152.x

Grit Scale
Duckworth, A.L, & Quinn, P.D. (2009). Development and validation of the Short Grit Scale (Grit-S). Journal of Personality Assessment, 91, 166-174.http://www.sas.upenn.edu/~duckwort/images/Duckworth%20and%20Quinn.pdf

Hierarquical Personality Inventory for Children (HiPIC)
Mervielde, I., & De Fruyt, F. (1999). Construction of the Hierarchical Personality Inventory for Children (HiPIC). In Personality psychology in Europe. Proceedings of the Eight European Conference on Personality Psychology/I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.).-Tilburg: Tilburg University Press, 1999 (pp. 107-127).

Norwick-Strickland Locus of Control Scale
Nowicki, Stephen; Strickland, Bonnie R. Journal of Consulting and Clinical Psychology, Vol 40(1), Feb 1973, 148-154. doi:10.1037/h0033978

Rosenberg Self-Esteem Scale
Rosenberg, M. (1979). Conceiving the Self. New York: Basic Books.

Self-Efficacy Questionnaire for Children
Muris, P. (2001). A brief questionnaire for measuring self-efficacy in youths. Journal of Psychopathology and Behavioral Assessment, 23(3), 145-149.

Social Skills Improvement System (SSIS)
Gresham & S. N. Elliott (2008) Social Skills Improvement System Rating Scales. Minneapolis, MN: NCS Pearson.

Strengths and Difficulties Questionnaire (SDQ)
Goodman R (1997) The Strengths and Difficulties Questionnaire: A Research Note. Journal of Child Psychology and Psychiatry, 38, 581-586.

Annexes

I. Tables and Graphs

TABLE A1: PRELIMINARY STUDY RESULTS - ENTIRE SAMPLE

Test	n.itens	Mean	item.p.min	N	SD	Median	Min	Max
AUTO	24	5,0	4.8	168	2.6	4	1	19
BFC	65	11.8	5.5	185	6.8	10	2	50
BFI	44	8.6	5.1	184	5.4	7	2	30
CORE	12	2.7	4.5	122	0.9	3	1	5
GRIT	8	3.1	2.6	170	2.3	2	1	16
LOCUS	21	6.1	3.4	170	3.4	5	1	27
ROSEMBERG	10	6.1	1.6	43	6.5	5	1	44
SDQ	25	5.2	4.8	166	2.7	5	2	22
Mean			**4.1**					
40m			**162.2**					

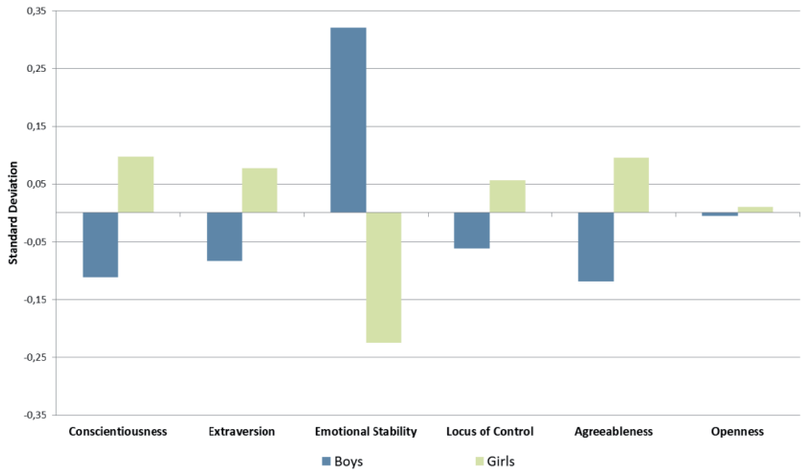
TABLE A2: PRELIMINARY STUDY RESULTS - 5TH GRADE

Test	n.itens	Mean	item.p.min	N	SD	Median	Min	Max
AUTO	24	8.6	2.8	21	3.8	9	3	19
BFC	65	20.3	3.2	31	8.8	19	5	50
BFI	44	15.2	2.9	32	6.1	15	4	26
GRIT	12	6.5	1.9	22	3.9	5	2	16
LOCUS	8	10.8	0.7	25	4.8	10	5	27
ROSEMBERG	21	5.4	3.9	23	2.6	5	1	13
SDQ	10	8.7	1.2	21	5.6	7	2	22
Mean			**2.4**					
40m			**94.3**					

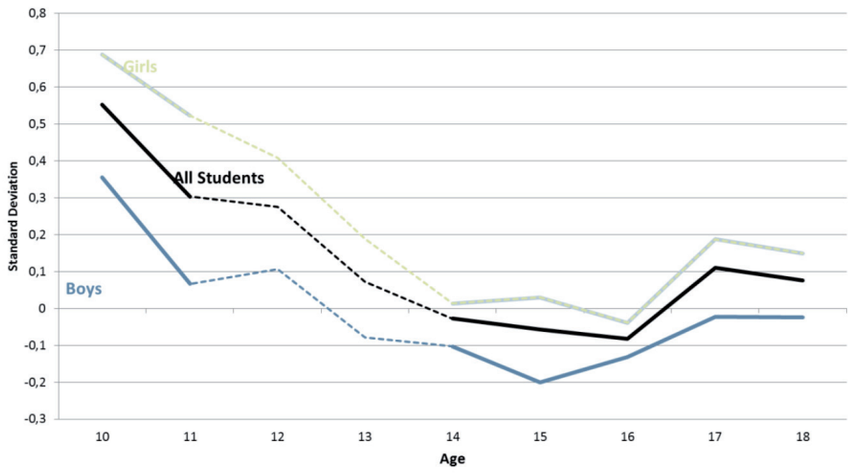
TABLE A3: PRELIMINARY STUDY RESULTS - AT THE 90TH PERCENTILE

5th grade		P90		10th grade		P90		
AUTO	24	14.6	1.6	0	24	6	4	0
BFC	65	27	2.4	0	65	14	4.64	0
BFI	44	23.7	1.9	0	44	9	4.89	0
GRIT	12	12	1	0	12	3	4	0
LOCUS	8	17.2	0.5	0	8	8	1	0
ROSEMBERG	21	8	2.6	0	21	0	0	0
CORE	0	0	0	0	0	4	0	0
SDQ	10	18.4	0.5	0	10	6	1.67	0
Mean			**1.5**				**3.4**	
40m			**60**				**135**	

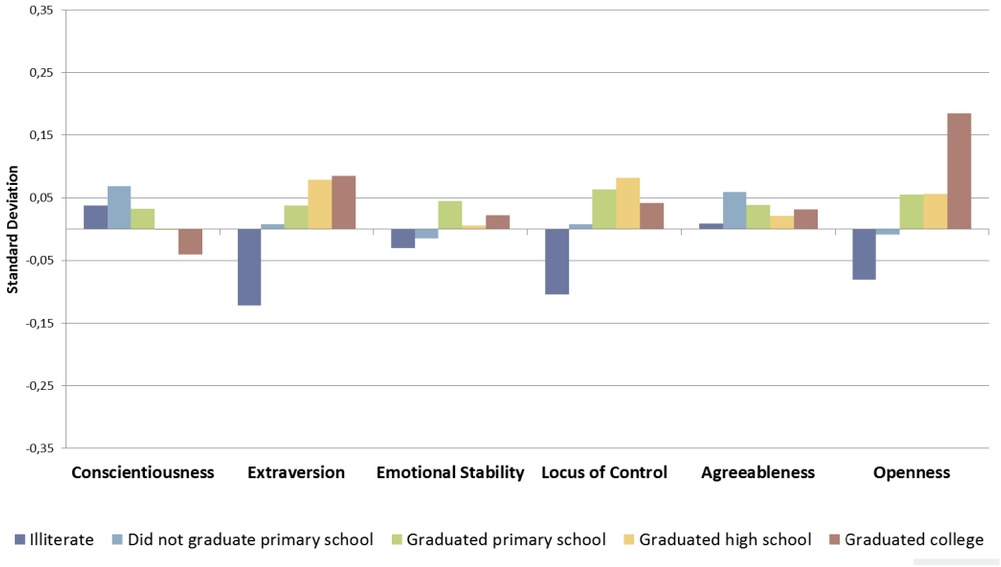
GRAPH A1: SOCIAL AND EMOTIONAL CHARACTERISTICS BY SEX



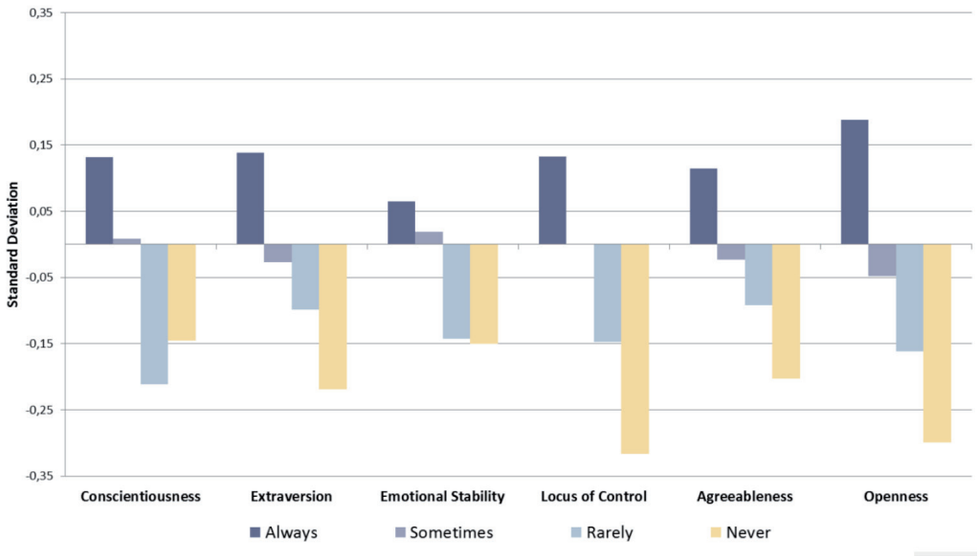
GRAPH A2: AVERAGE DEVELOPMENT OF CONSCIENTIOUSNESS AMONGST STUDENTS AT THE CORRECT AGE FOR THE GRADE



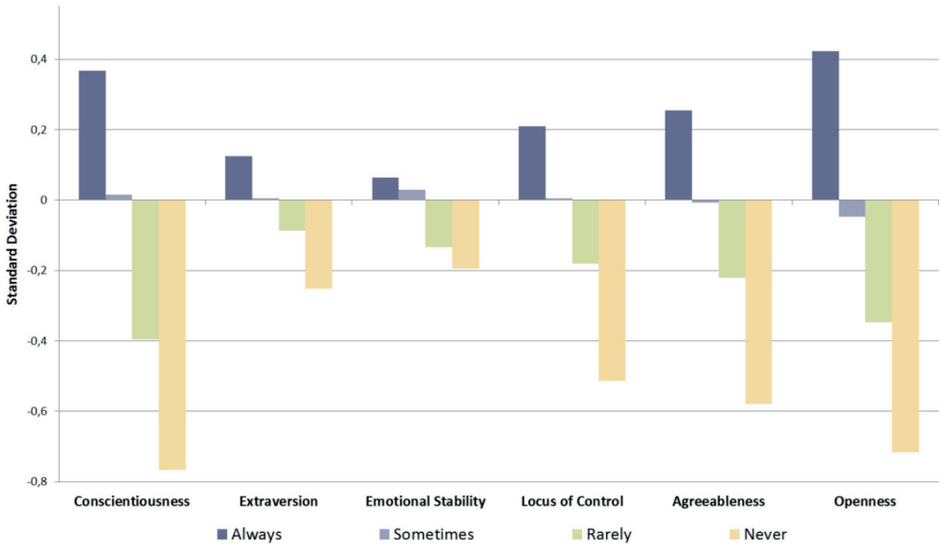
GRAPH A3: DIFFERENCES IN SOCIAL AND EMOTIONAL CHARACTERISTICS IN ACCORDANCE WITH THE MOTHER'S LEVEL OF EDUCATION



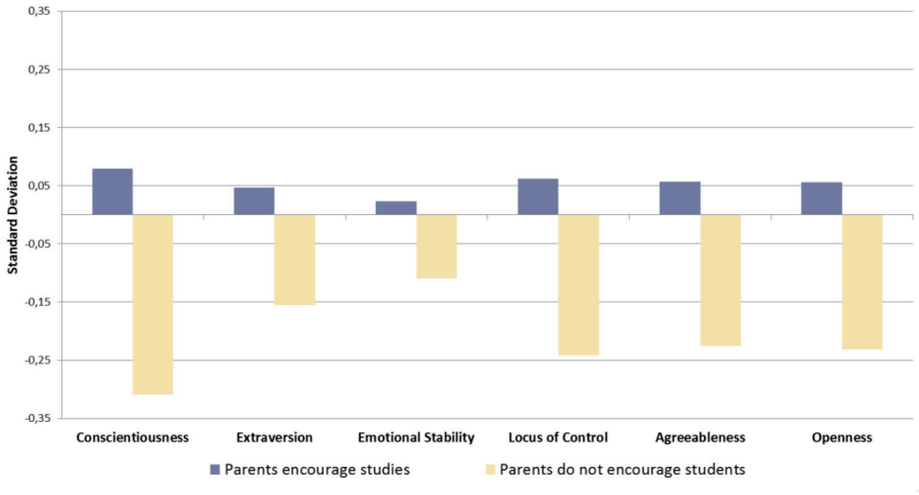
GRAPH A4A: AVERAGE OF SOCIAL AND EMOTIONAL CHARACTERISTICS ACCORDING TO THE FREQUENCY THAT CHILDREN SEE THEIR PARENTS READING



GRAPH A4B: STUDENT READING HABITS AND SOCIAL AND EMOTIONAL CHARACTERISTICS



GRAPH A5: MEASUREMENTS OF THE SOCIAL AND EMOTIONAL CHARACTERISTICS OF THE STUDENTS IN ACCORDANCE WITH THEIR UNDERSTANDING OF THE ENCOURAGEMENT TO STUDY THEY RECEIVE FROM THEIR PARENTS



II c) N.B. Uses for which the SENNA is intended and the issue of private use by psychologists

The SENNA is, essentially, a children personality test. As such, it conforms to the definition of a psychology test outlined in CFP Resolution 002/2003: "systematic observation and registration procedures of samples of behavior and responses from individuals with the objective of describing and/or measuring psychological characteristics and processes, traditionally recognized in the areas of emotion/affection, cognition/intelligence, motivation, personality, psychomotor skills, attention, memory and perception, amongst others, in their most diverse forms of expression, according to the patterns defined by the construction of instruments."

In Brazil, Article 13 of Law nr. 4119/62, that regulates the profession of Psychology, states that the use of psychological methods and techniques is exclusive of that profession when it's use involves the following objectives: a) psychological diagnosis; b) professional orientation and selection; c) psycho-educational orientation; d) solution of adjustment problems. The law understands psychological tests as being "methods and techniques". As such, the law restricts the use of these methods to psychologists with the abovementioned intentions.

We would like to emphasize this aspect and clarify that the non-restricted availability of the SENNA respects these regulations, since it presents the instrument for use in the monitoring of education systems (use of the instrument's results to monitor psychological characteristics in the long-term). The SENNA is therefore published for use in situations where the object under analysis is the educational institutions and public policies, and its unit of analysis is the classes and schools, not the individual. Its use is therefore much wider and does not come into conflict (on the contrary, it is complimentary) with private use designed for psycho-educational orientation or psychological diagnosis where the object is the student and the unit of analysis personal characteristics and inclinations in terms of behavior, thoughts and emotions.

III. Socio-economic questionnaire

Student Questionnaire – 10th grade

The questionnaires for the 5th grade and the 12th grade differ from that presented below only in item 4, which is adapted to the specific age groups.

Dear Student, to get to know you better, we would like your help by filling in this questionnaire. Your replies are very important!

Instructions:
Choose just **ONE ANSWER PER QUESTION!**
There are no right or wrong answers. Select the option that is closest to your reality.

1. What sex are you?
A) Male.
B) Female.

2. What color do you consider yourself to be?
A) White.
B) Mixed.
C) Black.
D) Asian.
E) Indigenous.

3. In what month is your birthday?
A) January
B) February
C) March
D) April
E) May
F) June
G) July
H) August
I) September
J) October
K) November
L) December

4. How old are you?
A) Under 12 years old.
B) 13 years old.
C) 14 years old.
D) 15 years old
E) 16 years old.
F) 17 years old.
G) 18 years old.
H) Over 19 years old.

5. Do you live with your mother?
A) Yes.
B) No.
C) No. I live with another woman who is responsible for me.

6. Does your mother or the other woman who is responsible for you know how to read and write?
A) Yes.
B) No.
C) I don't know.

7. Through until what grade/year did your mother or the other woman who is responsible for your study?
A) She never studied or did not finish 5th grade.
B) She finished 5th year, but did not finish 9th grade.
C) She finished 9th grade, but did not graduate from High School.
D) She graduated High School but did not graduate Higher Education.
E) She graduated Higher Education.
F) I don't know.

8. Does your father or the man responsible for you know how to read and write?
A) Yes.
B) No.
C) I don't know.

Considering where you live, please answer these questions:	Yes	No
9. Is your road asphalted or paved?	(A)	(B)
10. Does your house have electricity?	(A)	(B)
11. Does your house have running water?	(A)	(B)
12. Is the garbage collected in your road?	(A)	(B)
13. Does anyone who lives with you receive the '*Bolsa Família*' ('Family Allowance')?	(A)	(B)
14. Do you have a maid or cleaner who works at your house?	(A)	(B)

How many of the following exist in the place where you live?	None	1	2	3 or+
15. Bathroom	(A)	(B)	(C)	(D)
16. Fridge with separate freezer	(A)	(B)	(C)	(D)
17. Washing machine (not a sink with a washboard)	(A)	(B)	(C)	(D)
18. DVD player	(A)	(B)	(C)	(D)
19. Automobile (car/motorbike)	(A)	(B)	(C)	(D)
20. Portuguese language dictionary (and/or other language dictionary)	(A)	(B)	(C)	(D)

21. Not including school books, newspapers and magazines, how many books are there in the place where you live?
A) There are no books in my house.
B) Enough to fill a shelf (1 to 20 books).
C) Enough to fill a bookcase (21 to 100 books).
D) Enough to fill a number of shelves (more than 100 books).

22. How often do you see your parents or those responsible for you reading (newspapers, magazines, books, etc.)?
A) Always.
B) Sometimes.
C) Rarely.
D) Never.

23. Do your parents or those responsible for you encourage you to read (newspapers, magazines, books, etc.)?
A) Yes.
B) No.

24. How often do you read (newspapers, magazines, books, etc.)?
A) Always.
B) Sometimes.
C) Rarely.
D) Never.

25. Do you have a computer where you live?
A) Yes, with internet access.
B) Yes, but without internet access.
C) No.