



**PISA 2018
DRAFT ANALYTICAL FRAMEWORKS
MAY 2016**

TABLE OF CONTENTS

I. PISA 2018 – DRAFT READING LITERACY FRAMEWORK	5
INTRODUCTION	5
A. Reading as the major domain	5
B. Changes in the nature of reading literacy	5
C. The continuity and change in the framework from 2000 to 2015	6
D. Revising the framework for PISA 2018	8
E. The importance of digital reading literacy	8
F. Reading motivation, practices and metacognition	10
G. The structure of the reading literacy framework	10
1. DEFINING READING LITERACY	10
2. ORGANISING THE DOMAIN	14
A. Processes	15
B. Texts	21
3. ASSESSING READING LITERACY	24
A. Scenarios	24
B. Tasks	25
C. Factors affecting item difficulty	26
D. Factors improving the coverage of the domain	28
E. Responses formats	31
F. Assessing the ease and efficiency of reading simple texts for understanding	32
G. Assessing students' reading motivation, reading practices and awareness of reading strategies	34
H. Considerations for adaptive testing	36
4. REPORTING PROFICIENCY IN READING	36
A. Reporting scales	36
B. Interpreting and using the scales	37
C. Defining levels of reading literacy proficiency	37
APPENDIX A. MAIN CHANGES IN THE READING FRAMEWORK, 2000-2015	40
APPENDIX B. SAMPLE TASKS	41
REFERENCES	43

II. FINANCIAL LITERACY FRAMEWORK	50
INTRODUCTION	50
A. Policy interest in financial literacy	50
B. Expected benefits of financial education and improved levels of financial literacy	52
C. The need for data	56
D. The measurement of financial literacy in PISA	57
1. DEFINING FINANCIAL LITERACY	58
2. ORGANISING THE DOMAIN	60
A. Content	61
B. Processes	70
C. Contexts	75
3. ASSESSING FINANCIAL LITERACY	80
A. The structure of the assessment	80
B. Response formats and coding	81
C. Distribution of score points	83
D. The impact of other domain knowledge and skills on financial literacy	84
4. REPORTING FINANCIAL LITERACY	86
NOTES	88
REFERENCES	89
III. PISA 2018 – DRAFT QUESTIONNAIRE FRAMEWORK	92
INTRODUCTION	92
A. Aims of the Programme for International Student Assessment	92
B. PISA 2018	93
C. Outline of the Framework	94
1. DEFINING THE QUESTIONNAIRE CORE IN PISA 2018	95
A. Outline of content covered in questionnaires: a) general framework of the 2018 study and modular approach, b) constructs to be covered, c) global competence	96
B. Previous use of PISA context data: Measures that have been and will be important for analysis and reporting	102

2. COVERAGE OF POLICY ISSUES FOR 2018	104
A. Assessing non-cognitive and metacognitive constructs	104
B. Assessing student background	112
C. Assessing teaching and learning processes	114
D. Assessing school policies and governance	117
E. Dealing with response bias	120
REFERENCES	123

IV. FRAMEWORK DEVELOPED FOR PISA 2018 : MEASURING ADOLESCENT WELL-BEING IN PISA **133**

EXECUTIVE SUMMARY	133
INTRODUCTION	134
1. WELL-BEING AS A MULTI-DIMENSIONAL CONSTRUCT	137
2. ADDRESSING MEASUREMENT CHALLENGES	140
A. Balance single-item measures with multi-item indices	140
B. Use meaningful number of scale points and clearly distinguishable scale labels	141
C. Select measures that maximize cross-cultural comparability	142
D. Consider item formats beyond traditional self-reports	142
E. Consider alternative questionnaire designs to reduce respondent burden	143
3. SUGGESTED QUALITY OF LIFE INDICATORS	144
A. Quality of Life as a Whole	144
B. Self-related well-being	146
C. School-environment: School-related well-being	150
D. Out-of school environment: Well-being outside of school	154
E. Possible composite indicators	158
REFERENCES	160
APPENDIX: DETAILED ITEM-LEVEL ASSESSMENT PLAN	171

BOXES

Box 1.	The definition of reading literacy in earlier PISA cycles	11
Box 2.	Characteristics used to classify texts in the PISA 2009 reading framework	21
Box 3.	Text availability and its impact on comprehension in the design of tasks	27
Box 4.	Categorisation of situations	28
Box 5.	Classification of texts	30
Box 6.	The status of writing skills in PISA 2018 reading literacy assessment	31
Box 7.	OECD activities in relation to financial education	54

FIGURES

Figure 1.	Reading Literacy Sources of Influence _____	14
Figure 2.	PISA 2018 Reading Framework Processes _____	17
Figure 3.	A model for organising the domain for an assessment framework _____	61
Figure 4.	Item for the unit <i>AT THE MARKET</i> _____	63
Figure 5.	Item for the unit <i>SPENDING CHOICES</i> _____	64
Figure 6.	Item for the unit <i>MOTORBIKE INSURANCE</i> _____	66
Figure 7.	Item for the unit <i>BANK ERROR</i> _____	69
Figure 8.	Item for the unit <i>PAY SLIP</i> _____	70
Figure 9.	Item for the unit <i>SHARES</i> _____	71
Figure 10.	Item for the unit <i>AT THE MARKET</i> _____	72
Figure 11.	Item for the unit <i>NEW OFFER</i> _____	73
Figure 12.	Item for the unit <i>TRAVEL MONEY</i> _____	74
Figure 13.	Item for the unit <i>NEW BANK CARD</i> _____	77
Figure 14.	Relationship between the content of financial literacy and mathematical literacy in PISA _____	85
Figure 15.	PISA 2018 Questionnaire Modules _____	96
Figure 16.	Framework Overview _____	133
Figure 17.	Proposed PISA 2018 Well-being Framework _____	138

TABLES

Table 1.	Mapping of 2018 process typology to 2018 reporting scales and to former 2009-2015 cognitive aspects _____	21
Table 2.	Approximate distribution of tasks per targeted process and text availability _____	26
Table 3.	Item difficulty for tasks _____	27
Table 4.	An overview of reading proficiency levels as they were described in the PISA 2012 study _____	38
Table 5.	Approximate distribution of score points in financial literacy, by content _____	83
Table 6.	Approximate distribution of score points in financial literacy, by processes _____	83
Table 7.	Approximate distribution of score points in financial literacy, by contexts _____	84

I. PISA 2018 – DRAFT READING LITERACY FRAMEWORK

INTRODUCTION

A. Reading as the major domain

PISA 2018 marks the third time reading is a major domain and the third time that the framework receives a major revision. Such a revision must reflect the changing definition of reading literacy as well as the contexts in which reading is used in citizens' lives. Thus, the present revision of the framework builds on contemporary and comprehensive theories of reading literacy as well as considers how students acquire and use information across broad contexts.

We live in a rapidly changing world, in which both the quantity and variety of written materials are increasing and where more and more people are expected to use these materials in new and increasingly complex ways. It is now generally accepted that our understanding of reading literacy evolves along with changes in society and culture. The reading literacy skills needed for individual growth, educational success, economic participation and citizenship 20 years ago were different from those of today; and it is likely that in 20 years' time they will change further still.

The goal of education has continued to shift its emphasis from the collection and memorisation of information only to the inclusion of a broader concept of knowledge: "whether a technician or a professional person, success lies in being able to communicate, share, and use information to solve complex problems, in being able to adapt and innovate in response to new demands and changing circumstances, in being able to marshal and expand the power of technology to create new knowledge and expand human capacity and productivity" (Binkley et al., 2010, p. 1). The ability to locate, access, understand and reflect on all kinds of information is essential if individuals are to be able to participate fully in our knowledge-based society. Achievement in reading literacy is not only a foundation for achievement in other subject areas within the educational system, but also a prerequisite for successful participation in most areas of adult life (Cunningham & Stanovich, 1998; OECD, 2013a; Smith, Mikulecky, Kibby, & Dreher, 2000). The PISA framework for assessing the reading literacy of students towards the end of compulsory education, therefore, must focus on reading literacy skills that include finding, selecting, interpreting, integrating and evaluating information from the full range of texts associated with situations that reach beyond the classroom.

B. Changes in the nature of reading literacy

Evolving technologies have rapidly changed the ways in which people read and exchange information, both at home and in the workplace. Automation of routine jobs creates a demand for people who can adapt to quickly changing contexts and who can find and learn from diverse information sources. In 1997 when the first PISA framework for reading was starting to be discussed, just 1.7% of the world's population used the Internet. By 2014 the number had grown to a global penetration rate of 40.4%, representing almost three billion people (ITU, 2014a). Between 2007 and 2013, mobile phone subscriptions doubled: in 2013, there were almost as many active subscriptions as people on earth (95.5 subscriptions per 100 people) and mobile

broadband has increased to almost two billion subscriptions worldwide (ITU, 2014b). The Internet increasingly pervades the life of all citizens, from learning in and out of school, to working from real or virtual workplaces, to dealing with personal matters such as taxes, health care or holidays. As personal and professional development is becoming a lifelong undertaking, the students of tomorrow will need to be skilled with digital tools in order to succeed with the increased complexity and quantity of information available.

In the past, the primary and predominant interest in student reading literacy proficiency was the ability to understand, interpret and reflect upon single texts. While these skills remain important, greater emphasis on the integration of information technologies into citizens' social and work lives requires that the definition of reading literacy be updated and extended. It must reflect the broad range of newer skills associated with literacy tasks required in the 21st century (e.g. Ananiadou & Claro, 2009; Kirsch et al., 2002; Rouet, 2006; Spiro et al., 2015). This necessitates an expanded definition of reading literacy encompassing both the basic reading processes and higher-level digital reading skills while recognising that literacy will continue to change due to the influence of new technologies and changing social contexts (Leu et al., 2013, 2015).

As the medium through which we access textual information is moving from print to computer screens to smart phones, the structure and formats of texts have changed. This in turn requires readers to develop new cognitive strategies and clearer goals in purposeful reading. Therefore, success in reading literacy should no longer be defined by just being able to read and comprehend a single text. Although the ability to comprehend and interpret extended pieces of continuous texts - including literary texts - remains a valuable one, success will also come through deploying complex information-processing strategies, including analysing, synthesising, integrating and interpreting relevant information from multiple text (or information) sources. In addition, successful and productive citizens will need to use the information from across domains, such as science and mathematics, and employ technologies to effectively search, organise and filter a wealth of information. These will be the key skills, which are necessary for full participation in the labour market, in additional education as well as in social and civic life in the 21st Century (OECD, 2013b).

C. The continuity and change in the framework from 2000 to 2015

With the changes in the nature of reading literacy, the framework also has changed. Reading literacy was the major domain assessed in 2000 for the first PISA cycle (PISA 2000). For the fourth PISA cycle (PISA 2009), it was the first to be revisited as a major domain, requiring a full review of its framework and new development of the instruments that represent it. For the seventh PISA cycle (2018), it is again being revised.

The original reading literacy framework for PISA was developed for the PISA 2000 cycle (from 1998 to 2001) through a consensus building process involving reading experts selected by the participating countries to form the PISA 2000 reading expert group (REG). The definition of reading literacy evolved in part from the IEA Reading Literacy Study (1992) and the International Adult Literacy Survey (IALS, 1994, 1997 and 1998). In particular, it reflected the IALS emphasis on the importance of reading skills for active participation in society. It was also influenced by contemporary – and still current – theories of reading, which emphasise the multiple linguistic-cognitive processes involved in reading and their interactive nature (Britt, Goldman, & Rouet, 2012; Kamil, Mosenthal, Pearson, & Barr, 2000; Perfetti, 1985, 2007; Rayner & Reichle, 2010; Snow, 2002), models of discourse comprehension (Kintsch, 1998; Zwaan & Singer, 2003) and theories of performance in solving information problems (Kirsch, 2001; Kirsch & Mosenthal, 1990; Rouet, 2006).

Much of the substance of the PISA 2000 framework was retained in the PISA 2009 framework, respecting one of the central purposes of PISA: to collect and report trend information about performance in reading, mathematics and science. However, the PISA domain frameworks are designed to be evolving documents that will adapt to and integrate new developments in theory and practice over time. Thus, there has been an evolution, reflecting both an expansion in our understanding of the nature of reading and changes in the world. This evolution is shown in greater detail in Appendix A, which provides an overview of the primary changes in the reading framework from 2000 to 2015.

Changes in our concept of reading since 2000 have led to an expanded definition of reading literacy, which recognises motivational and behavioural characteristics of reading alongside cognitive characteristics. Both reading engagement and metacognition – an awareness and understanding of how one develops an understanding of text and uses reading strategies – were referred to briefly at the end of the first PISA framework for reading under “Other issues” (OECD, 2000). In the light of recent research, reading engagement and metacognition were featured more prominently in the PISA 2009 and 2015 reading frameworks as elements that can be developed, shaped and fostered as components of reading literacy.

A second major modification of the framework for PISA 2009 involved the inclusion of electronic texts in recognition of the increasing role digital texts play in the literacy skills needed for individual growth and active participation in society (OECD, 2011). This modification was also specifically developed for presentation of items on a computer screen. PISA 2009 was the first large-scale international study to assess electronic reading. Due to the rapidly evolving technologies and related practices, this initiative, which is grounded in current theory and best practices from around the world, was inevitably a first step.

For the 2015 cycle, reading was a minor domain and continued the description and illustration of reading literacy developed for PISA 2009. However, the 2015 cycle involved important changes in the test administration procedures, some of which required adjustments in the wording of the reading framework. For example, the reading assessment in the 2015 cycle was administered primarily on computer. As a result, the “environment” and “medium” dimensions were revisited and further elaborated with the inclusion of the terms “fixed” and “dynamic”.

D. Revising the framework for PISA 2018

The revisions to the reading literacy framework retain aspects of the 2009/2015 frameworks that are still relevant to PISA 2018. However, the framework is enhanced and revised in the following ways:

- The framework fully integrates reading in a traditional sense together with the new forms of reading that have emerged over the past decades and continue to emerge due to the spread of digital devices and digital texts.
- The framework incorporates constructs involved in basic reading processes. These constructs, such as fluent reading, literal interpretation, inter-sentence integration, extracting the central themes and inferencing, are critical skills for processing complex or multiple texts for specific purposes. If students fail at performing higher-level text processing functions, it is critical to know whether it was due to difficulties in these basic skills in order to provide targeted support to student populations within educational systems.
- The framework revisits the way in which the domain is organised to incorporate reading processes such as evaluating the veracity of texts, information seeking, reading from multiple sources and the integration/synthesis of information across sources. The revision rebalances the prominence of different reading processes to reflect the global importance of the different constructs, while ensuring there is a link to the prior frameworks in order to maintain trend.
- The revision considers how new technology options and the use of scenarios involving print and digital text can be harnessed to achieve a more authentic assessment of reading, consistent with the current use of texts around the world.

E. The importance of digital reading literacy

Reading in today's world is very different from just 20 years ago. Up to the mid-1990s, reading was mostly performed with paper. Printed matter existed in many different forms, shapes and texture, from children books to lengthy novels, from leaflets to encyclopaedia, from newspapers and magazines to scholarly journals, from administrative forms to notes on billboards.

In the early 1990s, a small percentage of people owned computers and most of those owned were mainframes or desktop PCs. Very few people owned laptops for their personal use, whereas digital tablets and smartphones were still mostly fiction. Computer-based reading was limited to specific types of users and uses, typically a specialised worker dealing with technical or scientific information. In addition, due to mediocre display quality, computer-based reading was slower, more error-prone and more tiring than reading on paper (Dillon, 1994). Initially acclaimed as a means to "free" the reader from the printed text "straightjacket", the emerging hypertext technology [(i.e. the linking of digital information pages allowing each reader to dynamically construct their own route through information chunks (Conklin, 1988))] was also generating syndromes of disorientation and cognitive overhead, as design of the Web was still in its infancy (Foltz, 1996; Nielsen, 1999; Rouet & Levonen, 1996). But then, only a very small fraction of the world population had access to the newly-born World Wide Web.

In less than 20 years, the number of computers in use worldwide grew to an estimated 2 billion in 2015 (ITU, 2014b). In 2013, 40% of the world's population had access to the Internet at home, with sharp contrasts between developed countries, where access reached 80% of the population, and some less developed countries; where access lagged below 20% (ITU, 2014b). The last decade has witnessed a dramatic expansion of portable digital devices, with wireless Internet access overtaking fixed broadband subscriptions in 2009 (OECD, 2012). By 2015, computer sales were slowing, whereas digital pads, readers and cell phones still grew at two-digit rates (Gartner, 2015).

As a notable consequence of the spread of information and communication technology (ICT) in the general public, reading is massively shifting from print to digital texts. For example, computers have become the second source of news for American citizens, after TV and before radio and printed newspapers and magazines (American Press Institute, 2014). Similarly, British children and teenagers prefer to read digital than printed texts (Clark, 2014), and a recent UNESCO report showed that two thirds of users of a phone-based reader across five developing countries indicated that their interest in reading and time spent reading increased once it was possible to read on their phones (UNESCO, 2014). This shift has important consequences for the definition of reading as a skill. Firstly, the texts that people read on line are quite different from traditional printed texts. In order to enjoy the wealth of information, communication and other services offered through digital devices, online readers have to cope with smaller displays, cluttered screens and challenging networks of pages. In addition, new genres of print-based communication have appeared, such as email, short messaging, forums and social networking applications. It is important to stress that the rise of digital technology means that people need to be selective in what they read while they must also read more, more often and for a broader range of purposes. Reading and writing are even replacing speech in some essential communication acts, such as telephoning and help desks. A consequence is that readers have to understand these new text-based genres and social-cultural practices.

Readers in the digital age also have to master several new skills. They have to be minimally ICT literate in order to understand and operate the devices and applications. They also have to search and access the texts they need to read through the use of search engines, menus, links, tabs and other paging and scrolling functions. Due to the uncontrolled profusion of information on the Internet, readers also have to be discerning in their choice of information sources and assessment of information quality and credibility. Finally, readers have to read across texts to corroborate information, to detect potential discrepancies and conflicts and to resolve them. The importance of these new skills was clearly illustrated in the OECD's PISA 2009 digital reading study, whose report noted the following:

Navigation is a key component of digital reading, as readers "construct" their text through navigation. Thus, navigational choices directly influence what kind of text is eventually processed. Stronger readers tend to choose strategies that are suited to the demands of the individual tasks. Better readers tend to minimise their visits to irrelevant pages and locate necessary pages efficiently. (OECD, 2011, p. 20)

In addition, a 2015 study of student use of computers in the classroom (OECD, 2015) shows for instance that "students' average navigation behaviour explains a significant part of the differences in digital reading performance between countries/economies that is not accounted for by differences in print-reading performance" (p. 119), (see also Nauman, 2015).

Thus, in many parts of the world skilful digital reading literacy is now key to one's ability to achieve one's goals and participate in society. The 2018 PISA reading framework is revised and expanded so as to encompass those skills that are essential for reading and interacting with digital texts.

F. Reading motivation, practices and metacognition

Individuals' reading practices, motivation and attitudes towards reading, as well as an awareness of how effective reading strategies are, play a prominent role in reading. Students who read more frequently, be it with print or on-screen, who are interested in reading, who feel themselves confident in their reading abilities and who know well which strategies to use, for instance, to summarise a text or search information on Internet, tend to be more proficient in reading. Moreover, if practices, motivation, and metacognition deserve close attention, it is not only because they are potential predictors of reading achievement and growth, it is also because they can be considered important goals or outcomes of education, potentially driving life-long learning. Furthermore, they are malleable variables, amenable to change. For instance, there is strong evidence that reading engagement and metacognition (awareness of strategies) can be enhanced through teaching and supportive classroom practices (Brozo & Simpson, 2007; Guthrie, Wigfield, & You, 2012; Guthrie, Ho, & Klauda, 2013; Reeve, 2012). Reading motivation, practices and metacognition are briefly discussed in the reading literacy framework since they are critical factors of reading, although they are assessed in the questionnaire and are covered in more detail in the questionnaire framework.

G. The structure of the reading literacy framework

Having addressed what is meant by the term "reading literacy" in PISA and introduced the importance of reading literacy in today's society in this introduction, the remainder of the framework is organised as follows. The second section defines reading literacy and elaborates on various phrases that are used in the reading framework, along with the assumptions underlying the use of these words. The third section focuses on the organisation of the domain of reading literacy and discusses the characteristics that will be represented in the tasks included in the PISA 2018 assessment. The fourth section discusses some of the operational aspects of the assessment and how it will be measured as well as presenting sample items. Finally, the last section describes how the reading literacy data will be summarised and outlines plans for reporting.

1. DEFINING READING LITERACY

Definitions of reading and reading literacy have changed over time to reflect changes in society, economy, culture and technology. Reading is no longer considered an ability acquired only in childhood during the early years of schooling. Instead it is viewed as an expanding set of knowledge, skills and strategies that individuals build on throughout life in various contexts, through interaction with their peers and the wider community. Thus, reading must be considered across the varied ways in which citizens interact with text-based artefacts and how reading is part of life-long learning.

Cognitively-based theories of reading emphasise the constructive nature of comprehension, the diversity of cognitive processes involved in reading and their interactive nature (Binkley, Rust, & Williams 1997; Kintsch, 1998; McNamara & Magliano, 2009; Oakhill, Cain, & Bryant, 2003; Snow and the RAND Reading Group, 2002; Zwaan & Singer, 2003). The reader generates meaning in response to text by using previous knowledge and a range of text and situational cues that are often socially and culturally derived. While constructing meaning, competent readers use various processes, skills and strategies to locate information, to monitor and maintain understanding (van den Broek, Ridsen, & Husbye-Hartmann, 1995) and to critically assess the relevance and validity of the information (Richter & Rapp, 2014). These processes and strategies are expected to vary with context and purpose as readers interact with multiple continuous and non-continuous texts both in print and when using digital technologies (Britt & Rouet, 2012; Coiro, Knobel, Lankshear, & Leu, 2008).

Box 1. The definition of reading literacy in earlier PISA cycles

The PISA 2000 definition of reading literacy was as follows:

Reading literacy is understanding, using and reflecting on written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.

The PISA 2009 definition of reading, continued for 2012 and 2015, added engagement in reading as part of reading literacy:

Reading literacy is understanding, using, reflecting on and engaging with written texts, in order to achieve one's goals, to develop one's knowledge and potential, and to participate in society.

For 2018 the definition of reading literacy adds in evaluation of texts as an integral part of reading literacy and removes the word "written".

The 2018 Definition of Reading Literacy

Reading literacy is understanding, using, evaluating, reflecting on and engaging with texts in order to achieve one's goals, to develop one's knowledge and potential and to participate in society.

Each part of the definition is considered in turn below, taking into account the original elaboration and some important developments in the definition of the domain that uses evidence from PISA and other empirical studies, from theoretical advances and from the changing nature of the world.

- Reading literacy . . .

The term "reading literacy" is used instead of the term "reading" because it is likely to convey to a non-expert audience more precisely what the survey is measuring. "Reading" is often understood as simply decoding, or even reading aloud, whereas the intention of this survey is to measure much broader and more encompassing constructs. Reading literacy includes a wide range of cognitive and linguistic competencies, from basic decoding to knowledge of words, grammar and larger linguistic and textual structures for comprehension, as well as integration of meaning with one's knowledge about the world. It also includes metacognitive competencies: the awareness of and ability to use a variety of appropriate strategies when processing texts. Metacognitive competencies are activated when readers think about, monitor and adjust their reading activity for a particular goal.

The term “literacy” typically refers to an individual’s knowledge of a subject or field, although it has been most closely associated with an individual’s ability to learn, use and communicate written and printed information. This definition seems close to the notion that the term “reading literacy” is intended to express in this framework: the active, purposeful and functional application of reading in a range of situations and for various purposes. PISA assesses a wide range of students. Some of these students will go on to a university, possibly to pursue an academic or professional career; some will pursue further studies in preparation for joining the labour force; and some will enter the workforce directly upon completion of secondary schooling. Regardless of their academic or labour-force aspirations, reading literacy will be important to their active participation in their community and economic and personal life.

- . . . *is understanding, using, evaluating, reflecting on* . . .

The word “understanding” is readily connected with the widely accepted concept of “reading comprehension”, that all reading involves some level of integrating information from the text with the reader’s knowledge structures. Even at the earliest stages, readers draw on symbolic knowledge to decode a text and require a knowledge of vocabulary to make meaning. However, this process of integration can also be much broader, such as developing mental models of how texts relate to the world. The word “using” refers to the notions of application and function – doing something with what we read. The term “evaluating” was added for PISA 2018 to incorporate the notion that reading is often goal-directed, and consequently the reader must weigh such factors as the veracity of the arguments in the text, the point of view of the author and the relevance of a text to the reader’s goals. “Reflecting on” is added to “understanding”, “using” and “evaluating” to emphasise the notion that reading is interactive: readers draw on their own thoughts and experiences when engaging with a text. Every act of reading requires some reflection, reviewing and relating of information within the text with information from outside the text. As readers develop their stores of information, experience and beliefs, they constantly test what they read against outside knowledge, thereby continually reviewing and revising their sense of the text. This evaluation can include determining the veracity of a text, checking the claims made by the author as well as inferring the author’s perspective. At the same time, incrementally and perhaps imperceptibly, readers’ reflections on texts may alter their sense of the world. Reflection might also require readers to consider the content of the text, apply their previous knowledge or understanding or think about the structure or form of the text. Each of these skills in the definition, “understanding”, “using”, “evaluating” and “reflecting on” are necessary, but none are sufficient for successful reading literacy.

- . . . *and engaging with* . . .

A reading literate person not only has the skills and knowledge to read well, but also values and uses reading for a variety of purposes. It is therefore a goal of education to cultivate not only proficiency but also engagement in reading. Engagement in this context implies the motivation to read and comprises a cluster of affective and behavioural characteristics that include an interest in and enjoyment of reading, a sense of control over what one reads, involvement in the social dimension of reading and diverse and frequent reading practices.

- . . . *texts* . . .

The phrase “texts” is meant to include all language as used in its graphic form: handwritten, printed or screen-based. In this definition, we exclude as texts those purely aural language artefacts such as voice recordings, as well as film, TV, animated visuals and pictures without words. Texts do include visual displays such as diagrams, pictures, maps, tables, graphs and comic strips, which include some written language (for example, captions). These visual texts can exist either independently or they can be embedded in larger texts.

Dynamic texts are distinguishable from fixed texts in a number of respects, including how they affect the ability to estimate the length and quantity of texts when physical cues (e.g. dimension of paper-based document are hidden in virtual space); the way different parts of a text and different texts are connected with one another through hypertext links; whether multiple abstracted texts are shown as a result of a search; and consequent upon all these text characteristics, the way that readers typically engage with dynamic texts. To a much greater extent than with what is printed, readers need to construct their own pathways to complete any reading activity associated with dynamic texts.

The term “texts” was chosen instead of the term “information” because of its association with written language and because it more readily connotes literary as well as information-focused reading.

- . . . *in order to achieve one’s goals, to develop one’s knowledge and potential and to participate in society.*

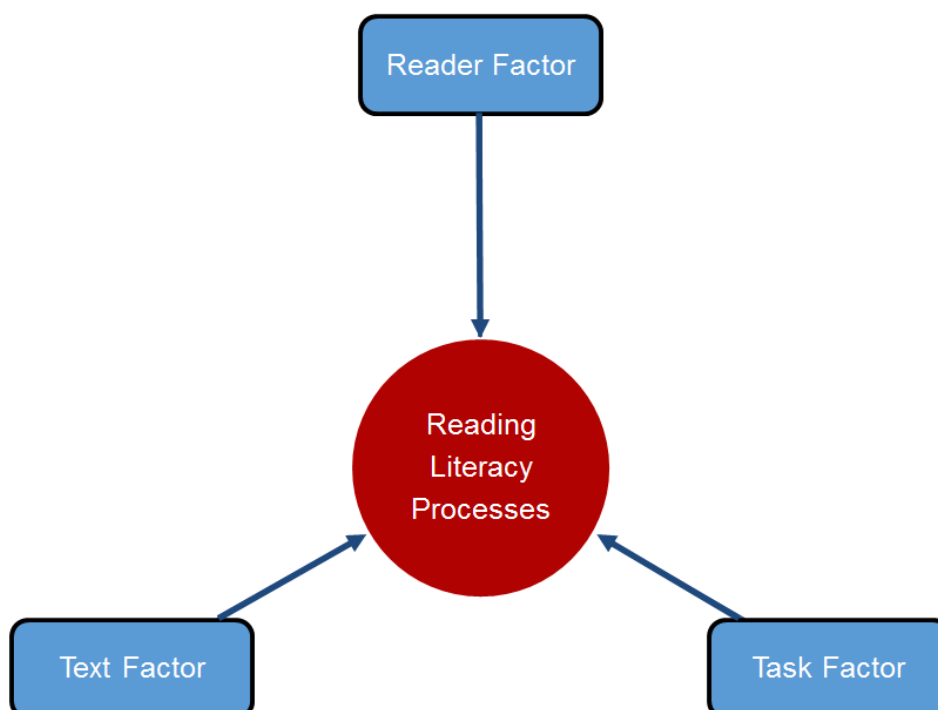
This phrase is meant to capture the full scope of situations in which reading literacy plays a role, from private to public, from school to work, from formal education to lifelong learning and active citizenship. “To achieve one’s goals and to develop one’s knowledge and potential” spells out the long-held idea that reading literacy enables the fulfilment of individual aspirations – both defined ones such as graduating or getting a job, and those less defined and less immediate that enrich and extend personal life and lifelong education (Gray & Rogers, 1956). The PISA concept of reading literacy also embraces the new challenges of reading in the 21st century. It conceives of reading literacy as the foundation for full participation in the economic, political, communal and cultural life of contemporary society. The word “participate” is used because it implies that reading literacy allows people to contribute to society as well as to meet their own needs: “participating” includes social, cultural and political engagement (Hofstetter, Sticht, & Hoffstetter, 1999). For instance, literate people have greater access to employment and more positive attitudes toward institutions (OECD, 2013). Higher levels of reading literacy have been found to be related to better health and reduced crime (Morrisroe, 2014). Participation may also include a critical stance, a step toward personal liberation, emancipation and empowerment (Lundberg, 1991).

2. ORGANISING THE DOMAIN

Reading as it occurs in everyday life is a pervasive and highly diverse activity. In order to design an assessment that adequately represents the many facets of reading literacy, the domain is organized according to a set of dimensions. The dimensions will in turn determine the test design and, ultimately, the evidence about student proficiencies that can be collected and reported.

Snow and the RAND group's (2002) influential framework defined reading comprehension as the joint outcome of three combined sources of influence: the reader, the text and the activity, task or purpose for reading. Reader, text and task dimensions interact within a broad sociocultural context, which can be thought of as the diverse range of situations in which reading occurs. For the purpose of PISA, we adopt a similar view of the dimensions of reading literacy. Figure 1 illustrates these dimensions. A reader brings a number of reader factors to reading, which can include motivation, prior knowledge, and other cognitive abilities. The reading activity is a function of text factors (i.e. the text or texts that are available to the reader at a given place and time). These factors can include the format of the text, the complexity of the language used, the number of texts a reader encounters, as well as others. Reading activity is also a function of task factors (i.e. the requirements or reasons that motivate the reader's engagement with text). Task factors also include the potential time and other practical constraints, the goals of the task (e.g. read for pleasure, read for deep understanding or skim) and the complexity or number of tasks to be completed. Based on their individual characteristics and their perception of text and task dimensions, readers apply a set of reading literacy processes in order to locate, extract information and construct meaning from texts to achieve the tasks.

Figure 1. Reading Literacy Sources of Influence



For the purpose of PISA reading literacy, the goal of the cognitive instrument is to measure students' mastery of reading literacy processes through manipulating task and text factors. The questionnaire further serves to assay some of the reader factors, such as motivation, disposition and experience.

In designing the PISA reading literacy assessment, the two most important considerations are, first, to ensure broad coverage of what students read and for what purposes they read, both in and outside of school, and, second, to represent a natural range of difficulty in texts and tasks. The PISA reading literacy assessment is built on three major characteristics: text – the range of material that is read; processes – the cognitive approach that determines how readers engage with a text; and scenarios – the range of broad contexts or purposes for which reading takes place with one or more thematically related texts. Within scenarios are tasks – the assigned goals that readers must achieve in order to succeed. All three contribute to ensuring broad coverage of the domain. In PISA, difficulty of tasks can be varied by manipulating text features and task goals, which then require deployment of different cognitive processes. Thus, the PISA reading literacy assessment aims at measuring students' mastery of reading processes (the possible cognitive approaches of readers to a text) by varying the dimensions of text (the range of material that is read) and scenarios (the range of broad contexts or purposes for which reading takes place) with one or more thematically related texts. While there may be individual differences in reader factors based on the skills and background of each reader, these are not manipulated in the cognitive instrument, but are captured through the assessment in the questionnaire.

In order to use these three characteristics in designing the assessment, they must be operationalised. That is, the various values that each of these characteristics can take on must be specified. This allows test developers to categorise the materials they are working with and the tasks they construct so that they can then be used to organise the reporting of the data and to interpret results.

A. Processes

The PISA typology of cognitive aspects involved in reading literacy was designed at the turn of the 21st Century (OECD, 2000). A revision of the “aspects” in the 2018 PISA reading literacy framework is needed for at least three reasons:

- a. A definition of reading literacy must reflect contemporary developments in school and societal literacy demands, namely, the increasing amount of text information available in print and digital forms and the increasing diversity and complexity of situations involving texts and reading. These evolutions are partly driven by the spread of digital information technology and in particular by increased access to the Internet worldwide.
- b. The PISA 2018 framework should also reflect recent developments in the scientific conceptualisation of reading and be as consistent as possible with the terminology used in current theories. There is a need to update the vocabulary that was used to designate the cognitive processes involved in reading, taking into account progress in the research literature.

- c. Finally a revision is needed to reassess the necessary trade-off between the precision of the aspects as described in the framework and the limited possibility to account for each of these individual aspects in a large-scale international assessment. Such a reassessment is particularly relevant in the context of PISA 2018 in which reading literacy is the main domain.

The 2018 framework replaces the phrase “cognitive aspects”, used in previous versions of the framework, with the phrase “cognitive processes”. The phrase “cognitive processes” aligns with the terminology used in reading psychology research and is more consistent with a description of reader skills and proficiencies. The term “aspects” tended to confound the reader's actual cognitive processes with the requirements of various types of tasks (e.g. demands of specific types of questions). A description of proficient reading processes permits the 2018 framework to map these processes to a typology of tasks.

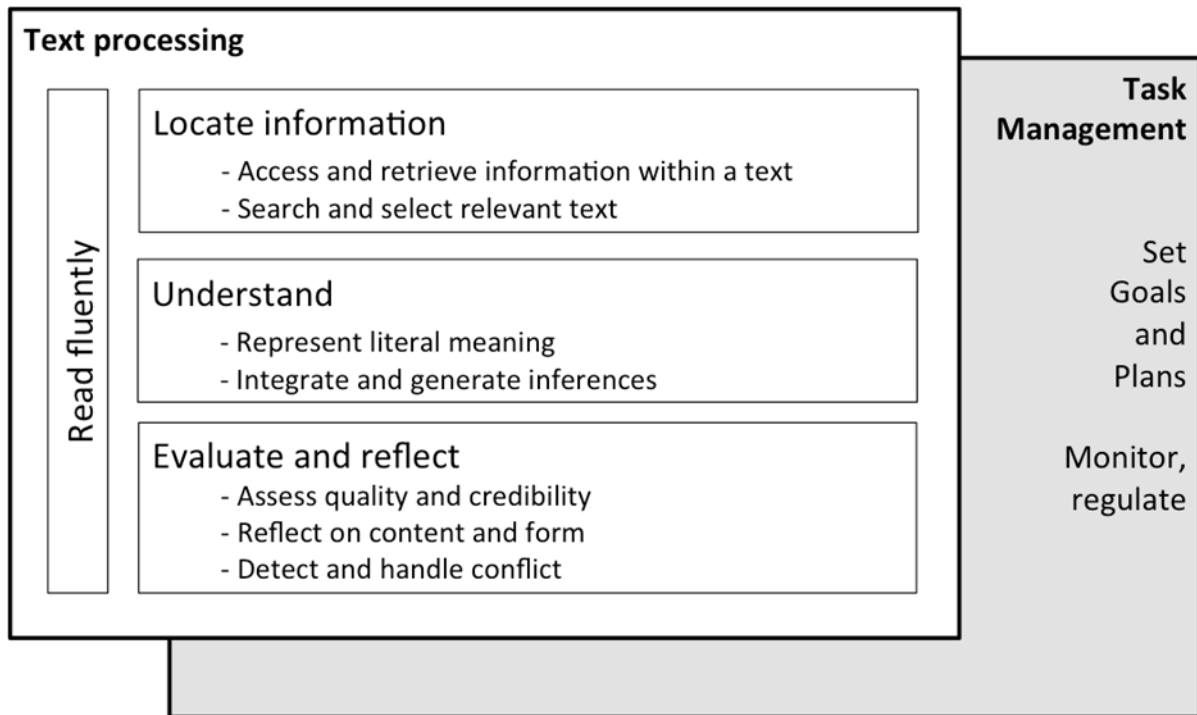
Recent theories of reading literacy emphasise the fact that "reading does not take place in a vacuum" (Snow and the RAND Reading Group, 2002; see also McCrudden & Schraw, 2007; Rouet & Britt, 2011). Indeed, most reading activities in people's daily lives are motivated by specific purposes and goals (White, Chen & Forsyth, 2010). Reading as a cognitive skill involves a set of specific reading processes that competent readers make use of when engaging with texts in order to achieve their goals. Goal setting and goal achievement drive not only readers' decisions to engage with texts, their selection of texts and passages of text, but also their decisions to disengage from a particular text, to reengage with a different text, to compare and to integrate information across multiple texts (Britt & Rouet, 2012; Goldman, 2004; Perfetti, Rouet, & Britt, 1999).

To achieve reading literacy as it is defined in this framework, an individual needs to be able to execute a wide range of processes. Effective execution of these processes, in turn, requires that the reader have the cognitive skills, strategies and motivation that support the processes.

The PISA 2018 reading framework acknowledges the goal-driven, critical and intertextual nature of reading literacy (McCrudden & Schraw, 2007; Rouet, 2006; Vidal-Abarca, Mañá, & Gil, 2010). Consequently, the former typology of reading aspects (OECD, 2000) is revised and extended so as to explicitly represent the fuller range of processes that skilled readers selectively draw from as a function of their particular task context and information environment.

More specifically, two broad categories of reading processes are defined for PISA 2018: text processing and task management (Figure 2). This distinction is consistent with current views of reading as a situated and purposeful activity (see e.g. Snow and the Rand Reading Group., 2002). The focus of the cognitive assessment is on processes identified in the text processing box.

Figure 2. PISA 2018 Reading Framework Processes



a. Text processing

The 2018 typology of reading process specifically identifies the process of reading fluently as distinct from other processes associated with text comprehension.

- Read Fluently

Reading fluency can be defined as an individual's ability to read words and connected text accurately and automatically and to phrase and process these words and texts in order to comprehend the overall meaning of the text (e.g. Kuhn & Stahl, 2003). In other words, fluency is the ease and efficiency of reading texts for understanding. There is considerable empirical evidence demonstrating a link between reading ease/efficiency/fluency to reading comprehension (Chard, Pikulski, & McDonagh, 2006; Jenkins et al., 2003 b; Kuhn; Wagner et al; Wayman et al., 2007; Woodcock, Mather, & McGrew, 2001). The chief psychological mechanism proposed to explain this relationship is that the ease and efficiency of reading text is indicative of expertise in foundational reading skills of decoding, word recognition and syntactic parsing of texts.

Fluent reading frees up attention and memory resources, which can be allocated to higher-level comprehension processes. Conversely, weaknesses in reading fluency divert resources from comprehension towards lower level processes necessary to process the printed text, resulting in weaker performance in reading comprehension (e.g. Cain & Oakhill, 2007; Perfetti, Marron, & Foltz, 1996). Acknowledging this strong link between fluency and comprehension, the National Reading Panel (2000) in the United States recommended fostering fluency in reading to enhance students' comprehension skills.

- *Locate information*

Competent readers can read a text entirely and carefully in order to comprehend the main ideas and reflect on the text as a whole. On a daily basis, however, readers most often use texts for purposes that require the location of specific information, with little or no consideration for the rest of the text (White et al., 2010). Furthermore, locating information is becoming a mandatory aspect of reading when people interact with complex digital information systems such as search engines and websites (Brand-Gruwel, Wopereis, Vermetten, 2005; Leu et al., 2013). The 2018 framework defines two processes whereby readers perform the selection of information within and across texts:

Access and retrieve information within a text. Locating information from tables, text chapters or whole books is a skill in and by itself (Dreher & Guthrie, 1990; Moore, 1995; Rouet & Coutelet, 2008). Locating information draws on readers' understanding of the task demands, their knowledge of text organisers and their ability to assess the relevance of text. The ability to locate information is grounded on readers' strategic awareness of their information needs and their capacity to quickly disengage from irrelevant passages (McCrudden & Schraw, 2007). In addition, readers sometimes have to skim through a series of paragraph to retrieve specific pieces of information. This requires an ability to modulate one's reading speed, depth of processing and consideration versus dismissal of the information (Duggan & Payne, 2009). In the context of PISA 2018, access and retrieve tasks require the reader to scan a single text in order to retrieve target information made of a few words, phrases or numerical values. There is little or no need to comprehend the text beyond the phrase level. The identification of target information is achieved through literal or close to literal matching of elements in the question and in the text.

Search and select relevant text. Proficient readers are able to select information from not just one, but also from several texts. In electronic environments, the amount of available information often largely exceeds the amount readers are able to actually process. In these multiple-text reading situations, readers have to make decisions as to which of the available texts is the most important, relevant, accurate or truthful (Rouet & Britt, 2011). These decisions are based on readers' assessment of the texts' qualities from partial and sometimes opaque indicators, such as the information contained in a web link. (Gerjets, Kammerer, & Wermer, 2011; Mason, Boldrin, & Ariasi, 2010; Naumann, 2015; Rieh, 2002). Thus, one's ability to search and select a text within a set is an integral component of reading literacy. In PISA 2018, text search and selection tasks involve the use of text descriptors such as headers, source information (e.g. author, medium, date), and embedded or explicit links such as search engine result pages.

- *Understand*

A large number of reading activities involve the parsing and integration of extended passages of text in order to form an understanding of the meaning conveyed in the passage. Text understanding (also called comprehension) may be seen as the construction by the reader of a mental representation of what the text is about, or "situation model" (Kintsch, 1998). A situation model is based on two core processes: the construction of a memory representation of the literal meaning of the text; and the integration of literal text contents with one's prior knowledge through mapping and inference processes (McNamara & Magliano, 2009; Zwaan & Singer, 2003).

Acquiring a literal meaning representation requires readers to comprehend sentences or short passages. Literal comprehension tasks involve a direct or paraphrase type of match between the question and target information within a passage. The reader may need to hierarchise or condense information at a local level (Note: tasks requiring integration at the level of an entire passage, such as identifying the main idea, summarizing, or giving a title, are considered integration; see below).

Constructing an integrated text representation involves materials ranging from a sentence to an entire passage. The reader needs to generate various types of inferences, ranging from simple connecting inferences (such as the resolution of anaphora) to more complex coherence relationships (e.g. spatial, temporal, causal or claim-argument links). Sometimes the inference involves several portions of the text; in other cases the inference is needed to connect the question and the passage. Finally, the production of inferences is also needed in tasks requesting the reader to identify an implicit main idea, in order to produce a summary or title for a given passage.

When readers are faced with more than one text, integration and inference generation may be performed based on pieces of information located in different texts (Perfetti, Rouet, & Britt, 1999). Integration of information across texts poses a specific problem when the texts provide inconsistent or conflicting information. In those cases readers must engage in evaluation processes in order to acknowledge and handle the conflict (Bråten, Strømsø, & Britt, 2009; Stadtler & Bromme, 2014; see below).

- Evaluate and reflect

Competent readers can reason beyond the literal or inferential meaning of the text. They can reflect on the content and form of the text and critically assess the quality and validity of information.

Assess quality and credibility. Competent readers can evaluate the quality and credibility of the text (e.g. whether the information is valid, up to date, accurate, unbiased). Proficient evaluation sometimes requires the reader to identify and assess the source of the information: whether the author is competent, well-informed and benevolent, the reader must be able to reflect critically on the content and form of the text. Evaluation and reflection were arguably always part of reading literacy, but their importance has increased with the increased amount and heterogeneity of information readers are faced with today.

Reflect on content and form. Competent readers must also be able to reflect on the quality and style of the writing. This reflection involves being able to evaluate the form of the writing and how the content and form together relate to, and effectively express, the author's purposes and point of view. Reflecting also involves drawing upon one's knowledge, opinions or attitudes beyond the text in order to relate the information provided within the text to one's own conceptual and experiential frames of reference. Reflect items may be thought of as those that require readers to consult their own experience or knowledge to compare, contrast or hypothesise different perspectives or viewpoints.

Detect and handle conflict. When facing multiple texts that contradict each other, readers need to become aware of the conflict and to find ways to deal with it (Britt & Rouet, 2012; Stadler & Bromme, 2013; 2014). Handling conflict typically requires readers to assign discrepant claims to their respective sources and to assess the soundness of the claims and/or the credibility of the sources. As these skills underlie much of contemporary reading, it is an issue of critical importance to measure the extent 15-year-olds can meet the new challenges of comprehending, comparing and integrating multiple texts (Bråten et al., 2011; Coiro et al., 2008; Goldman, 2004; Leu et al., 2015; Mason et al., 2010; Rouet & Britt, 2014).

b. Task management processes

In the context of any assessment, but also in many everyday reading situations (White et al., 2010), readers engage with texts because they receive some kind of assignment or external prompt to do so. Reading literacy involves one's ability to accurately represent the reading demands of a situation, to set up task-relevant reading goals and to monitor progress toward these goals throughout the activity. Task management processes to accomplish a reader's goals include the setting, self-monitoring and self-regulation of goals and strategies (see e.g. Hacker, 1998; Winne & Hadwin, 1998, for discussions of self-regulated reading).

Task-oriented goals fuel the reader's search for task-relevant texts and/or passages within a text (e.g. McCrudden & Schraw, 2007; Rouet & Britt, 2011; Vidal-Abarca, Mañá, & Gil, 2010). Finally, monitoring (metacognitive) processes enable the dynamic update of goals throughout the reading activity. Task management is represented in the background of text processing to emphasise the fact that it constitutes a different, metacognitive level of processing.

While readers' interpretation of task requirements are an important part of task management processes, it is important to stress that the construction of reading goals extends beyond the context of explicit task instructions, as goals may be self-generated based on one's own interests and initiative. However, the PISA reading literacy assessment only considers those goals that readers form upon receiving external prompts to accomplish a given task. In addition, due to implementation constraints, task management processes are represented but not directly and independently assessed as part of PISA 2018. However, portions of the background questionnaire will estimate readers' awareness of reading strategies. Future cycles may consider the use of computer-generated process indicators (such as visiting a particular page, number of question lookbacks) as part of the assessment of task management skills.

c. Summary of reading processes

To summarise, the 2018 framework features a comprehensive and detailed typology of the cognitive processes involved in purposeful reading activities as they unfold in single or multiple text environments. Due to design constraints, it is not possible to distinguish each of these processes in a separate proficiency scale. Instead, the framework also defines a smaller list of processes that will form the basis for scaling and reporting (Table 1).

It is worth noting that the 2018 process typology also permits an analysis of changes in students' proficiency at the level of broad reading processes, as the former "cognitive aspects" featured in previous frameworks can be mapped onto specific categories within the new typology. Table 1 shows the correspondence between the 2018 typology and the former 2009 typology (which was also used in 2012 and 2015). The distinction between single and multiple text processes is discussed in greater detail below.

Table 1. Mapping of 2018 process typology to 2018 reporting scales and to former 2009-2015 cognitive aspects

2018 Cognitive processes	Superordinate Category Used for Scaling in 2018	2009-2015 Aspects
Read fluently	Reported but not on PISA scale	Not assessed
Access and retrieve information within a text	Locate information	Access and retrieve
Search and select relevant text		
Represent literal meaning	Understand	Integrate and interpret
Integrate and generate inferences		
Assess quality and credibility	Evaluate and reflect	Reflect and evaluate
Reflect on content and form		
Detect and handle conflict		Complex

B. Texts

Reading requires material for the reader to read. In an assessment, that material – a text (or a set of texts) related to a particular task – must include sufficient information for the proficient reader to engage in meaningful comprehension and resolve the problem posed by the task. Although it is obvious that there are many different kinds of texts and that any assessment should include a broad range, there was never a single agreed-upon ideal categorisation of the many different kinds of text that readers encounter. With the advent of digital media and the profusion of new text genres and text-based communication services – some of which may not survive the next decade, some of which may be newly created in the same time span – this issue becomes even more complex.

Box 2. Characteristics used to classify texts in the PISA 2009 reading framework

The previous reference framework (2009) included four major dimensions to characterise texts:

Medium: print and electronic

Environment: authored and message-based

Text format: continuous, non-continuous, mixed and multiple

Text type: description, narration, exposition, argumentation, instruction and transaction

For the 2015 reading literacy assessment, only texts that had their origin as paper-based print documents were used, albeit presented on computer. For clarity, these were referred to as fixed and dynamic texts under the heading “text display space” instead of medium (clarifying that while their origin was paper-based print, students were in fact reading them on a computer screen, hence on an electronic medium). Because reading literacy was a minor domain in 2015, no new tasks were designed and implemented. Consequently, dynamic texts, i.e. texts such as websites designed to take advantage of hyperlinks, menus, and other navigational features of an electronic medium, were not part of PISA 2015¹

¹ Some dynamic navigation features were incidentally included in the 2015 assessment. This was a result of the adaptation of the trend of print documents to the electronic screen. Many of these so-called fixed texts used in previous cycles, although adapted to mimic as closely as possible the presentation of printed texts, had to be reformatted to cope with the smaller screen size typical of computer displays. Therefore, tabs and other very simple navigation tools were included to let the reader navigate from one page to another.

For 2018, reading is the major domain and a broader range of texts can be represented in the assessment. These will include texts that are typical of the print medium, and also the ever-expanding category of digital-native text genres. Just like printed texts, some digital texts are "static" in that they come with a minimal set of tools for interaction (scrolling, paging, and a find function). For instance, this is the case of documents intended for printing but displayed on a computer screen (e.g. word processing documents or pdf files). However, many digital texts come with innovative features that increase the possibilities for the reader to interact with the materials, hence the phrase "dynamic text", which is sometimes used to characterize these texts. Dynamic text features include embedded hyperlinks that take the reader to other sections, pages or web sites; advanced search functions that provide ad hoc indexes of the searched keyword and/or the highlighting of these words in the text; and social interaction like in interactive text-based communication media such as email, forums and instant messaging services.

The 2018 framework defines four dimensions of texts: source (single, multiple); organisation and navigation (static, dynamic); format (continuous, non-continuous, mixed); and type (description, narration, exposition, argument, instruction, interaction, transaction). The first three dimensions are typical of specific situations and tasks and may trigger the use of specific processes. In contrast, the fourth dimension is included mainly for purposes of domain coverage.

a. Source

In the PISA 2018 framework, a source is a unit of text. Single texts are defined by having a definite author (or group of authors), time of writing or publication date, and reference title or number. Authors may be defined precisely, like in most traditional printed books, or more vaguely like the pseudonyms in a blog post or the sponsors of a website. A single text may also be construed as such because it is presented to the reader in isolation from other texts, even if it does not explicitly bear any source indication. Multiple texts are defined by having different authors, or being published at different times, or bearing different titles or reference numbers. Note that in the PISA framework, "title" is meant in the sense of a bibliographical catalogue unit. Lengthy texts that feature several sections with titles and subtitles are still single texts, to the extent that they were written by a definite author (or group of authors) at a given date. Likewise, multi-page websites are single texts as long as there is no explicit mention of a different author or date. It is useful to point out that multiple texts may be represented on a single page. This is the case in printed newspapers and in many textbooks, but also in forums, customer reviews or question-and-answer websites. Finally, a single text may contain embedded sources, that is, references to various authors or texts (Rouet & Britt, 2014; Strømsø et al., 2013).

The source dimension in PISA 2018 replaces the distinction between "multiple" and the other types of "text formats" in the previous versions of the framework.

b. Organisation and navigation

Screen sizes vary dramatically in digital environments, from cell phone displays, which are smaller than a traditional index card, to large, multiple screen displays for simultaneously showing multiple screen windows of information. At the time of the drafting of this framework, however, the typical computer screen (such as the 15" or 17" that come with ordinary desktop and laptop computers) features a display resolution of 1024x768 pixels. Assuming a typical font size, this is enough to display about a half-page of A4 or US-Letter page; that is, a very short piece of text. Given the wide variation in the "landscape" available on screens to display text, digital texts come with a number of tools meant to let the user access and display specific passages. These tools range from generic tools, such as the scroll bar and tabs (also found in a

number of other software applications like spreadsheets and word processors) and tools to resize or position the text on the screen, to more specific devices such as menus, tables of contents and embedded hyperlinks to move between text segments. There is growing evidence that navigation in digital text requires specific skills (OECD, 2011; Rouet, Vörös, & Pléh, 2012). Therefore, it is important to assess readers' ability to deal with texts featuring a high density of navigation tools. For reasons of simplicity, the PISA 2018 framework distinguishes “static” texts, with a simple organisation and low density of navigation tools (typically, one or several screen pages arranged in a linear way), from “dynamic” texts, which feature a more complex, non-linear organisation and a higher density of navigation devices. Note that the term “density” is preferred to “number” to mark the fact that dynamic texts do not have to be longer than static texts.

For purposes of coverage, the 2018 framework also retains two former dimensions of texts, “format” and “type”, that remain for the most part unchanged from the previous framework.

c. Text format

An important classification of texts, and one at the heart of the organisation of the PISA 2000 framework and assessment, is the distinction between continuous and non-continuous texts. Continuous texts are typically composed of sentences that are, in turn, organised into paragraphs. These may fit into even larger structures such as sections, chapters and books. Non-continuous texts are most frequently organised in matrix format, based on combinations of lists.

Texts in continuous and non-continuous formats appear in both fixed and dynamic texts. Mixed and multiple format texts are also prevalent in both, particularly so in dynamic texts. Each of these four formats is elaborated below.

Other non-text-formatted objects are also commonly used in conjunction with fixed texts and particularly with dynamic texts. Pictures and graphic images occur frequently in fixed texts and can legitimately be regarded as integral to such texts. Static images as well as videos, animations and audio files regularly accompany dynamic texts and can, also, be regarded as integral to those texts. As a reading literacy assessment, PISA does not focus on non-text formatted objects independently, but any such objects may, in principle, appear in PISA as part of a (verbal) text. However, in practice the use of video and animation is very limited in the current assessment. Audio is not used at all because of practical limitations such as the need for headphones and audio translation.

Continuous texts. Continuous texts are formed by sentences organised into paragraphs. Examples of text objects in continuous text format include newspaper reports, essays, novels, short stories, reviews and letters, including on e-book readers.

Graphically or visually, organisation occurs by the separation of parts of the text into sentences and paragraphs with spacing (e.g. indentation) and punctuation conventions. Texts also follow a hierarchical structure signalled by headings and content that help readers to recognise the organisation of the text. These markers also provide clues to text boundaries (showing section completion, for example). The location of information is often facilitated by the use of different font sizes, font types such as italic and boldface or borders and patterns. The use of typographical and format clues is an essential subskill of effective reading.

Discourse markers also provide organisational information. Sequence markers (first, second, third, etc.), for example, signal the relation of each of the units introduced to each other and indicate how the units relate to the larger surrounding text. Causal connectors (therefore, for this reason, since, etc.) signify cause-effect relationships between parts of a text.

Non-continuous texts. Non-continuous texts are organised differently to continuous texts, and therefore require a different kind of reading approach. Most non-continuous texts are composed of a number of lists (Kirsch & Mosenthal, 1990). Some are single, simple lists, but most consist of several simple lists combined.

Examples of non-continuous text objects are lists, tables, graphs, diagrams, advertisements, schedules, catalogues, indexes and forms. These text objects occur in both fixed and dynamic texts.

Mixed texts. Many fixed and dynamic texts are single, coherent objects consisting of a set of elements in both a continuous and non-continuous format. In well-constructed mixed texts the components (for example, a prose explanation including a graph or table) are mutually supportive through coherence and cohesion links at the local and global level.

Mixed text is a common format in magazines, reference books and reports, where authors employ a variety of presentations to communicate information. In dynamic texts, authored web pages are typically mixed texts, with combinations of lists, paragraphs of prose and often graphics. Message-based texts, such as online forms, e-mail messages and forums, also combine texts that are continuous and non-continuous in format.

The “multiple” format defined in the previous versions of the framework is now represented as one modality of the new “source” dimension defined above.

3. ASSESSING READING LITERACY

The previous section outlined the conceptual framework for reading literacy. The concepts in the framework must in turn be represented in tasks and questions in order to collect evidence of students’ proficiencies in reading literacy.

In this section, we consider the use of scenarios, factors affecting item difficulty, dimensions ensuring coverage and some of the other major issues in constructing and operationalising the assessment.

A. Scenarios

Reading is a purposeful act that occurs within the context of particular reader goals. In many traditional reading assessments, test takers are presented with a series of unrelated passages on a range of general topics. Students answer a set of discrete items on each passage and then move on to the next unrelated passage. In this traditional design, students are effectively expected to “forget” what they read previously when answering questions on later passages. Consequently, there is no overarching purpose for reading other than to answer discrete questions (Rupp et al., 2006). In contrast to this approach, a scenario-based assessment approach can influence the ways in which students use the texts in order to assess specific processes (e.g. Sabatini et al., 2014, 2015).

The PISA 2018 assessment will include scenarios in which students are provided an overarching purpose for reading a collection of thematically related texts in order to complete a higher-level task (e.g. respond to some larger integrative question, write a recommendation based on a set of texts), along with traditional PISA reading units. The reading purpose sets up a collection of goals, or criteria, that students use to search for information, evaluate sources, read for comprehension and/or integrate across texts. The collection of sources can be diverse and may include a selection from literature, textbooks, e-mails, blogs, websites, policy documents, primary historical documents and so forth. Although the prompts and tasks that will evolve from this framework may not grant student test-takers freedom to choose their own purposes for reading and the texts related to those individual purposes, the goal of this assessment is to offer test-takers some freedom in choosing certain textual sources and paths after attending to initial prompts. In this way, within the constraints of a large-scale assessment, goal-driven reading can be assessed.

B. Tasks

Each scenario is made up of one or more tasks. For each task, students may be asked questions about the texts ranging from traditional comprehension items (locate information, perform an inference) to more complex tasks such as the synthesis and integration of multiple texts, evaluating web search results or corroborating information across multiple texts. Each task is designed to assess one or more processes identified in the framework. Tasks in a scenario can be sequenced starting with less difficult to more complex to provide information about different student abilities. For instance, a student might encounter an initial task in which the student must locate a particular document based on a search result. As a second task, the student might have to answer a question about information that is specifically stated in the text. As a third task, the student might need to determine if the author's point of view in the first text is the same as a second text. In each case, these tasks can be scaffolded so that if a student fails to find the correct document in the first task, the student is then provided with the correct document in order to complete the second task. In this way, complex multipart scenarios do not become an "all or none activity", but rather a way to help triangulate the level of different student skills within the context of realistic tasks. Thus, scenarios can be thought of as corresponding to units and tasks as items within units from previous PISA reading literacy assessments.

A scenario-based assessment mimics the way an individual interacts and uses literacy source material in a more authentic way than in traditional, decontextualised assessments. It presents students with realistic problems and issues to solve, and it involves the use of both basic and higher-level reading and reasoning skills (O'Reilly & Sabatini, 2013).

Scenarios make use of the affordances of computer-based assessments, such as the possibility of introducing stimulus material in a paced sequence, providing feedback to students and delivering items in a lockstep fashion, in order to provide scaffolding and manipulate difficulty. Scenarios represent a natural extension of the traditional, unit-based approach in PISA. A scenario-based approach was used in the PISA 2012 assessment of problem solving and the PISA 2015 assessment of collaborative problem solving. Tasks 2-4 in Appendix B illustrate a sample scenario with multiple items.

a. Distribution of tasks

Tasks are designed to assess the specific skills as were defined in the section on processes. Each task will primarily assess one process. As such, they can be thought of as individual assessment items. The approximate distribution of tasks for the 2018 reading literacy assessment are shown below in Table 2 and are contrasted with the distribution of tasks for the 2015 assessment.

Table 2. Approximate distribution of tasks per targeted process and text availability.

2015 FRAMEWORK	2018 FRAMEWORKS	
	SINGLE Text	MULTIPLE Text
Access and Retrieve 25%	Scan and Locate 15%	Search and select relevant text 10%
Integrate and Interpret 50%	Literal Comprehension 15% Inference Comprehension 15%	Inference Comprehension 15%
Reflect and Evaluate 25%	Assess quality and credibility 20% Reflect on content and form	Corroborate/Handle conflict 10%

Items will be reused from previous PISA reading literacy assessments in order to maintain trends. In order to achieve the desired representation of multiple text tasks, and because prior PISA assessments focused on single text tasks, the development of new items will mostly require the creation of tasks involving multiple texts (e.g. search, inference and corroborate/conflict). At the same time, a sufficient number of single-text items within the newly developed scenarios need to be present to ensure that future trend items cover the entire framework.

C. Factors affecting item difficulty

The purpose of the PISA reading literacy assessment is to monitor and report on the reading proficiency of 15-year-olds as they approach the end of compulsory education. Each task in the assessment is designed to gather a specific piece of evidence about that proficiency by simulating a reading activity that a reader might carry out either inside or outside school, as an adolescent or as an adult.

The PISA reading literacy tasks range from very straightforward locating and comprehension activities to quite sophisticated activities requiring integrating information across multiple texts. The difficulty of any reading literacy task depends on an interaction amongst several variables. Drawing on Kirsch and Mosenthal's work (see for example Kirsch, 2001; Kirsch & Mosenthal, 1990), the difficulty of items can be manipulated by applying knowledge of the process and text format variables. In Table 3 below we outline how the difficulty can be manipulated across the different types of tasks.

Table 3. Item difficulty for tasks.

Single	Multiple
<p>In scan and locate tasks, difficulty is conditioned by the number of pieces of information that the reader needs to locate, by the amount of inferencing required, by the amount and prominence of competing information and by the length and complexity of the text.</p>	<p>Multiple document search difficulty is conditioned by the number of texts, the complexity of the document hierarchy (depth and breadth), familiarity of the structure, the amount of non-hierarchical linking, the distance to the goal, the salience and relevance of the headers and the dissimilarity of each of the physical presentation/structure of the sources (lack of parallelism in different source texts)</p>
<p>In literal and explicit meaning and integrate and generate inferences tasks, difficulty is affected by the type of interpretation required (for example, making a comparison is easier than finding a contrast); by the number of pieces of information to be considered; by the degree and prominence of competing information in the text; and by the nature of the text: the less familiar and the more abstract the content and the longer and more complex the text, and the lower the coherence of the structure, the more difficult the task is likely to be.</p>	<p>In multiple documents, inference difficulty is conditioned on the number of texts, the salience of the headers, the similarity of content (e.g. discrepancy in text content/arguments, variability in point of view), the dissimilarity of the physical presentation/structure of the sources (lack of parallelism in different source texts) and the explicitness of source information.</p>
<p>In reflect on content and form tasks, difficulty is affected by the type of reflection or evaluation required (from least to most difficult with types of reflection being: connecting; explaining and comparing; hypothesising and evaluating); by the nature of the knowledge that the reader needs to bring to the text (a task is more difficult if the reader needs to draw on narrow, specialised knowledge rather than broad and common knowledge); by the relative abstraction and length of the text; and by the depth of understanding of the text required to complete the task.</p> <p>For assess quality and credibility tasks credibility and quality of a source can be conditioned by using text signals such as the explicitness of the source and the degree to which a text appears to be advertising or comes from a reputable source.</p>	<p>In multiple documents, corroborate/conflict/synthesize difficulty is conditioned on the number of texts, the dissimilarity of content (discrepancy in texts content/arguments), the dissimilarity of each of the physical presentation/structure of the sources (lack of parallelism in different source texts), the explicitness of the source information, and the degree of credibility of the source.</p>

Box 3. Text availability and its impact on comprehension in the design of tasks

In the last decade, there has been some debate whether memory-based measures of reading comprehension, i.e. answering comprehension question while the text is not available to students after initial reading, might be a better indicator of students' reading comprehension skills than questions with text availability. From a theoretical point of view, arguments can be made for both, with- and without-text availability questions. Answering comprehension questions with a text might be more ecologically valid because many reading settings (especially in the digital age) potentially allow the reader to refer back to the text. In addition, if the text is not available to students, their performance on the comprehension questions might be confounded with their memory skills, i.e. their ability to remember the content of the text. On the other hand, answering comprehension questions without text availability is also a common reading situation (e.g. commenting on a newspaper article over lunch that has been read in the morning) and might be less confounded by students' motivational and test taking strategies. Empirically, recent studies (Ozuru et al., 2007; Schroeder, 2011) do provide some evidence that comprehension questions without text availability might indeed be more sensitive to the quality of the processes that are executed while students are reading a text and the strength of the resulting memory representation. At the same time, however, both kinds of measures are highly correlated and are thus difficult to dissociate empirically. At present, therefore, there is not enough evidence that justifies any major changes in the way PISA is administered. However, it is encouraged to include further measures in the analysis, e.g. time on task, time of initial reading of a text, etc., in order to further explore this issue.

D. Factors improving the coverage of the domain

a. Situations

Scenarios can be developed across a wide range of potential situations. *Situation* is used to define the contexts and uses for which the author constructed the text. The manner in which the situation variable is specified is therefore about supposed audience and purpose, and is not simply based on the place where, or the purpose for which, the reading activity is carried out.

The framework categorises situations using a typology adapted from the Common European Framework of Reference (CEFR) developed for the Council of Europe (Council of Europe, 1996). The situations are personal, public, occupational and educational contexts and are defined in Box 4. In contrast to previous PISA reading literacy assessments, texts from the different situations may be mixed within a scenario. For example, a student may perform a set of tasks that require relating an *educational* text that provides historical content to *personal* texts that provide first person accounts of the events.

Box 4. Categorisation of situations

A personal situation relates to texts that are intended to satisfy an individual's personal interests, both practical and intellectual. This category also includes texts that are intended to maintain or develop personal connections with other people. It includes personal letters, fiction, biography and informational texts that are intended to be read to satisfy curiosity, as a part of leisure or recreational activities. In the electronic medium, it includes personal e-mails, instant messages and diary-style blogs.

A public situation describes the reading of texts that relate to activities and concerns of the larger society. The category includes official documents as well as information about public events. In general, the texts associated with this category assume a more or less anonymous contact with others; therefore they also include message boards, news websites and public notices that are encountered both on line and in print.

The content of educational situations makes use of texts designed specifically for the purpose of instruction. Printed textbooks, electronic textbooks and interactive learning software are typical examples of material generated for this kind of reading. Educational reading normally involves acquiring information as part of a larger learning task. The materials are often not chosen by the reader, but instead assigned by an instructor.

Many 15-year-olds will move from school into the labour force within one to two years. A typical occupational reading situation is one that involves the accomplishment of some immediate task. It might include searching for a job, either in a print newspaper's classified advertisement section or online; or following workplace directions. Texts written for these purposes, and the tasks based on them, are classified as occupational in PISA. While only some of the 15-year-olds who are assessed will currently have to read at work, it is important to include tasks based on texts that are related to work since the assessment of young people's readiness for life beyond compulsory schooling and their ability to use their knowledge and skills to meet real-life challenges is a fundamental goal of PISA.

Many texts used in classrooms are not specifically designed for classroom use. For example, a piece of literary text may typically be read by a 15-year-old in a mother-tongue language or literature class, yet the text was written (presumably) for readers' personal enjoyment and appreciation. Given its original purpose, such a text is classified as being used in a personal situation in PISA. As Hubbard (1989) has shown, some kinds of reading usually associated with out-of-school settings for children, such as rules for clubs and records of games, often take place informally at school as well. These are classified as public situations in PISA. Conversely, textbooks are read both in schools and in homes, and the process and purpose probably differ little from one setting to another. These are classified as educational situations in PISA.

It should be noted that many texts can be cross-classified to different situations. In practice, for example, a text may be intended both to delight and to instruct (personal and educational); or to provide professional advice, which is also general information (occupational and public). While content is not a variable that is specifically manipulated in this study, by sampling texts across a variety of situations the intent is to maximise the diversity of content that will be included in the PISA reading literacy test.

b. Text types

Text types further describe the diversity of texts in a way to cover a wide range of types of reading that students would encounter: description, narration, exposition, argumentation, instruction and transaction². Texts as they are found in the world typically resist categorisation, as they are usually not written with text type rules in mind, and tend to cut across categories. For example, a chapter in a textbook might include some definitions (exposition), some directions on how to solve particular problems (instruction), a brief historical account of the discovery of the solution (narration) and descriptions of some typical objects involved in the solution (description). Nevertheless, in an assessment like PISA it is useful to categorise texts according to the text type, based on the predominant characteristics of the text, in order to ensure that the instrument samples across a range of texts that represent different types of reading.

The classification of texts used in PISA is adapted from the work of Werlich (1976) and are shown in Box 5.

² In the first version of the reading framework, these text types were located as subcategories of the continuous text format. In the PISA 2009 cycle it was acknowledged that non-continuous texts (and the elements of mixed and multiple texts) also have a descriptive, narrative, expository, argumentative or instructional purpose.

Box 5. Classification of texts

Description is the type of text where the information refers to properties of objects in space. The typical questions that descriptive texts provide an answer to are what questions. Descriptions can take several forms. Impressionistic descriptions present information from the point of view of subjective impressions of relations, qualities and directions in space. Technical descriptions present information from the point of view of objective observation in space. Frequently, technical descriptions use non-continuous text formats such as diagrams and illustrations. Examples of text objects in the text type category description are a depiction of a particular place in a travelogue or diary, a catalogue, a geographical map, an online flight schedule or a description of a feature, function or process in a technical manual.

Narration is the type of text where the information refers to properties of objects in time. Narration typically answers questions relating to *when*, or *in what* sequence. Why characters in stories behave as they do is another important question that narration typically answers. Narration can take different forms. *Narratives* present change from the point of view of subjective selection and emphasis, recording actions and events from the point of view of subjective impressions in time. *Reports* present change from the point of view of an objective situational frame, recording actions and events which can be verified by others. *News* stories intend to enable the readers to form their own independent opinion of facts and events without being influenced by the reporter's references to his own views. Examples of text objects in the text type category *narration* are a novel, a short story, a play, a biography, a comic strip and a newspaper report of an event.

Exposition is the type of text in which the information is presented as composite concepts or mental constructs, or those elements into which concepts or mental constructs can be analysed. The text provides an explanation of how the different elements interrelate in a meaningful whole and often answers questions about *how*. Expositions can take various forms. *Expository* essays provide a simple explanation of concepts, mental constructs or conceptions from a subjective point of view. *Definitions* explain how terms or names are interrelated with mental concepts. In showing these interrelations, the definition explains the meaning of words. *Explications* are a form of analytic exposition used to explain how a mental concept can be linked with words or terms. The concept is treated as a composite whole that can be understood by breaking it down into its constituent elements and then naming the interrelations of those elements. *Summaries* are a form of synthetic exposition used to explain and communicate texts in a shorter form than the original text requires. *Minutes* are a record of the results of meetings or presentations. *Text interpretations* are a form of both analytic and synthetic exposition used to explain the abstract concepts that are realised in a particular (fictional or non-fictional) text or group of texts. Examples of text objects in the text type category *exposition* are a scholarly essay, a diagram showing a model of memory, a graph of population trends, a concept map and an entry in an online encyclopaedia.

Argumentation is the type of text that presents the relationship among concepts or propositions. Argumentative texts often answer *why* questions. An important subclassification of argumentative texts is persuasive and opinionative texts, referring to opinions and points of view. *Comment* relates the concepts of events, objects and ideas to a private system of thoughts, values and beliefs. *Scientific argumentation* relates concepts of events, objects and ideas to systems of thought and knowledge so that the resulting propositions can be verified as valid or non-valid. Examples of text objects in the text type category *argumentation* are a letter to the editor, a poster advertisement, the posts in an online forum and a web-based review of a book or film.

Instruction (sometimes called injunction) is the type of text that provides directions on what to do. *Instructions* present directions for certain behaviours in order to complete a task. *Rules*, *regulations* and *statutes* specify requirements for certain behaviours based on impersonal authority, such as practical validity or public authority. Examples of text objects in the text type category *instruction* are a recipe, a series of diagrams showing a procedure for giving first aid and guidelines for operating digital software.

Transaction represents the kind of text that aims to achieve a specific purpose outlined in the text, such as requesting that something is done, organising a meeting or making a social engagement with a friend. Before the spread of electronic communication, this kind of text was a significant component of some kinds of letters and, as an oral exchange, the principal purpose of many phone calls. This text type was not included in Werlich's (1976) categorisation, used until now for the PISA framework.

The term transactional is used in PISA not to describe the general process of extracting meaning from texts (as in reader-response theory), but the type of text written for the kinds of purposes described here. Transactional texts are often personal in nature, rather than public, and this may help to explain why they do not appear to be represented in some of the corpora used to develop many text typologies. For example, this kind of text is not commonly found on websites, which are frequently the subject of corpus linguistics studies (for example, Santini, 2006). With the extreme ease of personal communication using e-mail, text messages, blogs and social networking websites, this kind of text has become much more significant as a reading text type in recent years. Transactional texts often build on common and possibly private understandings between communicators – though clearly, this feature is difficult to explore in a large-scale assessment. Examples of text objects in the text type transaction are everyday e-mail and text message exchanges between colleagues or friends that request and confirm arrangements.

Narration occupies a prominent position in many national and international assessments. Some texts are presented as being accounts of the world as it is (or was) and therefore claim to be factual or non-fictional. Fictional accounts bear a more metaphorical relation to the world as it is, appearing either as accounts of how it might be or of how it seems to be. In other large-scale reading studies, particularly those for school students: the National Assessment of Educational Progress (NAEP); the IEA Reading Literacy Study (IEARLS); and the IEA Programme in International Reading Literacy Study (PIRLS), the major classification of texts is between fictional or literary texts and non-fictional texts (*reading for literary experience and reading for information or to perform a task* in NAEP; *literary experience and acquire and use information* in PIRLS). This distinction is increasingly blurred as authors use formats and structures typical of factual texts in creating their fictions. The PISA reading literacy assessment includes both factual and fictional texts, and texts that may not be clearly one or the other. PISA, however, does not attempt to measure differences in reading proficiency between one type and the other. In PISA, fictional texts are

E. Response formats

The form in which the evidence is collected – the *response format* – varies according to what is considered appropriate given the kind of evidence that is being collected, and also according to the pragmatic constraints of a large-scale assessment. As in any large-scale assessments the range of feasible item formats is limited. However, with computers for assessment, the types of response formats can include interactions with text, such as highlighting and drag-and-drop, as well as multiple choice and short constructed response items (to which students write their own answer).

Response formats can be differentially sensitive to individual differences. For example, cloze and sometimes multiple choice are typically more dependent on decoding skills, because readers have to decode distractors or items, when compared to open constructed response items (Cain & Oakhill, 2006). Several studies based on PISA data suggest that the response format has a significant effect on the performance of different groups: for example, students at different levels of proficiency (Routitsky & Turner, 2003); students in different countries (Grisay & Monseur, 2007); students with different levels of intrinsic reading motivation (Schwabe, McElvany & Trendtel, 2015), and boys and girls (Lafontaine & Monseur, 2006, 2006b; Schwabe, et al., 2015). Given this variation, in measuring trends over time, it is important to maintain a similar proportion of tasks in multiple choice and constructed response formats from one administration to the next. A further significant consideration in the context of reading literacy is that open constructed response items are particularly important for the reflection and evaluation aspect, where the intent is often to assess the quality of thinking rather than the conclusion itself. Nevertheless, because the focus of the assessment is on reading and not on *writing*, constructed response items should not be designed to put great emphasis on assessing writing skills, such as spelling, grammar, etc. Finally, students in different countries are more or less familiar with various response formats. Including items in a variety of formats is likely to provide some balance between more and less familiar formats for all students, regardless of nationality.

In summary, to ensure proper coverage of the ability ranges in different countries, to ensure fairness given the inter-country and gender differences observed and to ensure a valid assessment of the *reflect and evaluate* aspect, both multiple choice and open constructed response items continue to be used in PISA reading literacy assessments regardless of the change in delivery mode. Any major change in the distribution of item types in print reading might also impact the measurement of trends.

Box 6. The status of writing skills in PISA 2018 reading literacy assessment

Skilled readers are often required to write comments, essays or explanations in response to questions, or choose to make notes, outlines and summaries, or simply write down their thoughts and reflections about texts, towards achieving their reading goals. They also routinely engage in written communication with others (e.g. teachers, student peers, acquaintances) in learning (e.g. an email assignment from a teacher) or social (e.g. a chat with peers about text or school literacy contexts). The PISA 2018 reading framework considers writing to be an important correlate of reading literacy. Test design and administration constraints prohibit the inclusion of the assessment of writing skills, where writing is in part defined as the quality and organization of the production. However, a significant proportion of test items requires readers to articulate their thinking into written answers. Thus, the assessment of reading skills also draws on readers' ability to communicate their understanding in writing, although such aspects as spelling, quality of writing and organization are not measured in PISA.

F. Assessing the ease and efficiency of reading simple texts for understanding

The PISA 2018 reading literacy assessment will include the assessment of reading fluency, defined as the ease and efficiency with which students can read simple texts for understanding. This will provide a valuable indicator for use in describing or understanding differences between students, especially for individuals in the lower reading proficiency levels. Students with low levels of foundational reading skills may be exerting so much attention and cognitive effort on lower level skills of decoding, word recognition and sentence parsing that they have diminished resources to perform higher-level comprehension tasks with single or multiple texts. This finding applies to developing as well as teenage readers (Rasinski et al., 2005; Scamacca et al., 2006).

The computerized administration and scoring in PISA 2018 allows the measurement of the ease and efficiency with which 15-year-olds can read simple texts for understanding. While not all slow reading is poor reading, as noted above, a large body of evidence documents how and why a lack of automaticity in one's basic reading processes can be a bottleneck to higher-level reading proficiency and is associated with poor comprehension (e.g. Rayner et al., 2001). Thus, it is valuable to have an indicator of the ease and efficiency with which 15-year-olds can read simple texts accurately for understanding to better describe and interpret very low-level performance on PISA comprehension tasks.

It is further worth noting that with the exponential expansion of text content available on the Internet, there is an ever greater need for 21st century students to not only be proficient readers, but also efficient readers (OECD, 2011). Thus, a basic indicator of reading rate under low demand conditions can also be used descriptively for other purposes, such as investigating how much students regulate their rate or strategic processes in the face of more complex tasks or larger volumes of text.

While there are many variations in how to define, operationalize and measure reading ease, efficiency or fluency, the most common evidence collected when using silent reading tasks are indicators of accuracy and rate. Oral reading fluency measures can also be used to estimate prosody and expressiveness of the reader, but these attributes are more challenging to measure in silent reading tasks and there is less agreement concerning their added value over and above strong indicators of accuracy and rate (Eason et al., 2013; Kuhn, Schwanenflugel, & Meisinger, 2010). In addition, it is not currently feasible to implement and score oral reading tasks in all the languages in which PISA is available. Thus, a silent reading task design is recommended.

In order to better understand the challenges facing 15-year-olds scoring at lower levels on the PISA reading literacy task, a specific task can be administered near the start of the assessment to measure reading ease and efficiency. Performance on this task can be scaled and reported independently from the main proficiency scales. As noted, inefficient reading can be a symptom of low foundational skills. However, there may be individuals who are relatively slow readers, yet possess compensatory or strategic processes that permit them to be higher-level readers when given sufficient time to complete complex tasks. This may be especially the case for non-native speakers of a language, who may be relatively slower than native speakers, but score comparably to more proficient students on untimed tasks. Thus, it seems most prudent to use the ease of reading indicator as a descriptive variable to help differentiate students who may have foundational skill deficits from those who are slow, but nonetheless proficient readers.

In addition, an index of ease and efficiency of reading could be, as one of several indicators, used for placing students in a level for adaptive testing (see section below on “Considerations for adaptive testing”). For the reasons cited in the previous paragraph, the index may not be suitable as a sole indicator of reading level, however, when combined with other evidence, inefficiency in basic processing may be helpful in placing students in appropriate levels.

A task design that has been used effectively as an indicator of reading ease and efficiency in other empirical research requires students to read a sentence and make a judgment of the plausibility of the sentence in relation to world knowledge or internal logical consistency of the sentence. The measure takes into account both accuracy of understanding the text and the time it takes to read and respond. This sentence task structure has been used in the Woodcock Johnson Subtest of Reading Fluency (Woodcock, McGrew, & Mather, 2001) and the Tests of Silent Reading Efficiency and Comprehension (TOSREC) (Wagner, Torgesen, Rashotte, & Pearson, 2010). It is also the task type used in the PIAAC Reading Components task set (OECD, 2013a; Sabatini & Bruce, 2009), and in two PISA countries (Bruce & Sabatini, 2013) with success. A similar task has been used in the Austrian PISA 2000 assessment and showed high correlations ($r = .64$) with students’ final test score (Landerl & Reiter, 2002). This task design thus has a proven empirical foundation as an index of reading ease and efficiency in international study contexts. Task 1 in Appendix B shows a sample item taken from the PIAAC Reading Components task.

While it may be possible in future cycles of PISA to use log-file data based on complex reading literacy tasks as the sole source for measuring ease and efficiency, this option is not recommended for the current cycle. In order to ensure that students complete tasks under conditions that yield a valid indicator of efficiency, the design and instructions accompanying the task should target the desired construct. The texts need to be simple and short in order to maximize reading efficiency versus strategic or compensatory processes. In addition, the task demands should take minimal reasoning so as to not confound individual differences in decision time with basic reading rate information. It will therefore be difficult to ensure that the reading rates and accuracy observed in tasks that were designed for different measurement purposes that are executed by students under these constraints. The more complex the task, the more likely that students will deploy strategic or compensatory processes that interfere with measuring ease and efficiency of basic understanding.

Thus, it is recommended that the log files from this cycle be analysed to evaluate whether there are indicators within the new PISA Reading Literacy task set that are strongly correlated with the sentence level efficiency task proposed. The probability is low that there is sufficient valid evidence in the field test log files – essentially psychometric equivalence with the sentence task – from initial item trials of the new reading literacy tasks. On the other hand, such log file correlational evidence would serve as cross-validation evidence for the ease and efficiency task.

G. Assessing students' reading motivation, reading practices and awareness of reading strategies

Since PISA 2000, the importance of motivational attributes of the reader (such as their attitude toward reading) and of their reading practices (e.g. the readers' factors in Figure 1) has been highlighted in the reading literacy framework; accordingly, items and scales have been developed to measure these important constructs in the student questionnaire. It is important to note that reading motivation and reading strategies may vary as a function of the contexts and types of texts considered. Therefore, questionnaire items assessing motivation and strategies should refer to a range of situations that represent students' practices. In addition to increased theoretical relevance, items referring to more specific and concrete situations are known to decrease the risk of response bias that may come with ratings and self-reports.

a. *Intrinsic motivation and interest in reading*

“While *motivation* refers to goals, values, beliefs in a given area, such as reading, *engagement* refers to behavioural displays of effort, time, and persistence in attaining desired outcomes” (Klauda & Guthrie, 2015, p. 240). Reading engagement, motivation and practices have been shown in a number of studies to be strongly linked with reading proficiency (Becker, McElvany, & Kortenbruck, 2010; Guthrie, Wigfield, Metsala, & Cox, 2004; Klauda & Guthrie, 2014; Mol & Bus, 2011; Morgan & Fuchs, 2007; Pfof, Dörfler, & Artelt, 2013; Schaffner, Philipp, & Schiefele, 2014; Schiefele, Schaffner, Möller, & Wiegfield, 2012). In PISA 2000, engagement in reading (comprising interest, intrinsic motivation, avoidance and practices) were strongly correlated with reading proficiency, stronger even than the association between reading literacy and socio-economic status (OECD, 2002; 2010a). In other studies, reading engagement has been shown to explain reading achievement more than any other variable besides previous reading achievement (Guthrie & Wigfield, 2000). Critically, perseverance as a characteristic of engagement has also been linked to successful learning and achievement outside of school (Heckman & Kautz, 2012). Thus, motivation and engagement are powerful variables and levers on which one can act in order to enhance reading proficiency and reduce gaps between groups of students.

In previous PISA cycles in which reading literacy was the major domain (PISA 2000 and PISA 2009), the main motivational construct investigated was *interest in reading and intrinsic motivation*. The scale measuring interest and intrinsic motivation also captured *reading avoidance*, which is lack of interest or motivation and has shown strong associations with achievement especially among struggling readers (Klauda & Guthrie, 2015; Legault, & al., 2006). For PISA 2018, in accordance with what was done in other domains, two other prominent motivational constructs will be investigated as part of the PISA questionnaire, namely *self-efficacy*, the individual's perceived capacity of doing specific tasks, and *self-concept*, the individual's own perceived abilities related to a domain.

b. *Reading practices*

Beside motivation, *reading practices* have previously been measured as *the self-reported frequencies of reading different types of texts in various media, including online reading*. In PISA 2018, the list of online reading practices scales will be updated and extended in order to take into account emerging practices (e.g. e-books, online search, short messaging and social networking).

c. Awareness of reading strategies

Metacognition, an individual's ability to think about and control his or her reading and comprehension strategies, has both a significant correlation with reading proficiency and is responsive to teaching and learning. A number of studies have found an association between reading proficiency and metacognitive strategies (Artelt, Schiefele, & Schneider, 2001; Brown, Palincsar, & Armbruster, 1984). Explicit or formal instruction of reading strategies leads to an improvement in text understanding and information use (Cantrell et al., 2010). More specifically, it is assumed that the reader becomes independent of the teacher after these strategies have been acquired and are applied without much effort. By using these strategies, the reader can effectively interact with the text by conceiving reading as a problem-solving task that requires the use of strategic thinking and by thinking strategically about solving reading comprehension problems. In previous PISA cycles, engagement and metacognition proved to be robust predictors of reading achievement, mediators of gender or socioeconomic status (OECD, 2010, b vol. III) and also potential levers to reduce achievement gaps. In the questionnaire framework, the measures of these motivational, metacognition and reader practices are updated and extended in order to take into account the recent and emerging practices (e.g. e-books, online search, social networking) as well as to better cover measurement of teaching practices and classroom support that support reading growth.

Skilled reading requires students to know and employ strategies in order to make the best use of text given their purposes and goals. For instance, students must know when it is appropriate to scan a passage or when the task requires the sustained and complete reading of the passage. In PISA 2009, information about reading strategies was collected. Two reading scenarios were presented to students. In the first scenario, students were asked to evaluate the effectiveness of different reading and text comprehension strategies to reach the goal of *summarising information*; in the second, students had to evaluate the effectiveness of other strategies for *understanding and remembering a text*. For PISA 2018, in accordance with the new frame of reading processes (see Figure 2), information will also be collected about knowledge of reading strategies specifically linked to the goal of “*assessing the quality and credibility of sources*”, which is particularly prominent in digital reading and when confronted with multiple texts.

d. Teaching practices and classroom support for reading growth and engagement

There is strong research evidence showing that classroom practices, such as the direct teaching of reading strategies, contribute to growth in reading skill (Pressley, 2000; Rosenshine & Meister, 1997; Waters & Schneider, 2010). In addition, teachers' scaffolding and support for autonomy, competence and ownership improve students' reading proficiency, awareness of strategies, and engagement in reading (Guthrie, Ho, & Klauda, 2013; Guthrie, Wigfield, & You, 2012). While in most educational systems, reading is no longer taught as a subject matter to 15-year-old students in the same way as are mathematics and science, some reading instruction may be explicitly or incidentally given in language lessons and in other disciplines (e.g. social science, science, foreign languages, civic education, ICT). Yet the dispersed nature of reading instruction represents a challenge for articulating questions that capture the classroom practices and opportunities to learn to which students may be exposed. Despite these challenges, it is thought extremely important to capture through the student questionnaire the relevant instructional processes – opportunity-to-learn and teaching practices – that might support the development of students' reading skills, practices and motivation.

H. Considerations for adaptive testing

The deployment of computer-based assessment in PISA creates the opportunity to implement adaptive testing. Adaptive testing enables higher levels of measurement precision using fewer items per individual student. This is accomplished by targeting more items that are aligned to the ability range of students at different points in the ability distribution.

Adaptive testing has the potential to increase the resolution and sensitivity of the assessment, most particularly at the lower end of the distribution of student performance. For example, students who perform low on items that assess their ease and efficiency of reading (e.g. reading fluency) will likely struggle on highly complex multiple text items. Thus, there would be benefit in providing additional lower-level texts for those students to better assess specific aspects of their comprehension.

4. REPORTING PROFICIENCY IN READING

A. Reporting scales

PISA reports students' results in terms of proficiency scales that are interpretable in educational policy terms. In PISA 2000, when reading was the major domain, the results of the reading literacy assessment were first summarised on a single composite reading literacy scale having a mean of 500 and a standard deviation of 100. In addition to the composite scale, student performance was also represented on five subscales: three process (aspect) subscales (retrieving information, interpreting texts, and reflection and evaluation) and two text format subscales (continuous and non-continuous) (OECD, 2002). These five subscales made it possible to compare mean scores and distributions among subgroups and countries by various components of the reading literacy construct. Although there is a high correlation between these subscales, reporting results on each subscale revealed interesting deviations among the participating countries. Where such deviations occur, they can be examined and linked to the curriculum and teaching methodology used. In some countries, the important question may be how to teach the current curriculum better. In others, the question may be not only how to teach but also what to teach. In PISA 2009, reading was again the major domain. A reporting scheme including subscales as well as a composite scale was used.

In both PISA 2003 and 2006, and 2012 when reading was a minor domain, and fewer reading items were administered to participating students, a single reading literacy trend scale was reported based upon the overall composite scale (OECD, 2004, 2007, 2014). In 2018 reading is the major domain, and reporting on subscales is again possible.

For PISA 2018, the reporting subscales will be (see also Table 1):

1. Locate information, which is composed of tasks that require students to search and select relevant texts, and access relevant information within texts.
2. Understand, which is composed of tasks that require students to represent the explicit meaning of texts as well as integrate information and generate inferences.
3. Evaluate and reflect, which is composed of tasks that require the student to assess the quality and credibility of information, reflect on the content and form of a text and detect and handle conflict within and across texts.

As described above, a separate subscore for reading fluency can also be provided as a measure of students' ease and efficiency of reading. This subscore will not be reported on the PISA scale, but can be used to help interpreting student's performance.

B. Interpreting and using the scales

Just as students can be ordered from the least proficient to the highly skilled on a single scale, reading literacy tasks are arranged along a scale that indicates progressively the level of difficulty for students and the level of skill required to answer each item correctly. By comparing the position of students and items on these scales, we can summarise both the proficiency of a person in terms of his or her ability and the complexity of an item in terms of its difficulty.

Reading literacy tasks used in PISA vary widely in situation, text format and task requirements, and they also vary in difficulty. This range is captured through what is known as an item map. The item map provides a visual representation of the reading literacy skills demonstrated by students at different points along the scale.

Tasks at the lower end of the reading scale and subscales differ from those at the higher end. Difficulty is in part determined by the length, structure and complexity of the text itself. However, while the structure of a text contributes to the difficulty of an item, what the reader has to do with that text, as defined by the question or instruction, interacts with the text and affects the overall difficulty. A number of variables that can influence the difficulty of any reading literacy task have been identified, including the complexity and sophistication of the mental processes integral to the aspect of the task (retrieving, interpreting or reflecting), the amount of information to be assimilated by the reader and the familiarity or specificity of the knowledge that the reader must draw on both from within and from outside the text.

C. Defining levels of reading literacy proficiency

In an attempt to capture this progression of complexity and difficulty in PISA 2000, the composite reading literacy scale and each of the subscales were divided into six levels (Below level 1, 1, 2, 3, 4, 5). These levels as they were defined for PISA 2000 were kept for the composite scale used to measure trends in PISA 2009 and 2015. However, newly constructed items helped to improve descriptions of the existing levels of performance and to furnish descriptions of levels of performance above and below those established in PISA 2000. Thus, the scales were extended to level 6, and level 1b was introduced at the bottom of the scale (OECD, 2012).

The levels provide a useful way to explore the progression of reading literacy demands within the composite scale and each subscale. The scale summarises both the proficiency of a person in terms of his or her ability and the complexity of an item in terms of its difficulty. The mapping of students and items on one scale represents the idea that students are more likely to be able to successfully complete tasks mapped at the same level on the scale (or lower), and less likely to be able to successfully complete tasks mapped at a higher level on the scale.

As an example, the reading proficiency scale for the PISA 2012 study is represented in Table 4. The left-hand column shows the level number, the lower score limit, and the percentage of students who are able to perform tasks at each level or above (OECD average). The right-hand column describes what students can do at each level (adapted from OECD, 2013).

Table 4. An overview of reading proficiency levels as they were described in the PISA 2012 study.

Level	What Students Can Do
<p>6 698 1.1%</p>	<p>Readers at level 6 typically can make multiple inferences, comparisons and contrasts that are both detailed and precise. They demonstrate a full and detailed understanding of one or more texts and may integrate information from more than one text. Tasks may require the reader to deal with unfamiliar ideas, in the presence of prominent competing information, and to generate abstract categories for interpretations. Students can hypothesise about or critically evaluate a complex text on an unfamiliar topic, taking into account multiple criteria or perspectives, and applying sophisticated understandings from beyond the text. A salient condition for access and retrieve tasks at this level is precision of analysis and fine attention to detail that is inconspicuous in the texts.</p>
<p>5 626 8.4%</p>	<p>At level 5, readers can locate and organise several pieces of deeply embedded information, inferring which information in the text is relevant. Reflective tasks require critical evaluation or hypothesis, drawing on specialised knowledge. Both interpretative and reflective tasks require a full and detailed understanding of a text whose content or form is unfamiliar. For all aspects of reading, tasks at this level typically involve dealing with concepts that are contrary to expectations.</p>
<p>4 553 29.5%</p>	<p>At level 4, readers can locate and organise several pieces of embedded information. They can also interpret the meaning of nuances of language in a section of text by taking into account the text as a whole. In other interpretative tasks, students demonstrate understanding and application of categories in an unfamiliar context. In addition, students at this level can use formal or public knowledge to hypothesise about or critically evaluate a text. Readers must demonstrate an accurate understanding of long or complex texts whose content or form may be unfamiliar.</p>
<p>3 480 58.6%</p>	<p>Readers at level 3 can locate, and in some cases recognise the relationship between, several pieces of information that must meet multiple conditions. They can also integrate several parts of a text in order to identify a main idea, understand a relationship or construe the meaning of a word or phrase. They need to take into account many features in comparing, contrasting or categorising. Often the required information is not prominent or there is much competing information; or there are other text obstacles, such as ideas that are contrary to expectation or negatively worded. Reflective tasks at this level may require connections, comparisons, and explanations, or they may require the reader to evaluate a feature of the text. Some reflective tasks require readers to demonstrate a fine understanding of the text in relation to familiar, everyday knowledge. Other tasks do not require detailed text comprehension but require the reader to draw on less common knowledge.</p>
<p>2 407 82.0%</p>	<p>Readers at level 2 can locate one or more pieces of information, which may need to be inferred and may need to meet several conditions. They can recognize the main idea in a text, understand relationships, or construe meaning within a limited part of the text when the information is not prominent and the reader must make low-level inferences. Tasks at this level may involve comparisons or contrasts based on a single feature in the text. Typical reflective tasks at this level require readers to make a comparison or several connections between the text and outside knowledge, by drawing on personal experience and attitudes.</p>
<p>1a 335 94.3%</p>	<p>Readers at level 1a can locate one or more independent pieces of explicitly stated information; they can recognise the main theme or author's purpose in a text about a familiar topic, or to make a simple connection between information in the text and common, everyday knowledge. Typically the required information in the text is prominent and there is little, if any, competing information. The student is explicitly directed to consider relevant factors in the task and in the text.</p>
<p>1b 262 98.7%</p>	<p>Readers at level 1b can locate a single piece of explicitly stated information in a prominent position in a short, syntactically simple text with a familiar context and text type, such as a narrative or a simple list. Texts in level 1b tasks typically provide support to the reader, such as repetition of information, pictures or familiar symbols. There is minimal competing information. Level 1b readers can interpret texts by making simple connections between adjacent pieces of information.</p>

Given that the top of the reading literacy scale currently has no bounds, there is arguably some uncertainty about the upper limits of proficiency of extremely high performing students. However such students are likely to be capable of performing tasks characterised by the highest level of proficiency. For students who are at the bottom end of the reading literacy scale, there is a greater issue. Although it is possible to measure the reading proficiency of students performing below Level 1, at this stage their proficiency cannot be described. The independent measure of reading ease and efficiency, however, may provide additional information about those students performing below Level 1. In developing new material for PISA 2018, an effort should be made to design items that measure reading skills and understandings located at or below the current Level 1.

APPENDIX A. MAIN CHANGES IN THE READING FRAMEWORK, 2000-2015.

	2000	2009	2015
TEXT			
Format	Continuous, Non-continuous, Mixed	Same as 2000, plus Multiple	Same as 2009
Type	Argumentation, Description, Exposition, Narration, Instruction	Same as 2000, plus "Transactional"	Same as 2009
Environment	N/A	Authored, Message-based	N/A
Medium	N/A	Print, Electronic	N/A
Space	N/A	N/A	Fixed, Dynamic
SITUATIONS	Educational, Personal, Professional, Public	Same as 2000	Same as 2000
ASPECT	Access and retrieve Integrate and interpret Reflect and evaluate	Same as 2000, plus "complex"	Same as 2000

APPENDIX B. SAMPLE TASKS

Task 1. Sample of reading ease and efficiency task.

The sentence-processing items are timed tasks that require the respondent to assess whether a sentence makes sense in terms of the properties of the real world or the internal logic of the sentence. The respondent reads the sentence and circles YES if the sentence makes sense or NO if the sentence does not make sense. This task is adapted from PISA 2012 and PIAAC's Reading Components sentence processing items.

Directions: Circle YES if the sentence makes sense. Circle NO if the sentence does not make sense

The red car had a flat tire	YES	NO
Airplanes are made of dogs	YES	NO
The happy student read the book last night	YES	NO
If the cat had stayed out all night, it would not have been in the house at 2 a.m.	YES	NO
The man who is taller than the woman and the boy is shorter than both of them	YES	NO

Tasks 2-4. Sample scenario with three embedded tasks.

In this scenario, students are asked to read three sources: a blog post, the comments section that follows and an article that is referenced by one of the commenters. The articles and comments all discuss space exploration now and in the future. Students are asked to answer several questions that assess different reading processes.

Task 2. Scan and locate (single text).

PISA 2018

Unit Title: Space Exploration
Question 1/5

Refer to Scott Huffington's Blog on the right. Select a choice to answer the question.

According to Scott Huffington in the article "Is the Golden Era of Space Exploration Over?" what effect have private companies had on space exploration?

- Private companies have shown that they can better manage space exploration projects.
- People are questioning whether government space programs are necessary.
- Government agencies are losing funding to private companies offering the same service.
- Government agencies and private companies are collaborating effectively.

Text 1
Is the Golden Era of Space Exploration Over?
by Scott Huffington • May 16, 201

Beginning with the launch of Sputnik in 1957 the focus of space exploration had one aim: be the first to go where no human had gone before. In 1961 Yuri Gagarin became the first man in space sparking an intense competition where astronauts and cosmonauts battled to break records, expand frontiers, and bring notoriety to their countries of origin. However, since July 22nd 1969 and Neil Armstrong's historic leap for mankind, space exploration has slowed.

Since then, space programs have focused on creating a sustainable presence in low-Earth Orbit through the development and maintenance of space craft, space stations, and satellites. The Russian space station Mir and the US Skylab were the first space stations but proved too expensive to operate independently. We now have the International Space Station (ISS), an impressive international collaborative effort led by the United States, Russia, Canada, and Japan. Yet, the station was meant to be a stepping stone to bolder projects including a manned mission to Mars. Thirty years later, we are still maintaining the space station but we are no closer to achieving a manned mission to Mars.

For decades, the idea of human space exploration has widely been seen as the exclusive domain of government agencies like the Russian Federal Space Agency (RKA), the National Aeronautics and Space Administration (NASA) in the United States, and the European Space Agency (ESA) with 22 member countries. However, the rise of private companies making serious steps toward successful commercial space flights has many people questioning the relevance and necessity of government run and publicly funded space exploration programs. Add the highly publicized U.S. space shuttle disasters in 1986 and 2003 and the enthusiasm and commitment for space exploration has further eroded.

All of this leads me to conclude that the world has lost the focus and drive to explore new frontiers. I fear that the golden age of space exploration has passed, and we are rapidly progressing toward a decidedly Earth bound future.

Task 3: Multiple text inference

PISA 2018

Unit Title: Space Exploration
Question 2/5

Refer to both Scott Huffington's article and the comment section that followed it to answer the next question. Select all the choices that apply.

The author Scott Huffington and commenters Yoshi Kubota and Claude Messier disagree on some issues while agreeing on others.

Based on what you have learned, select the radio button for each person that would agree with the issue statement.

Issue Statement	Scott Huffington	Yoshi Kubota	Claude Messier
Enthusiasm for space exploration has decreased	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Advances resulting from space exploration have slowed in recent years	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
Both human and robotic missions are vital to space exploration programs.	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

Text 1 Text 2

COMMENTS:

Yoshi Kubota: 05/17/2015 09:42 CDT

The perception that enthusiasm and commitment for space exploration has eroded is simply false. While funding for governmental agencies has been challenging due to a sluggish worldwide economy, enthusiasm for exploration on an international level is still high. Please note that 14 space agencies from around the world came together in 2007 to draft *The Global Exploration Strategy: the Framework for Coordination*. The purpose of the framework is to create a globally coordinated vision for human and robotic space exploration. Together, our space agencies have a very clear plan for space exploration. In fact, the framework was updated in 2014. Please read the attached copy of the global exploration strategy.

Randall M. Kay: 05/18/2015 08:31 CDT

Scott, like Yoshi Kubota, I think you have misunderstood the current state of global space exploration. The ISS is allowing us to develop the skills and technology needed for deep space exploration. NASA, one of the members of the International Space Exploration Coordination Group that drafted *The Global Exploration Strategy* that Yoshi shared, has published a detailed plan for developing the technologies needed to send a manned mission to Mars. Low Earth orbit is the initial step, not the final goal.

Claude Messier: 05/19/2015 12:42 CDT

This discussion is fascinating, but I do feel the need to point out a few corrections. It is precisely through space exploration that we will solve the problems of the world. With that said, Scott does have a point that advances in the Global Exploration Strategy are slow to materialize. Part of the reason is cost, but also a flexibility built into the Global Exploration Strategy. The article referenced by Yoshi presents a strategy that the moon is our next stop. However, the organization is currently exploring two strategies: moon next and asteroid next.

Task 4. Evaluate and reflect.

PISA 2018

Unit Title: Space Exploration
Question 5/5

Refer to the articles on the right. Type your answer to the questions in the space provided.

Think about how Scott Huffington wrote his article and the commenters responded. Based on this information, write a comment that explains two primary benefits of space exploration? Support your answer with details from the articles.

Text 1 Text 2 Text 3

Is the Golden Era of Space Exploration Over?
by Scott Huffington 🍷 May 16, 201

Beginning with the launch of Sputnik in 1957 the focus of space exploration had one aim: be the first to go where no human had gone before. In 1961 Yuri Gagarin became the first man in space sparking an intense competition where astronauts and cosmonauts battled to break records, expand frontiers, and bring notoriety to their countries of origin. However, since July 22nd 1969 and Neil Armstrong's historic leap for mankind, space exploration has slowed.

Since then, space programs have focused on creating a sustainable presence in low-Earth Orbit through the development and maintenance of space craft, space stations, and satellites. The Russian space station Mir and the US Skylab were the first space stations but proved too expensive to operate independently. We now have the International Space Station (ISS), an impressive international collaborative effort led by the United States, Russia, Canada, and Japan. Yet, the station was meant to be a stepping stone to bolder projects including a manned mission to Mars. Thirty years later, we are still maintaining the space station but we are no closer to achieving a manned mission to Mars.

For decades, the idea of human space exploration has widely been seen as the exclusive domain of government agencies like the Russian Federal Space Agency (RKA), the National Aeronautics and Space Administration (NASA) in the United States, and the European Space Agency (ESA) with 22 member countries. However, the rise of private companies making serious steps toward successful commercial space flights has many people questioning the relevance and necessity of government run and publicly funded space exploration programs. Add the highly publicized U.S. space shuttle disasters in 1986 and 2003 and the enthusiasm and commitment for space exploration has further eroded.

All of this leads me to conclude that the world has lost the focus and drive to explore new frontiers. I fear that the golden age of space exploration has passed, and we are rapidly progressing toward a decidedly

REFERENCES

- American Press Institute. (2014). *How Americans get their news*. Retrieved from <http://www.americanpressinstitute.org/publications/reports/survey-research/how-americans-get-news/>
- Ananiadou, K., & Claro, M. (2009). *21st century skills and competences for new millennium learners in OECD countries* [OECD Education Working Papers, No. 41]. Paris: OECD Publishing. <http://dx.doi.org/10.1787/218525261154>
- Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education*.
- Barth, A. E., Catts, H. W., & Anthony, J. A. (2009). The component skills underlying reading fluency among adolescent readers: A latent variable approach. *Reading and Writing*, 22, 567–590. doi:10.1007/s11145-008-9125-y
- Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: A longitudinal study. *Journal of Educational Psychology*, 102, 773-785. <http://dx.doi.org/10.1037/a0020084>
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., & Rumble, M. (2010). *Draft white paper 1: Defining 21st century skills*. The University of Melbourne: Assessment and Teaching of 21st Century Skills.
- Binkley, M., Rust, K., & Williams, T. (Eds.). (1997). *Reading literacy in an international perspective*. Washington, D.C.: US Department of Education.
- Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, 21, 487-508.
- Bråten, I., Britt, M. A., Strømsø, H. I., & Rouet, J.-F. (2011). The role of epistemic beliefs in the comprehension of multiple expository texts: Toward an integrated model. *Educational Psychologist*, 46(1), 48-70.
- Bråten, I., Strømsø, H.I., & Britt, M.A. (2009). Trust matters: Examining the role of source evaluation in students' construction of meaning within and across multiple texts. *Reading Research Quarterly*, 44, 6–28.
- Britt, M. A., Goldman, S. R., & Rouet, J.-F. (Eds.). (2012). *Reading: From words to multiple texts*. New York: Routledge
- Britt, M. A., & Rouet, J.-F. (2012). Learning with multiple documents: Component skills and their acquisition. In M. J. Lawson & J. R. Kirby (Eds.), *The quality of learning: Dispositions, instruction, and mental structures*. New York: Cambridge University Press.
- Brown, A. L., Palincsar, A. S & Armbruster, B. (1984). Inducing comprehension-fostering activities in interactive learning situations. In H. Mandel, N. Stein, & T. Trabasso (Eds.), *Learning from texts*. Hillsdale, N.J.: Lawrence Erlbaum.
- Brozo, W. G., & Simpson, M. L. (2007). *Content literacy for today's adolescents: Honoring diversity and building competence* (5th ed.). Upper Saddle River, NJ: Merrill/Prentice Hall.
- Cain, K., & Oakhill, J. (2006). Assessment matters: Issues in the measurement of reading comprehension. *British Journal of Educational Psychology*, 76, 697-708.
- Cain, K., & Oakhill, J. (2007). *Children's Comprehension Problems in Oral and Written Language: A Cognitive Perspective*. New York: Guilford Press.
- Cantrell, S. C., Almasi, J. F., Carter, J. C., Rintamaa, M., & Madden, A. (2010). The impact of a strategy-based intervention on the comprehension and strategy use of struggling adolescent readers. *Journal of Educational Psychology*, 102(2), 257-280.
- Carver, R. P. (1997). Reading for one second, one minute, or one year from the perspective of reading theory. *Scientific Studies of Reading*, 1(1), 3-43.
- Chard, D. J., Pikulski, J. J., & McDonagh, S. H. (2006). Fluency: The link between decoding and comprehension for struggling readers. In T. Rasinski, C. Blanchowicz, & K. Lems (Eds.), *Fluency instruction: Research-based best practices* (pp. 39–61). New York: Guilford Press.
- Clark, C. (2014). *Children's and young people's reading in 2013. Findings from the 2013 National Literacy Trust's annual survey*. London: National Literacy Trust.
- Coiro, J., Knobel, M., Lankshear, C., & Leu, D. (2008). Central issues in new literacies and new literacies research. In J. Coiro, M. Knobel, C. Lankshear, & D. Leu (Eds.), *The handbook of research on new literacies*. Mahwah, NJ.: Erlbaum.

- Conklin, J. (1987). Hypertext: An introduction and survey. *Computer*, 20, 17-41.
- Council of Europe. (1996). *Modern languages: Learning, teaching, assessment. A common European framework of reference*. Strasbourg: CC LANG (95) 5 Rev. IV.
- Council of Europe. (2001). *The Common European framework of reference for languages: Learning, teaching, assessment*. Cambridge: Cambridge University Press.
- Cunningham, A. E., & Stanovich, K. E. (1998). Early reading acquisition and its relation to reading experience and ability 10 years later. *Developmental Psychology*, 33, 934-945.
- Darling-Hammond, L. (2001). *The right to learn: A blueprint for creating schools that work* (1st ed.). San Francisco, CA: Jossey-Bass/ Wiley.
- Dillon, A. (1994). *Designing usable electronic text: Ergonomic aspects of human information usage*. London: Taylor & Francis.
- Dreher, M. J., & Guthrie, J. T. (1990). Cognitive processes in textbook chapter search tasks. *Reading Research Quarterly*, 25, 323-339.
- Duggan, G. B., & Payne, S. J. (2009) Text skimming: the process and effectiveness of foraging through text under time pressure. *Journal of Experimental Psychology: Applied*, 15(3). 228-242.
- Eason, S., Sabatini, J., Goldberg, L. F., Bruce, K., & Cutting, L. E. (2012). Examining the relationship between word reading efficiency and oral reading rate in predicting comprehension among different types of readers. *Scientific Studies of Reading*, 17, 199-223.
- Foltz, P. W. (1996). Comprehension, coherence, and strategies in hypertext and linear text. In J. Levonen, J.-F. Rouet, A. Dillon, & R. Spiro (Eds.), *Hypertext and cognition* (pp. 109-136). Hillsdale, NJ: Lawrence Erlbaum Associates.
- Gartner. (2015). *Forecast: PCs, ultramobiles and mobile phones, worldwide, 2011-2018, 4Q14 update*. Retrieved from <http://www.gartner.com/document/2945917>.
- Gerjets, P., Kammerer, Y., & Werner, B. (2011). Measuring spontaneous and instructed evaluation processes during web search: Integrating concurrent thinking-aloud protocols and eye-tracking data. *Learning and Instruction*, 21, 220-231. doi: 10.1016/j.learninstruc.2010.02.005
- Goldman, S. R. (2004). Cognitive aspects of constructing meaning through and across multiple texts. In N. Shuart-Faris & D. Bloome, (Eds.), *Uses of intertextuality in classroom and educational research*, (pp. 317–351). Greenwich, CT: Information Age.
- Gray, W. S., & Rogers, B. (1956). *Maturity in reading*. University of Chicago Press, Chicago.
- Grisay, A., & Monseur, C. (2007). Measuring the equivalence of item difficulty in the various versions of an international test. *Studies in Educational Evaluation*, 33, 69-86.
- Guthrie, J. T. (1988). Locating information in documents: Examination of a cognitive model. *Reading Research Quarterly*, 23, 178-199.
- Guthrie, J. T., Ho, A. N., & Klauda, S. L. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly*, 48, 9-26.
- Guthrie, J. T., & Klauda, S. L. (2014). Effects of classroom practices on reading comprehension, engagement, and motivations for adolescents. *Reading Research Quarterly*, 49(4), 387-416. doi:10.1002/rrq.81
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and motivation in reading. In M. L. Kamil & P. B. Mosenthal (Eds.), *Handbook of reading research* (Vol. 3, pp. 403-422). Mahwah, NJ: Erlbaum.
- Guthrie, J. T., Wigfield, A., Metsala, J. L., & Cox, K. E. (1999). Motivational and cognitive predictors of text comprehension and reading amount. *Scientific Studies of Reading*, 3(3), 231-256. doi:10.1207/s1532799xssr0303_3
- Guthrie, J. T., Wigfield, A., & Klauda, S. L. (2012). *Adolescents' engagement in academic literacy* (Report N°7). University of Maryland, College Park. Retrieved from <http://www.corilearning.com/research-publications>.
- Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In S. L. Christensen, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 601-634). New York: Springer Science. doi:10.1007/978-1-4614-2018-7_29
- Hacker, D.J. (1998). Self-regulated comprehension during normal reading. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.) *Metacognition in educational theory and practice* (pp. 165-191). Mahwah, NJ: Erlbaum.

- Heckman, J. J., & Kautz, T. (2012). *Hard evidence on soft skills*. Bonn, Germany: Institute for the Study of Labor.
- Hofstetter, C. R., Sticht, T. G., & Hofstetter, C. H. (1999). Knowledge, literacy and power, *Communication Research*, 26, 58–80.
- Hubbard, R. (1989). Notes from the underground: Unofficial literacy in one sixth grade. *Anthropology and Education Quarterly*, 20, 291-307.
- International Telecommunications Union. (2014a). *Measuring the Information Society Report 2014*. Geneva (Switzerland): ITU. Retrieved from <http://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2014.aspx>
- International Telecommunications Union. (2014b). *Key 2005-2014 ICT data for the world, by geographic regions and by level of development* [Excel file]. Geneva (Switzerland): ITU. Retrieved from <http://www.itu.int/en/ITU-D/Statistics/Pages/publications/mis2014.aspx>
- Jenkins, J. R., Fuchs, L. S., van den Broek, P., Espin, C., & Deno, S. L. (2003). Sources of individual differences in reading comprehension and reading fluency. *Journal of Educational Psychology*, 95(4), 719-729.
- Kame'enui, E. J., & Simmons, D. C. (2001). Introduction to this special issue: The DNA of reading fluency. *Scientific Studies of Reading*, 5(3), 203-210.
- Kamil, M. J., Mosenthal, P. B., Pearson, P. D., & Barr, R. (Eds.). (2000). *Handbook of reading research* (Vol. 3). Mahwah, NJ: Lawrence Erlbaum Associates.
- Kim, Y., Park, C., & Wagner, R. (2014). Is oral/text reading fluency a “bridge” to reading comprehension? *Reading and Writing*, 27(1), 79-99.
- Kintsch, W. (1998). *Comprehension: A paradigm for cognition*. Cambridge, MA: Cambridge University Press.
- Kirsch, I. (2001). *The international adult literacy survey: Understanding what was measured*, Educational Testing Service, Princeton, NJ.
- Kirsch, I., & Mosenthal, P. B. (1990). Exploring document literacy: Variables underlying the performance of young adults. *Reading Research Quarterly*, 25(1), pp. 5-30.
- Kirsch, I., De Jong, J., Lafontaine, D., McQueen, J., Mendelovits, J., & Monseur, C. (2002). *Reading for change: Performance and engagement across countries: Results from PISA 2000*. Paris: Organisation for Economic Co-operation and Development.
- Klauda, S. L., & Guthrie, J. T. (2015). Comparing relations of motivation, engagement, and achievement among struggling and advanced adolescent readers. *Reading and Writing*. doi:10.1007/s11145-014-9523-2
- Kuhn, M. R., Schwanenflugel, P. J., & Meisinger, E. B. (2010). Aligning theory and assessment of reading fluency: Automaticity, prosody, and definitions of fluency. *Reading Research Quarterly*, 45, 230–251. doi:10.1598/RRQ.45.2.4
- Kuhn, M. R., & Stahl, S. (2003). Fluency: A review of developmental and remedial practices. *The Journal of Educational Psychology*, 95, 3-21.
- Lafontaine, D., & Monseur, C. (2006a). *Impact of item choice on the measurement of trends in educational achievement*. Paper presented at the AERA meeting, San Francisco.
- Lafontaine, D., & Monseur, C. (2006b). *Impact of test characteristics on gender equity indicators in the Assessment of Reading Comprehension*. University of Liège, Liège.
- Lai, A. S., Benjamin, R. G., Schwanenflugel, J. P., & Kuhn, R. M. (2014). The longitudinal relationship between reading fluency and reading comprehension skills in second-grade children. *Reading & Writing Quarterly: Overcoming Learning Difficulties*, 30, 116-138. DOI:10.1080/10573569.2013.789785.
- Landerl, K., & Reiter, C. (2002). Lesegeschwindigkeit als Indikator für basale Lesefertigkeiten. [Reading fluency as an indicator for basic reading skills]. In C. Wallner-Paschon & G. Haider (Eds.), *PISA Plus 2000. Thematische Analysen nationaler Projekte [PISA Plus 2000. Thematic analyses of national projects]*. Innsbruck: Studien Verlag.
- Language and Reading Research Consortium. (2015). Learning to read: Should we keep things simple? *Reading Research Quarterly*, 50, 151-169.
- Lathrop, R. (2005). *Democratic schools: Empowering students through active learning and applied civic education*. ProQuest Information and Learning Company. (UMI No. 1428169).

- Legault, L., Green-Demers, I., & Pelletier, L. (2006). Why do high school students lack motivation in the classroom? Toward an understanding of academic motivation and the role of social support. *Journal of Educational Psychology, 98*, 567-582. doi:10.1037/0022-0663.98.3.567.
- Leu, D. J., Kinzer, C. K., Coiro, J., Castek, J., & Henry, L. A. (2013). New literacies: A dual-level theory of the changing nature of literacy instruction and assessment. In D. E. Alvermann, N. J. Unrau, & R. B. Rudell (Eds.), *Theoretical models and processes of reading* (6th edition, pp. 1150-1181). Newark: International Reading Association.
- Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2015). The new literacies of online reading and comprehension: Rethinking the reading achievement gap. *Reading Research Quarterly, 50*(1), 37-59.
- Lundberg, I. (1991). Reading as an individual and social skill. In I. Lundberg & T. Høien (Eds.), *Literacy in a world of change: Perspectives on reading and reading disability*. Stavanger: Center for Reading Research/UNESCO.
- Mason, L., Boldrin, A., & Ariasi, N. (2010). Searching the Web to learn about a controversial topic: are students epistemically active? *Instructional Science, 38*, 607-633.
- McCrudden, M. T., Magliano, J., & Schraw, G. (Eds.). (2011). *Text relevance and learning from text*. Greenwich, CT: Information Age Publishing.
- McCrudden, M. T., & Schraw, G. (2007). Relevance and goal-focusing in text processing. *Educational Psychology Review, 19*(2), 113-139.
- McNamara, D. S., & Magliano, J. (2009). Toward a comprehensive model of comprehension. *Psychology of Learning and Motivation, 51*, 297-384.
- Mol, S., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*, 267-296. doi:10.1037/a0021890.
- Moore, P. (1995). Information problem-solving: A wider view of library skills. *Contemporary Educational Psychology, 20*, 1-31.
- Morgan, P., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children, 73*, 165-183.
- Morrisroe, J. (2014). *Literacy changes lives: A new perspective on health, employment and crime*. London: National Literacy Trust.
- National Reading Panel (2000). *Teaching children to read: An evidence-based assessment of the scientific research literature on reading and its implications for reading instruction*. National Institute of Child Health and Human Development.
- Naumann, J. (2015). A model of online reading engagement: Linking engagement, navigation, and performance in digital reading. *Computers in Human Behavior, 53*, 263-277.
- Nielsen, J. (1999). *Designing web usability: The practice of simplicity*. New Riders Publishing, Indianapolis.
- Naumann, J. (2015). A model of online reading engagement: Linking engagement, navigation, and performance in digital reading. *Computers in Human Behavior, 53*, 263-277. doi:10.1016/j.chb.2015.06.051
- Oakhill, J. V., Cain, K., & Bryant, P. E. (2003). The dissociation of word reading and text comprehension: Evidence from component skills. *Language and Cognitive Processes, 18*, 443-468.
- OECD. (2000). *Measuring student knowledge and skills. The PISA 2000 Assessment of Reading, Mathematical and Scientific Literacy*. Paris: OECD. Retrieved from <http://www.oecd.org/education/school/programme-for-international-student-assessment-pisa/33692793.pdf>
- OECD. (2002). *Reading for change - Performance and engagement across countries*, OECD, Paris.
- OECD. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD.
- OECD. (2007). *PISA 2006 science competencies for tomorrow's world, volume 1: Analysis*. Paris: OECD.
- OECD. (2010a). *PISA 2009 assessment framework - Key competencies in reading, mathematics and science*. Paris: Author. Retrieved from <http://www.oecd.org/pisa/pisaproducts/pisa2009assessmentframework-keycompetenciesinreadingmathematicsandscience.htm>

- OECD. (2010b). *PISA 2009 results: Learning to learn – Student engagement, strategies and practices (volume III)*. PISA/ OECD Publishing. Retrieved from http://www.oecd-ilibrary.org/education/pisa-2009-results-learning-to-learn_9789264083943-en
- OECD. (2011). PISA: Do students today read for pleasure? *PISA in Focus*, vol.8, p. 20. Paris: OECD.
- OECD. (2011). PISA 2009 results: Students on Line. Digital technologies and performance. <http://dx.doi.org/10.1787/9789264112995-en>
- OECD. (2012). The OECD Internet economy outlook. Retrieved from http://www.oecd-ilibrary.org/science-and-technology/oecd-internet-economy-outlook-2012_9789264086463-en
- OECD. (2013a). *OECD skills outlook 2013: First results from the Survey of Adult Skills*. OECD Publishing. <http://dx.doi.org/10.1787/9789264204256-en>
- OECD. (2013b). *PISA 2015 draft frameworks*. Retrieved from <http://www.oecd.org/pisa/pisaproducts/pisa2015draftframeworks.htm>
- OECD. (2014). *PISA 2012 results: What students know and can do (Volume I, Revised edition, February 2014): Student Performance in Mathematics, Reading and Science*, PISA, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264208780-en>
- OECD. (2015). *Students, computers and learning: Making the connection*. PISA, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264239555-en>
- O'Reilly, T., & Sabatini, J. (2013). *Reading for understanding: How performance moderators and scenarios impact assessment design* (ETS Research Report No. RR-13-31). Retrieved from <http://www.ets.org/Media/Research/pdf/RR-13-31.pdf>
- Ozuru, Y., Best, R., Bell, C., Witherspoon, A., & McNamara, D. S. (2007). Influence of question format and text availability on the assessment of expository text comprehension. *Cognition and Instruction*, 25, 399-438. doi:10.1080/07370000701632371
- Perfetti, C. A. (1985). *Reading Ability*. New York: Oxford University Press.
- Perfetti, C. A. (2007). Reading ability: Lexical quality to comprehension. *Scientific Studies of Reading*, 11, 357-383.
- Perfetti, C.A., Marron, M.A., & Foltz, P.W. (1996). Sources of comprehension failure: Theoretical perspectives and case studies. In C. Cornoldi & J. Oakhill (Eds.), *Reading comprehension difficulties: Processes and intervention*. Lawrence Erlbaum; Mahwah, NJ: 1996.
- Perfetti, C. A., Rouet, J.-F., & Britt, M. A. (1999). Toward a theory of documents representation. In H. van Oostendorp & S. Goldman (Eds.), *The construction of mental representations during reading* (pp. 99-122). Mahwah, NJ: Erlbaum.
- Pfost, M., Dörfler, T., & Artelt, C. (2013). Students' extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences*, 26, 89-102. doi:10.1016/j.lindif.2013.04.008
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. B. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research*, (Vol. III, pp. 545-563). NJ: L. Erlbaum.
- Rasinski, T.V., Padak, N.D., McKeon, C.A., Wilfong, L.G., Friedauer, J.A., & Heim, P. (2005). Is reading fluency a key for successful high school reading? *Journal of Adolescent and Adult Literacy*, 49, 22-27.
- Rayner, K. (1997). Understanding eye movements in reading. *Scientific Studies of Reading*, 1(4), 317.
- Rayner, K., Chace, K. H., Slattery, T. J., & Ashby, J. (2006). Eye movements as reflections of comprehension processes in reading. *Scientific Studies of Reading*, 10(3), 241-255.
- Rayner, K., Foorman, B. R., Perfetti, C. A., Pesetsky, D., & Seidenberg, M. S. (2001). How psychological science informs the teaching of reading. *Psychological Science in the Public Interest*, 2(2), 31-74.
- Reeve, J. (2012). A self-determination theory perspective on student engagement. In S.L. Christensen, A.L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149-173). New York: Springer Science. doi:10.1007/978-1-4614-2018-7_7
- Reschly, A. M., Busch, T. W., Betts, J., Deno, S. L., & Long, J. D. (2009). Curriculum-based measurement oral reading as an indicator of reading achievement: A meta-analysis of the correlational evidence. *Journal of School Psychology*, 47, 427-469.
- Richter, T., & Rapp, D. N. (2014). Comprehension and validation of text information: Introduction to the special issue. *Discourse Processes*, 51, 1-6.

- Rosenshine, B., & Meister, C. (1997). Cognitive strategy instruction in reading. In A. Stahl & A. Hayes (Eds.), *Instructional models in reading*, (pp. 85-107). Mahwah, NJ: L. Erlbaum.
- Rouet, J.-F. (2006). *The skills of document use: From text comprehension to web-based learning*. Mahwah, NJ: Lawrence Erlbaum Associates.
- Rouet, J.-F., & Britt, M. A. (2011). Relevance processes in multiple document comprehension. In M.T. McCrudden, J. P. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 19-52). Greenwich, CT: Information Age Publishing.
- Rouet, J.-F., & Britt, M. A. (2014). Learning from multiple documents. In Mayer, R.E. (Ed.) *Cambridge handbook of multimedia learning*, (2nd edition). Cambridge, MA, Cambridge University Press.
- Rouet, J.-F., & Coutelet, B. (2008). The acquisition of document search strategies in grade school students. *Applied Cognitive Psychology*, 22: 389-406. doi: 10.1002/acp.1415
- Rouet, J.-F., & Levonen, J. J. (1996). Studying and learning with nonlinear documents: Empirical studies and their implications. In J.-F. Rouet, J.J., Levonen, A.P. Dillon, & R.J. Spiro (Eds.), *Hypertext and cognition* (pp. 9-24). Mahwah, NJ: Lawrence Erlbaum Associates.
- Rouet, J.-F., Vörös, Z., & Pléh, C. (2012). Incidental learning of links during navigation: The role of visuo-spatial capacity. *Behaviour and Information Technology*, 31, 71-81.
- Routitsky, A., & Turner, R. (2003). *Item format types and their influences on cross-national comparisons of student performance*. Paper presented at the annual meeting of the American Educational Research Association (AERA). Chicago: IL.
- Rupp, A., Ferne, T., & Choi, H. (2006). How assessing reading comprehension with multiple-choice questions shapes the construct: A cognitive processing perspective. *Language Testing*, 23, 441-474.
- Sabatini, J. P., & Bruce, K. M. (2009). *PIAAC Reading Components: A conceptual framework* (OECD Educational Working paper No. 33). Retrieved from [http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=edu/wkp\(2009\)12](http://www.oecd.org/officialdocuments/publicdisplaydocumentpdf/?doclanguage=en&cote=edu/wkp(2009)12)
- Sabatini, J., Petscher, Y., O'Reilly, T., & Truckenmiller, A. (2015). Improving comprehension assessment for middle and high school students: Challenges and opportunities. In D. Reed and K. Santi (Eds.), *Improving reading comprehension of middle and high school students*, (pp. 119-151) New York: Springer.
- Sabatini, J., O'Reilly, T., Halderman, L., & Bruce, K. (2014). Broadening the scope of reading comprehension using scenario-based assessments: Preliminary findings and challenges. *International Journal Topics in Cognitive Psychology*, 114, 693-723.
- Santini, M. (2006), Web pages, text types, and linguistic features: Some issues. *International Computer Archive of Modern and Medieval English (CAME)*, 30, 67-86.
- Scammacca, N., Roberts, G., Vaughn, S., Edmonds, M., Wexler, J., Reutebuch, C.K., & Torgesen, J.K. (2006). *Interventions for adolescent struggling readers. A meta-analysis with implications for practice*. Portsmouth, NH: RMC Research Corporation, Center on Instruction.
- Schaffner, E., Philipp, M., & Schiefele, U. (2014). Reciprocal effects between intrinsic reading motivation and reading competence? A cross-lagged panel model for academic track and non-academic track students. *Journal of Research on Reading*, 00 (00), 1-18. DOI:10.1111/1467-9817.12027.
- Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, 47, 427-463.
- Schroeder, S. (2011). What readers have and do: Effects of students' verbal ability and reading time components on comprehension with and without text availability. *Journal of Educational Psychology*, 103, 877-896.
- Schwabe, F., McElvany, N., & Trendtel, M. (2015). The school age gender gap in reading achievement: Examining the influences of item format and intrinsic reading motivation. *Reading Research Quarterly*, 50(1), 1–14. doi: 10.1002/rrq.92
- Simon, H. A. (1996). *Observations on the sciences of science learning*. Paper prepared for the Committee on Developments in the Science of Learning for the Sciences of Science Learning: An Interdisciplinary Discussion. Department of Psychology, Carnegie Mellon University.
- Singer, M. (2006). Verification of text ideas during reading. *Journal of Memory and Language*, 54, 574-591.
- Smith, M. C., Mikulecky, L., Kibby, M. W., Dreher, M. J., & Dole, J. A. (2000). What will be the demands of literacy in the workplace in the next millennium? *Reading Research Quarterly*, 35(3), 378-383.

- Snow and the Rand Corporation. (2002). *Reading for understanding: Toward an R&D program in reading comprehension*. Rand Corporation Report.
- Solis, M., Miciak, J., Vaughn, S., & Fletcher, J. M. (2014). Why intensive interventions matter: Longitudinal studies of adolescents with reading disabilities and poor reading comprehension. *Learning Disability Quarterly*. Advance online publication. doi:10.1177/0731948714528806
- Spiro, J. J., Deschryver, M., Hagerman, M. S., Morsink, P., & Thompson, P. (Eds.). (2015). *Reading at a crossroads? Disjunctures and continuities in current conceptions and practices*. New York: Routledge.
- Stadtler, M., & Bromme, R. (2013). Multiple document comprehension: An approach to public understanding of science. *Cognition and Instruction*, 31, 122-129.
- Stadtler, M., & Bromme, R. (2014). The content–source integration model: A taxonomic description of how readers comprehend conflicting scientific information. In D. N. Rapp & J. Braasch (Eds.), *Processing inaccurate information: Theoretical and applied perspectives from cognitive science and the educational sciences* (pp. 379-402). Cambridge, MA: MIT Press.
- Strømsø, H. I., Bråten, I., Britt, M. A., & Ferguson, L. E. (2013). Spontaneous sourcing among students reading multiple documents. *Cognition and Instruction*, 31, 176-203.
- Torgesen, J. K., Wagner, R. K., & Rashotte, C. A. (1999). *Test of Word Reading Efficiency (TOWRE)*. Austin, TX: Pro-Ed.
- UNESCO. (2014). *Reading in the mobile era: A study of mobile reading in developing countries*. Paris: UNESCO.
- van den Broek, P., Risdén, K., & Husbye-Hartmann, E., (1995). The role of readers' standards of coherence in the generation of inferences during reading. In R. F. Lorch, Jr., & E. J. O'Brien (Eds.), *Sources of coherence in text comprehension* (pp. 353-373). Hillsdale, NJ: Erlbaum.
- van den Broek, P., Virtue, S., Everson, M., Tzeng, Y., & Sung, Y. (2002). Comprehension and memory of science texts: Inferential processes and the construction of a mental representation. In J. Otero, J. Leon, & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 131-154). Mahwah, NJ: Erlbaum.
- Vidal-Abarca, E., Mañá, A., & Gil, L. (2010). Individual differences for self-regulating task-oriented reading activities. *Journal of Educational Psychology*, 102(4), 817-826. doi: 10.1037/a0020062.
- Wagner, R.K.; Torgesen, J.; Rashotte, C.A.; Pearson, N. *Test of Sentence Reading Efficiency and Comprehension*. Pro-Ed; Austin, TX: 2010.
- Waters, H. S., & Schneider, W. (Eds.). (2010). *Metacognition, strategy use, and instruction*. New York, NY, US: Guilford Press.
- Wayman, M. M., Wallace, T., Wiley, H. I., Ticha, R., & Espin, C. A. (2007). Literature synthesis on curriculum-based measurement in reading. *The Journal of Special Education*, 41(2), 85-120.
- Werlich (1976). *A text grammar of English*. Heidelberg: Quelle and Meyer.
- White, S., Chen, J., & Forsyth, B. (2010). Reading-related literacy activities of American adults: Time spent, task types, and cognitive skills used. *Journal of Literacy Research*, 42:3, 276-307.
- Winne, P. H., & Hadwin, A. F. (1998). Studying as self-regulated learning. In D. Hacker, J. Dunlosky, & A. Graesser (Eds.) *Metacognition in educational theory and practice* (pp. 277-304). Mahwah, NJ: Erlbaum.
- Woodcock, R. W., McGrew, K. S., & Mather, N. (2001). *Woodcock-Johnson III. Tests of Achievement*. Itasca, IL: Riverside Publishing.
- Zwaan, R. A., & Singer, M. (2003). Text comprehension. In A. C. Graesser, M. A. Gernsbacher, & S. R. Goldman (Eds.), *Handbook of discourse processes* (pp. 83-122). Mahwah, NJ: Erlbaum.

II. FINANCIAL LITERACY FRAMEWORK

PISA 2018 provides an opportunity to offer an optional assessment of financial literacy for a third time. PISA 2012 was the first large-scale international study to assess the financial literacy of young people, and indicated wide variations in levels of financial literacy within and across countries. The ongoing PISA 2015 assessment and the planned assessment for 2018 will provide information about trends, as well as data from countries that have not previously participated in this option.

This framework is a revised version of that developed for the 2012 exercise. It provides the necessary detail for developing new items, designing the instrument and providing a common language with which to discuss financial literacy. It also contains a financial literacy definition for youth and organises the domain around the content, processes and contexts that are relevant for the assessment of 15-year-old students. Content areas described in the framework include *money and transactions, planning and managing finances, risk and reward and financial landscape*. The framework covers *identify financial information, analyse information in a financial context, evaluate financial issues, apply financial knowledge and understanding, and the education and work, home and family, individual and societal contexts*. These areas are illustrated with 10 items drawn from early phases of the assessment. Additionally, the framework discusses the relationship of financial literacy to non-cognitive skills and to both mathematics and reading literacy, and the measurement of students' financial behaviour and experience.

INTRODUCTION

A. Policy interest in financial literacy

In recent years, developed and emerging countries and economies have become increasingly concerned about the level of financial literacy of their citizens. This has stemmed in particular from shrinking public and private support systems, shifting demographic profiles including the ageing of the population, and wide-ranging developments in the financial marketplace. Concern was also heightened by the challenging economic and financial context with the recognition that lack of financial literacy was one of the factors contributing to ill-informed financial decisions and that these decisions could, in turn, have tremendous negative spill-overs (OECD INFE, 2009; OECD, 2009a; see also Gerardi et al., 2010, for empirical analysis of financial literacy and mortgage delinquency). As a result, financial literacy is now globally acknowledged as an important element of economic and financial stability and development. This is reflected in the G20 endorsement of the OECD/INFE (International Network on Financial Education) High-level Principles on National Strategies for Financial Education (G20, 2012; OECD INFE, 2012), the G20 leaders' call for the OECD/INFE to undertake additional work on national strategies for financial education and core competencies on financial literacy, and their statement supporting the widespread use of instruments to measure youth financial literacy including the PISA financial literacy assessment (G20, 2013).

A series of tangible trends underpin the rising global interest in financial literacy as a key life skill. These are summarised below.

a. Demographic and cultural shifts

In most countries longevity is increasing, and in many the birth-rate is falling. At the same time, female participation in the labour force and the proportion of people entering higher education are both increasing, and grown-up children are less likely to continue to live in close proximity to older family members than previous generations. The likely outcome of these shifts will be a greater need for financial security in retirement and professional care in old age, resulting in additional government expenditure (Colombo et al., 2011). Working-age adults may be expected to shoulder the tax burden to finance this expenditure whilst at the same time also saving for their own retirement, potentially repaying their own student loans, and managing increasingly varied working-life trajectories which may include periods of inactivity, self-employment or retraining.

b. Risk shift and increased individual responsibility

There has been a widespread transfer of risk from both governments and employers to individuals, meaning that now many people face the financial risks associated with longevity, investment, credit, out-of-pocket healthcare and long-term-care. The number of financial decisions that individuals have to make, and the significance of these decisions, is increasing as a consequence of changes in the market and the economy. For instance, longer life expectancy means individuals need to ensure that they accumulate savings to cover much longer periods of retirement than previous generations, despite the steadily rising age of retirement in many countries. Traditional pay-as-you-go (PAYG) public pension schemes are supplemented by private funded schemes in which the individual may be responsible for making investment decisions, including the contribution rate, the investment allocation and the type of pay out product. Moreover, defined-contribution pension plans are quickly replacing defined-benefit pension plans for new entrants, shifting onto workers the risks of uncertain investment performance and of longer life expectancy.

Even when individuals use the services of financial intermediaries and advisors, they need to be financially literate in order to understand what is being offered or advised, and to manage the products they choose. They should also be aware that some advisors may face a conflict of interest. Depending on the national legal framework for financial advice, individuals may be fully responsible for the financial product they decide to purchase, facing all the direct consequences of their choice.

Surveys show that a majority of workers are unaware of the risks they now have to face, and have neither sufficient financial knowledge nor the skills to manage such risks adequately, even if they are aware of them (OECD, 2008; Money and Pensions panel, 2013; Barrett et al, 2013).

c. Increased supply of a wide range of financial products and services

In addition, in all countries, growing numbers of consumers have access to a wide range of financial products and services from a variety of providers and delivered through various channels³. Improved levels of financial inclusion in emerging economies, as well as worldwide developments in technology and deregulation have resulted in widening access to all kinds of financial products, from current accounts and remittances products to revolving credit and equity portfolios. The products available are also becoming more complex, and individuals are required to make comparisons across a number of factors such as the fees charged, interest rates paid or received, length of contract and exposure to risk. They must also identify appropriate providers and delivery channels from the vast array of possibilities, including community groups, traditional financial institutions, online banks and mobile phone companies.

d. Increased demand for financial products and services

Economic and technological developments have brought greater global connectedness and massive changes in both the methods and frequency of communications and financial transactions, as well as in social interactions and consumer behaviour. Such changes have made it more important that individuals be able to interact with financial providers and their intermediaries. In particular, consumers often need access to financial services (including banks and other providers such as post offices) in order to make and receive electronic payments like income, remittances and online transactions, and even to conduct face-to-face transactions in instances where cash or cheques are no longer favoured. Together, these trends have transferred the responsibility of major financial decisions to individuals, enlarged the options for the majority of the population (including new financial consumers) and increased the level of complexity they face. Against this backdrop, individuals are expected to be sufficiently financially literate to take the necessary steps to protect themselves and their relatives and ensure their financial well-being.

B. Expected benefits of financial education and improved levels of financial literacy

Existing empirical evidence shows that young people and adults in both developed and emerging economies who have been exposed to good quality financial education are subsequently more likely than others to plan ahead, save and engage in other responsible financial behaviours (Atkinson et al. 2015 forthcoming; Bruhn et al. 2013; Miller et al. 2014). This evidence suggests a possible causal link between financial education and outcomes and indicates that improved levels of financial literacy can lead to positive behaviour change.

Other research indicates a number of potential benefits of being financially literate. There is mounting evidence that in developed countries those with higher financial literacy are better able to manage their money, participate in the stock market and perform better on their portfolio choice, and that they are more likely to choose mutual funds with lower fees (Hastings and Tejada-Ashton, 2008; Hilgert et al., 2003; Lusardi and Mitchell, 2008, 2011; Stango and Zinman, 2009; van Rooij et al., 2011; Yoong, 2011). In emerging economies, financial literacy is shown to be correlated with holding basic financial products like bank accounts and insurance take-up (OECD/INFE, 2013; Xu and Zia, 2012); similarly, 15-year-old students with bank accounts have higher levels of financial literacy than those without on average across the OECD countries participating in the 2012 PISA exercise (OECD, 2014c). Moreover, adults who have greater

³ Financial inclusion has increased from 51% of the adult population with an account at a financial institution or mobile money service in 2011, to 62% in 2014. However, two billion adults remain unbanked (Demirguc-Kunt et al., 2015).

financial knowledge are more likely to accumulate higher amounts of wealth (Lusardi and Mitchell, 2011).

Higher levels of financial literacy have been found to be related not only to asset building but also to debt and debt management, with more financially literate individuals opting for less costly mortgages and avoiding high interest payments and additional fees (Gerardi et al., 2010; Lusardi and Tufano, 2009a, 2009b; Moore, 2003).

In addition to the benefits identified for individuals, widescale financial literacy can be expected to improve economic and financial stability for a number of reasons (OECD, 2005). Financially literate consumers can make more informed decisions and demand higher quality services, which can, in turn, encourage competition and innovation in the market. As they can protect themselves to a greater extent against income or expenditure shocks and are less likely to default on credit commitments, macro level shocks are likely to have a lower impact on financially literate populations. Financially literate consumers are also less likely to react to market conditions in unpredictable ways, less likely to make unfounded complaints and more likely to take appropriate steps to manage the risks transferred to them. All of these factors can lead to a more efficient financial services sector. They can also ultimately help in reducing government aid (and taxation) aimed at assisting those who have taken unwise financial decisions – or no decision at all.

Box 7. OECD activities in relation to financial education

In 2002, the OECD initiated a far-reaching financial education project to address governments' emerging concerns about the potential consequences of low levels of financial literacy. This project is serviced by the OECD Committee on Financial Markets (CMF) and the Insurance and Private Pensions Committee (IPPC) in coordination with other relevant bodies including the Education Policy Committee (EDPC) on issues related to schools. The project takes a holistic approach to financial-consumer issues that highlights how, alongside improved financial access, adequate consumer protection and regulatory frameworks, financial education has a complementary role to play in promoting the outcome of financial literacy.

One of the first milestones of the financial education project was the adoption of the *Recommendation on Principles and Good Practices for Financial Education and Awareness* by the OECD Council (OECD, 2005a). Alongside these recommendations, the publication *Improving Financial Literacy: Analysis of Issues and Policies* details the reasons for focusing on financial education, and provides a first international overview of financial education work being undertaken in various countries (OECD, 2005b). This book also includes principles and good practices for policy makers and other stakeholders seeking to improve levels of financial literacy in their country. It is complemented by a global clearinghouse on financial education, the OECD International Gateway for Financial Education (www.financial-education.org/home.html), which gathers data, resources, research and news on financial education issues and programmes from around the world.

Recognising the increasingly global nature of financial literacy and education issues, in 2008 the OECD created the International Network on Financial Education (INFE) to benefit from and encompass the experience and expertise of developed and emerging economies. More than 240 public institutions from more than 110 countries are economies are members of the INFE (2015 figures). Members meet twice yearly to discuss the latest developments in their country, share their expertise, and collect evidence, as well as to develop analytical and comparative studies, methodologies, good practice, policy instruments and practical guidance on key priority areas.

Financial education for youth and in schools

The 2005 OECD Recommendation advised that “financial education should start at school. People should be educated about financial matters as early as possible in their lives” (OECD, 2005a). Two main reasons underpin the OECD recommendation: the importance of focusing on youth in order to provide them with key life skills before they start to become active financial consumers, and the relative efficiency of providing financial education in schools rather than attempting remedial actions in adulthood.

At the time the OECD Recommendation was published, there was a lack of guidance on ways of implementing financial education initiatives for youth and in schools. The OECD/INFE therefore subsequently created a dedicated expert subgroup in order to develop policy and practical tools. The resulting publication was welcomed by G20 leaders in September 2013 (OECD, 2014b). The publication includes guidelines for financial education in schools and guidance on financial education learning frameworks, which have also been supported by the Ministers of Finance of the Asia-Pacific Economic Cooperation (APEC) in August 2012.

Youth are increasingly seen as an important target group for financial education. A survey of individual financial literacy schemes supported by the European Commission (Habschick et al., 2007) found that most were directed at children and young people, and stock-take exercises launched by the OECD/INFE demonstrated that many OECD and non-OECD countries have developed or are developing programmes in schools to varying extents (OECD 2014b; Messy and Monticone, 2015 forthcoming, 2016 forthcoming).

a. Focus on youth

People form habits and behaviours from a young age, learning from their parents and others around them, indicating the importance of early interventions to help shape beneficial behaviours and attitudes (Whitebread and Bingham, 2013). Furthermore, young people need financial knowledge and understanding from an early age in order to operate within the complex financial landscape they are likely to find themselves in, often before reaching adulthood. Younger generations are not only likely to face ever-increasing complexity in financial products, services and markets, but as noted above, they are more likely to have to bear more financial risks in adulthood than their parents. In particular, as the previous discussion illustrates, they are likely to bear more responsibility for the planning of their own retirement savings and investments, and the coverage of their healthcare needs; and they will have to deal with more sophisticated and diverse financial products.

Young people may learn beneficial behaviours from their friends and family, such as prioritising their expenditure or putting money aside for a rainy day, but the recent changes in the financial marketplace and social welfare systems mean it is unlikely that they can gain sufficient knowledge or information from such people unless they work in related fields⁴. The majority of young people will have to apply their skills to search for information and solve problems, and know when to make informed use of professional financial advice. Efforts to improve financial knowledge in the workplace or in other settings can be severely limited by a lack of early exposure to financial education and by a lack of awareness of the benefits of continuing financial education. It is therefore important to provide early opportunities for establishing the foundations of financial literacy.

In addition to preparing young people for their adult life, financial education for youth and in schools can also address the immediate financial issues facing young people. Children are often consumers of financial services from a young age. The results of the 2012 PISA financial literacy assessment revealed that, on average across the 13 OECD participating countries and economies, almost 60% of 15-year-old students have a bank account (OECD, 2014c). Moreover, it is not uncommon for them to have accounts with access to online payment facilities or to use mobile phones (with various payment options) even before they become teenagers, and it is clear that financial literacy skills would be of benefit to them when using such products. Before leaving school, they may also need to make decisions about issues such as scooter or car insurance, savings products and overdrafts.

In many countries, at around the age of 15 to 18, young people (and their parents) face one of their most important financial decisions: that is, whether or not to invest in tertiary education. The gap in wages between college and non-college educated workers has widened in many economies (OECD, 2014a). At the same time, the education costs borne by students and their families have increased, often resulting in large student loans to repay, and potentially leading towards a reliance on credit (Bradley, 2012; OECD, 2014b; Ratcliffe and McKernan, 2013; Smithers, 2010).

b. Efficiency of providing financial education in schools

Research suggests that there is a link between financial literacy and family economic and educational background: those who are more financially literate disproportionately come from highly educated families that hold a wide range of financial products (Lusardi et al., 2010). Results of the 2012 PISA financial literacy assessment show that on average across OECD countries and economies, 14% of the variation in student performance in financial literacy within each country and economy is associated with their families' economic, social and cultural status, and that students with at least one parent who has tertiary level education have higher scores, on average, than other students (OECD, 2014c). In order to provide equality of opportunity, it is important to offer financial education to those who would not otherwise have access to it. Schools are well positioned to advance financial literacy among all demographic groups and reduce financial literacy gaps and inequalities (including across generations).

Recognising both the importance of financial literacy for youth and the unique potential to create more skilled and knowledgeable future generations, an increasing number of countries have embarked on the development of financial education programmes for children and young people. These are either dedicated to youth generally or to (some part of) the school population, and include those at national, regional and local levels as well as pilot exercises.

⁴ PISA 2012 indicates that students with a parent working in the financial services sector have higher levels of financial literacy on average, although the data are only available for a limited number of countries.

C. The need for data

Policy makers, educators and researchers need high-quality data on levels of financial literacy in order to inform financial education strategies and the implementation of financial education programmes in schools by identifying priorities and measuring change across time.

Several countries have undertaken national surveys of financial literacy across their adult population and the OECD has developed a questionnaire designed to capture levels of financial literacy amongst adults at an international level, which was first piloted in 2010 and is currently being used for a second international comparative study (Atkinson and Messy, 2012; OECD INFE, 2011; OECD/INFE, 2015). However, until financial literacy was included in the 2012 PISA assessment, there were few data collection efforts on the levels of financial literacy amongst young people under the age of 18, and none that could be compared across countries.

A robust measure of financial literacy amongst young people provides information at a national level that can indicate whether the current approach to financial education is effective. In particular, it can help to identify issues that need addressing through schools or extra curricula activities or programmes that will enable young people to be properly and equitably equipped to make financial decisions in adulthood. It can also be used as a baseline from which to measure success and review school and other programmes in future years.

An international study provides additional benefits to policy makers and other stakeholders. Comparing levels of financial literacy across countries makes it possible to see which countries have the highest levels of financial literacy and begin to identify particularly effective national strategies and good practices. It also makes it possible to recognise common challenges and explore the possibility of finding international solutions to the issues faced.

Against this backdrop, the collection of robust and internationally comparable financial literacy data in the student population provides policy makers, educators, curriculum and resource developers, researchers, and others with:

- international evidence on how young people are distributed across the financial literacy proficiency scale, which can be used to inform the development of more targeted programmes and policies;

- an opportunity to compare financial education strategies across countries and explore good practice; and, ultimately,

- comparable data over time to track trends in financial literacy and potentially assess the association between financial literacy and the availability of financial education in schools.

There are other advantages to be gained from an international measurement exercise on financial literacy. In particular, the development of a financial literacy assessment framework that is applicable across countries provides national authorities with detailed guidance about the scope and operational definition of financial literacy without having to fund national studies. As noted in the article “Financial Literacy and Education Research Priorities”, there has been a gap in the research on financial literacy related to the lack of consistency among researchers in how to define and measure programme success. There is a need for researchers to develop a clear understanding of what it means to be “financially educated” (Schuchardt et al., 2009).

D. The measurement of financial literacy in PISA

PISA 2012 was the first large-scale international study to assess the financial literacy of young people. PISA assesses the readiness of students for their life beyond compulsory schooling, and, in particular, their capacity to use knowledge and skills, by collecting and analysing cognitive and other information from 15-year-olds in many countries and economies.

PISA financial literacy data provides a rich set of comparative data that policy makers and other stakeholders can use to make evidence-based decisions. International comparative data on financial literacy can answer questions such as, “How well are young people prepared for the new financial systems that are becoming more global and more complex?” and “Who are the leaders in terms of financial literacy amongst youth?”

As with the core PISA domains of reading, mathematics and science, the main focus of the financial literacy assessment in PISA is on measuring the proficiency of 15-year-old students in demonstrating and applying knowledge and skills. And like other PISA domains, financial literacy is assessed using an instrument designed to provide data that are valid, reliable and interpretable.

The PISA financial literacy assessment framework developed in 2012 (OECD, 2013) provided a first step in constructing an assessment that satisfies these three broad criteria. The main benefit of constructing an assessment framework is improved measurement, as it provides an articulated plan for developing the individual items and designing the instrument that will be used to assess the domain. A further benefit is that it provides a common language for discussion of the domain, and thereby increases understanding of what is being measured. It also promotes an analysis of the kinds of knowledge and skills associated with competency in the domain, thus providing the groundwork for building a described proficiency scale or scales that can be used to interpret the results.

The development of the PISA frameworks can be described as a sequence of the following six steps:

- development of a definition for the domain and a description of the assumptions that underlie that definition;
- identification of a set of key characteristics that should be taken into account when constructing assessment tasks for international use;
- operationalisation of the set of key characteristics that will be used in test construction, with definitions based on existing literature and experience in conducting other large scale assessments;
- evaluation of how to organise the set of tasks constructed in order to report to policy makers and researchers on achievement in each assessment domain for 15-year-old students in participating countries;
- validation of the variables and assessment of the contribution each makes to understanding task difficulty across the various participating countries; and
- preparation of a described proficiency scale for the results.

The framework revision undertaken ahead of the 2018 assessment maintains the definition for the domain whilst updating the operationalisation of the domain to ensure it is in line with recent developments in financial markets and the latest research findings.

1. DEFINING FINANCIAL LITERACY

In developing a working definition of financial literacy that can be used to lay down the groundwork for designing an international financial literacy assessment, the Financial Literacy Expert Group (FEG) looked both to existing PISA domain definitions of literacies, and to articulations of the nature of financial education.

PISA conceives of literacy as the capacity of students to apply knowledge and skills in key subject areas and to analyse, reason and communicate effectively as they pose, solve and interpret problems in a variety of situations. PISA is forward looking, focusing on young people's ability to use their knowledge and skills to meet real-life challenges, rather than merely on the extent to which they have mastered specific curricular content (OECD, 2010a).

In its *Recommendation on Principles and Good Practices for Financial Education and Awareness*, the OECD defined financial education as “the process by which financial consumers/investors improve their understanding of financial products, concepts and risks and, through information, instruction and/or objective advice, develop the skills and confidence to become more aware of financial risks and opportunities, to make informed choices, to know where to go for help, and to take other effective actions to improve their financial well-being” (OECD, 2005a).

The FEG agreed that “understanding”, “skills” and the notion of applying understanding and skills (“effective actions”) were key elements of this definition. It was recognised, however, that the definition of financial education describes a process – education – rather than an outcome. What was required for the assessment framework was a definition encapsulating the outcome of that process in terms of competency or literacy.

The definition of financial literacy for PISA is as follows:

Financial literacy is knowledge and understanding of financial concepts and risks, and the skills, motivation and confidence to apply such knowledge and understanding in order to make effective decisions across a range of financial contexts, to improve the financial well-being of individuals and society, and to enable participation in economic life.

This definition, like other PISA domain definitions, has two parts. The first part refers to the kind of thinking and behaviour that characterises the domain. The second part refers to the purposes for developing the particular literacy.

In the following paragraphs, each part of the definition of financial literacy is considered in turn to help clarify its meaning in relation to the assessment.

- *Financial literacy...*

Literacy is viewed as an expanding set of knowledge, skills and strategies, which individuals build on throughout life, rather than as a fixed quantity, a line to be crossed, with illiteracy on one side and literacy on the other. Literacy involves more than the reproduction of accumulated knowledge, although measuring prior financial knowledge is an important element in the assessment. It also involves a mobilisation of cognitive and practical skills, and other resources such as attitudes, motivation and values. The PISA assessment of financial literacy draws on a range of knowledge and skills associated with development of the capacity to deal with the financial demands of everyday life and uncertain futures within contemporary society.

- *...is knowledge and understanding of financial concepts and risks...*

Financial literacy is thus contingent on some knowledge and understanding of fundamental elements of the financial world, including key financial concepts as well as the purpose and basic features of financial products. This also includes risks that may threaten financial well-being as well as insurance policies and pensions. It can be assumed that 15-year-olds are beginning to acquire this knowledge and gain experience of the financial environment that they and their families inhabit and the main risks they face. All of them are likely to have been shopping to buy household goods or personal items; some will have taken part in family discussions about money and whether what is wanted is actually needed or affordable; and a sizeable proportion of them will have already begun to earn and save money. Some students already have experience of financial products and commitments through a bank account or a mobile phone contract. A grasp of concepts such as interest, inflation, and value for money are soon going to be, if they are not already, important for their financial well-being.

- *...and the skills,...*

These skills include generic cognitive processes such as accessing information, comparing and contrasting, extrapolating and evaluating – applied in a financial context. They include basic skills in mathematical literacy such as the ability to calculate a percentage, undertake basic mathematical operations or convert from one currency to another, and language skills such as the capacity to read and interpret advertising and contractual texts.

- *...motivation and confidence...*

Financial literacy involves not only the knowledge, understanding and skills to deal with financial issues, but also non-cognitive attributes: the motivation to seek information and advice in order to engage in financial activities, the confidence to do so and the ability to manage emotional and psychological factors that influence financial decision making. These attributes are considered as a goal of financial education, as well as being instrumental in building financial knowledge and skills.

- *...to apply such knowledge and understanding in order to make effective decisions...*

PISA focuses on the ability to activate and apply knowledge and understanding in real-life situations rather than the reproduction of knowledge. In assessing financial literacy, this translates into a measure of young people's ability to transfer and apply what they have learnt about personal finance into effective decision-making. The term "effective decisions" refers to informed and responsible decisions that satisfy a given need.

- ...*across a range of financial contexts...*

Effective financial decisions apply to a range of financial contexts that relate to young people's present daily life and experience, but also to steps they are likely to take in the near future as adults. For example, young people may currently make relatively simple decisions such as how they will use their pocket money or, at most, which mobile phone contract they will choose; but they may soon be faced with major decisions about education and work options with long-term financial consequences.

- ...*to improve the financial well-being of individuals and society...*

Financial literacy in PISA is primarily conceived of as literacy around personal or household finance, distinguished from economic literacy, which includes concepts such as the theories of demand and supply, market structures and so on. Financial literacy is concerned with the way individuals understand, manage and plan their own and their households' – which often means their families' – financial affairs. It is recognised, however, that good financial understanding, management and planning on the part of individuals has some collective impact on the wider society, in contributing to national and even global stability, productivity and development.

- ...*and to enable participation in economic life.*

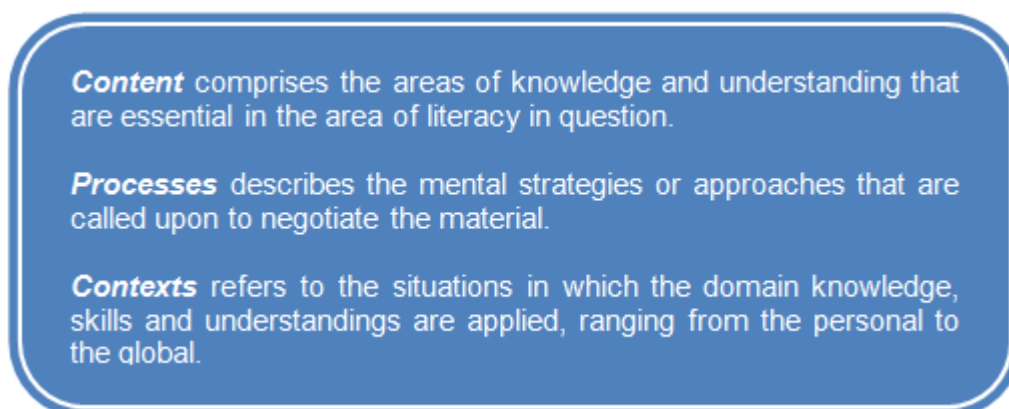
Like the other PISA literacy definitions, the definition of financial literacy implies the importance of the individual's role as a thoughtful and engaged member of society. Individuals with a high level of financial literacy are better equipped to make decisions that are of benefit to themselves, and also to constructively support and critique the economic world in which they live.

2. ORGANISING THE DOMAIN

How the domain is represented and organised determines the assessment design, including item development and, ultimately, the evidence about student proficiencies that can be collected and reported. Many elements are part of the concept of financial literacy, not all of which can be taken into account and varied in an assessment such as PISA. It is necessary to select the elements that will best ensure construction of an assessment comprising tasks with an appropriate range of difficulty and a broad coverage of the domain.

A review of approaches and rationales adopted in previous large-scale studies, and particularly in PISA, shows that most consider the relevant content, processes and contexts for assessment as they specify what they wish to assess. Content, processes and contexts can be thought of as three different perspectives on the area to be assessed, as shown in Figure 3.

Figure 3. A model for organising the domain for an assessment framework



The steps of identifying and weighting the different categories within each perspective, and then ensuring that the set of tasks in the assessment adequately reflects these categories, are used to ensure coverage and validity of the assessment. The three perspectives are also helpful in thinking about how achievement in the area is to be reported.

The following section presents a discussion of each of the three perspectives and the framework categories into which they are divided. The section includes the types of tasks that a student may be asked to complete. Examples of items drawn from the PISA 2012 field trial are included to illustrate the perspectives and categories. While they are representative of those used in the main survey, these particular items are not used in the assessment instrument: only secure, unpublished items are used for this purpose, to protect the integrity of the data that is collected to measure student proficiency.

A. Content

The content of financial literacy is conceived of as the areas of knowledge and understanding that must be drawn upon in order to perform a particular task. A review of the content of existing financial literacy learning frameworks from Australia, Brazil, England, Japan, Malaysia, the Netherlands, New Zealand, Northern Ireland, Scotland, South Africa and the United States indicated that there is some consensus on the financial literacy content areas (OECD 2014b). The review showed that the content of financial education in schools was – albeit with cultural differences – relatively similar, and that it was possible to identify a series of topics commonly included in these frameworks. These form the four content areas for PISA financial literacy: *money and transactions*, *planning and managing finances*, *risk and reward*, and *financial landscape*. Further work, undertaken by the OECD/INFE to develop a core competencies framework on financial literacy for youth provides additional guidance on how these content areas map to desired outcomes (OECD, 2015 forthcoming).

a. Money and transactions

This content area includes awareness of the different forms and purposes of money and managing monetary transactions, which may include actions such as spending or making payments taking into account value for money, and the use of bank cards, cheques, bank accounts and currencies. It also covers practices such as taking care of cash and other valuables, calculating value for money, and filing documents and receipts.

Tasks in this content area can, for example, ask students to show that they:

- Are aware of the different forms and purposes of money:
 - recognise bank notes and coins;
 - understand that money can be exchanged for goods and services;
 - identify different ways to pay for items purchased in person or at a distance (from a catalogue, or online for example);
 - recognise that there are various ways of receiving money from other people and transferring money between people or organisations, such as cash, cheques, card payments in person or online or electronic transfers online or via SMS; and
 - understand that money can be borrowed or lent, and the purpose of interest (taking into account that the payment and receipt of interest is forbidden in some religions).
- Are confident and capable at handling and monitoring transactions:
 - can use cash, cards and other payment methods to purchase items;
 - can use cash machines to withdraw cash or to get an account balance;
 - can calculate the correct change;
 - can work out which of two consumer items of different sizes would give better value for money, taking into account the individual's specific needs and circumstances; and
 - can check transactions listed on a bank statement and note any irregularities.

The following example from the unit *AT THE MARKET* illustrates a task that requires students to apply the concept of value for money. In this question, and in many others, the unit of currency is the imaginary Zed. PISA questions often refer to situations that take place in the fictional country of Zedland, where the Zed is the unit of currency. This artifice (about which students are informed at the beginning of the testing session) has been introduced to enhance comparability across countries.

Figure 4. Item for the unit **AT THE MARKET**

You can buy tomatoes by the kilogram or by the box.

QUESTION

The box of tomatoes is better value for money than the loose tomatoes.

Using an everyday context, shopping for groceries, this item addresses the basic concept of value for money. Questions about the buying of goods are generally categorised as being in the content area of money and transactions. To gain credit for this item, students have to show they have compared the two ways of buying tomatoes using a common point of comparison. Examples of answers that would receive credit are:

It is 2.75 zeds/kg for loose tomatoes but only 2.2 zeds/kg for boxed tomatoes.

Because 10 kg of loose tomatoes would cost 27.50 zeds.

You get more tomato for each zed you spend when buying the box.

In the field trial, three-quarters of all students were able to analyse the information and explain that the price per kilogram of the boxed tomatoes is less than the price per kilogram of the loose tomatoes.

b. Planning and managing finances

Income, expenditure and wealth need planning and managing over both the short term and long term. This content area therefore reflects the process of managing, planning and monitoring income and expenses and understanding ways of enhancing wealth and financial well-being. It includes content related to credit use as well as savings and wealth creation.

This content area includes:

- Knowledge and ability to monitor and control income and expenses:
 - identify various types of income (e.g. allowances, salary, commission, benefits,) and ways of discussing income (such as hourly wage and gross or net annual income) and
 - draw up a budget to plan regular spending and saving and live within it.

- Knowledge and ability to make use of income and other available resources in the short and long terms to enhance financial well-being:
 - understand how to manipulate various elements of a budget, such as identifying priorities if income does not meet planned expenses, or finding ways to increase savings, such as reducing expenses or increasing income;
 - assess the impact of different spending plans and be able to set spending priorities in the short and long term;
 - plan ahead to pay future expenses: for example, working out how much money needs to be saved each month to make a particular purchase or pay a bill;
 - understand the purposes of accessing credit and the ways in which expenditure can be smoothed over time through borrowing or saving;
 - understand the idea of building wealth, the impact of compound interest on savings, and the pros and cons of investment products;
 - understand the benefits of saving for long term goals or anticipated changes in circumstance (such as living independently); and understand how government taxes and benefits impact on personal and household finances.

The example *SPENDING CHOICES* presented below illustrates an item addressing *planning and managing finances* in a context that is relevant to 15-year-olds as they think about their lives in the near future.

- *Illustrative PISA Financial literacy item 2 – SPENDING CHOICES*

Figure 5. Item for the unit *SPENDING CHOICES*

Claire and her friends are renting a house.
They have all been working for two months.
They do not have any savings.
They are paid monthly and have just received their wages.
They have made this “To do” list.

To do
Get cable TV
Pay the rent
Buy outdoor furniture

➤ **QUESTION**

- Which of the tasks on the list are likely to need prompt attention from Claire and her friends?

Circle “Yes” or “No” for each task.

Task	Is the task likely to need prompt attention?
Get cable TV	Yes / No
Pay the rent	Yes / No
Buy outdoor furniture	Yes / No

This question in *SPENDING CHOICES* asks students to evaluate spending priorities for the home when working within a budget, distinguishing between wants and needs. It therefore falls under the content area of *planning and managing finances*. Over three-quarters of all students in the field trial scored full credit on this item by circling “No”, “Yes” and “No” in that order, thus identifying that, of the three tasks, only paying the rent requires prompt attention for Claire and her housemates.

Another instance from the *planning and managing finances* content category is provided later, in the example *TRAVEL MONEY* which asks students to plan spending and saving in order to pay a future expense.

c. Risk and reward

Risk and reward is a key area of financial literacy, incorporating the ability to identify ways of balancing and covering risks and managing finances in uncertainty and an understanding of the potential for financial gains or losses across a range of financial contexts. There are two types of risk of particular importance in this domain. The first relates to financial losses that an individual cannot bear, such as those caused by catastrophic or repeated costs. The second is the risk inherent in financial products, such as credit agreements with variable interest rates, or investment products. This content area therefore includes knowledge of the types of products that may help people to protect themselves from the consequences of negative outcomes, such as insurance and savings, as well as being able to make a general assessment of the level of risk and reward related to different products, purchases, behaviours or external factors.

This content category includes:

- Recognising that certain financial products (including insurance) and processes (such as saving) can be used to manage and offset various risks (depending on different needs and circumstances):
 - knowing how to assess whether certain insurance policies may be of benefit.
- Applying knowledge of the benefits of contingency planning, diversification and the dangers of default on payment of bills and credit agreements to decisions about:
 - limiting the risk to personal capital;
 - various types of investment and savings vehicles, including formal financial products and insurance products, where relevant; and
 - various forms of credit, including informal and formal credit, unsecured and secured, rotating and fixed term, and those with fixed or variable interest rates.
- Knowing about and managing risks and rewards associated with life events, the economy and other external factors, such as the potential impact of:
 - theft or loss of personal items, job loss, birth or adoption of a child, deteriorating health or mobility;
 - fluctuations in interest rates and exchange rates; and other market changes.
- Knowing about the risks and rewards associated with substitutes for financial products; in particular:
 - saving in cash, or buying property, livestock or gold as a store of wealth; and
 - taking credit or borrowing money from informal lenders.
- Knowing that there may be unidentified risks and rewards associated with new financial products (examples may currently include innovative digital finance or crowd funding', but by definition, such list will change over time).

An illustration from the *risk and reward* content category is provided in the example **MOTORBIKE INSURANCE**.

- *Illustrative PISA Financial literacy item 3 – MOTORBIKE INSURANCE*

Figure 6. Item for the unit *MOTORBIKE INSURANCE*

Last year, Steve's motorbike was insured with the PINSURA insurance company.

The insurance policy covered damage to the motorbike from accidents and theft of the motorbike.

➤ **QUESTION**

- Steve plans to renew his insurance with PINSURA this year, but a number of factors in Steve's life have changed since last year.
- How is each of the factors in the table likely to affect the cost of Steve's motorbike insurance this year?

Circle "Increases cost", "Reduces cost" or "Has no effect on cost" for each factor.

Factor	How is the factor likely to affect the cost of Steve's insurance?
Steve replaced his old motorbike with a much more powerful motorbike	Increases cost / Reduces cost / Has no effect on cost
Steve has painted his motorbike a different colour	Increases cost / Reduces cost / Has no effect on cost
Steve was responsible for two road accidents last year	Increases cost / Reduces cost / Has no effect on cost

Motorbike insurance falls under the content area of *risk and reward* because insurance is a product designed specifically to protect individuals against risks and financial losses that they would not otherwise be able to bear. The question relies on students understanding that the higher their risk exposure is with regards to measurable criteria, the more it will cost them to buy appropriate insurance. Half of all students scored full credit on this question in the field trial by recognising that the first and third factors increase the cost of the insurance while the second factor has no effect.

Another illustration of the *risk and reward* content category is provided in the example *SHARES*, which requires students to consider information about an investment which has fluctuated in value over time.

d. Financial landscape

This content area relates to the character and features of the financial world. It covers awareness of the role of regulation and consumer protection, knowing the rights and responsibilities of consumers in the financial marketplace and within the general financial environment, and the main implications of financial contracts. Information resources are also topics relevant to this content area. In its broadest sense, *financial landscape* also incorporates an understanding of the consequences of changes in economic conditions and public policies, such as changes in interest rates, inflation, taxation or welfare benefits for individuals, households and society. Tasks associated with this content area include:

Awareness of the role of regulation and consumer protection

- Knowledge of rights and responsibilities, and the ability to apply it to:
 - understand that buyers and sellers have rights, such as being able to apply for redress;
 - understand that buyers and sellers have responsibilities, such as:
 - consumers/investors giving accurate information when applying for financial products;
 - providers disclosing all material facts; and
 - consumers/investors being aware of the implications of one of the parties not doing so.
 - recognise the importance of the legal documentation provided when purchasing financial products or services and the importance of understanding the content.
- Knowledge and understanding of the financial environment, including:
 - identifying which providers are trustworthy, and which products and services are protected through regulation or consumer protection laws;
 - identifying whom to ask for advice when choosing financial products, and where to go for help or guidance in relation to financial matters; and
 - awareness of existing financial crimes such as identity theft and scams, knowledge of how to take appropriate precautions to protect personal data and avoid other scams, and knowledge of their rights and responsibilities in the event that they are a victim.
 - awareness of the potential for new forms of financial crime and alertness to the risks.
- Knowledge and understanding of the impact of their own financial decisions on themselves and others:
 - understand that individuals have choices in spending and saving and each action can have consequences for the individual and for society; and
 - recognise how personal financial habits, actions and decisions impact at an individual, community, national and international level.
- Knowledge of the influence of economic and external factors:
 - aware of the economic climate and understand the impact of policy changes such as reforms related to the funding of post-school training or compulsory savings for retirement;
 - understand how the ability to build wealth or access credit depends on economic factors such as interest rates, inflation and credit scores; and
 - understand that a range of external factors, such as advertising and peer pressure, can affect individuals' financial choices and outcomes.

An illustration of an item that reflects financial landscape by focusing on financial crimes is presented in the example *BANK ERROR*.

- Illustrative PISA Financial literacy item 4 – BANK ERROR

Figure 7. Item for the unit **BANK ERROR**

David banks with ZedBank. He receives this e-mail message.

Dear ZedBank member,

There has been an error on the ZedBank server and your Internet login details have been lost.

As a result, you have no access to Internet banking.

Most importantly your account is no longer secure.

Please click on the link below and follow the instructions to restore access. You will be asked to provide your Internet banking details.

<https://ZedBank.com>

➤ **QUESTION**

- Which of these statements would be good advice for David?

Circle “Yes” or “No” for each statement.

Statement	Is this statement good advice for David?
Reply to the e-mail message and provide his Internet banking details	Yes / No
Contact his bank to inquire about the e-mail message	Yes / No
If the link is the same as his bank’s website address, click on the link and follow the instructions	Yes / No

Internet banking is part of the broader *financial landscape* in which students are likely to participate, either now or in the near future. In this environment they may be exposed to financial fraud. *BANK ERROR* investigates whether they know how to take appropriate precautions. In this question, students are asked to respond appropriately to a financial scam e-mail message. They must evaluate the presented options and recognise that the second piece of advice is the only one that can be considered good advice. In the field trial, just over 40% of the students gained full credit for this item, by responding “No”, “Yes”, “No” in that order.

B. Processes

The process categories relate to cognitive processes. They are used to describe students' ability to recognise and apply concepts relevant to the domain, and to understand, analyse, reason about, evaluate and suggest solutions. In PISA financial literacy, four process categories have been defined: *identify financial information*, *analyse information in a financial context*, *evaluate financial issues* and *apply financial knowledge and understanding*. While the verbs used here bear some resemblance to those in Bloom's taxonomy of educational objectives (Bloom, 1956), an important distinction is that the processes in the financial literacy construct are not operationalised as a hierarchy of skills. They are, instead, parallel essential cognitive approaches, all of which are part of the financially literate individual's repertoire. The order in which the processes are presented here relates to a typical sequence of thought processes and actions, rather than to an order of difficulty or challenge. At the same time, it is recognised that financial thinking, decisions and actions are most often dependent on a recursive and interactive blend of the processes described in this section. For the purposes of the assessment, each task is identified with the process that is judged most central to its completion.

a. Identify financial information

This process is engaged when the individual searches and accesses sources of financial information, and identifies or recognises its relevance. In PISA 2018 the information is in the form of texts such as contracts, advertisements, charts, tables, forms and instructions displayed on screen. A typical task might ask students to identify the features of a purchase invoice, or recognise the balance on a bank statement. A more difficult task might involve searching through a contract that uses complex legal language to locate information that explains the consequences of defaulting on loan repayments. This process category is also reflected in tasks that involve recognising financial terminology, such as identifying "inflation" as the term used to describe increasing prices over time.

Example 5, *PAY SLIP*, shows an item that focuses on identifying and interpreting financial information.

- Illustrative PISA Financial literacy item 5 – PAY SLIP

Figure 8. Item for the unit PAY SLIP

Each month, Jane's employer pays money into Jane's bank account.

This is Jane's pay slip for July.

EMPLOYEE PAY SLIP: Jane Citizen

Position: Manager	1 July to 31 July
Gross salary	2 800 zeds
Deductions	300 zeds
Net salary	2 500 zeds
Gross salary to date this year	19 600 zeds

➤ **QUESTION**

- How much money did Jane’s employer pay into Jane’s bank account on 31 July?
 - A. 300 zeds
 - B. 2 500 zeds
 - C. 2 800 zeds
 - D. 19 600 zeds

Students are asked to *identify financial information* in a simple pay slip. The correct answer, 2 500 zeds, was selected by just over half of all students in the field trial.

b. Analyse information in a financial context

This process covers a wide range of cognitive activities undertaken in financial contexts, including interpreting, comparing and contrasting, synthesising, and extrapolating from information that is provided. Essentially it involves recognising something that is not explicit: identifying the underlying assumptions or implications of an issue in a financial context. For example, a task may involve comparing the terms offered by different mobile phone contracts, or working out whether an advertisement for a loan is likely to include unstated conditions. An example in this process category is provided below, in the unit *SHARES*.

- *Illustrative PISA Financial literacy item 6 – SHARES*

Figure 9. Item for the unit *SHARES*

This graph shows the price of one Rich Rock share over a 12-month period.

➤ **QUESTION**

- Which statements about the graph are true?

Circle “True” or “False” for each statement.

Statement	Is the statement true or false?
The best month to buy the shares was September.	True / False
The share price increased by about 50% over the year.	True / False

There are two parts to this question, which asks students to *analyse information in a financial context* by considering the information in a line graph about an investment product. The graph shows how the price of shares has changed over a year. The first part of the question assesses a student's understanding that shares should be bought when the price is low (in this case, September). The second part of the question assesses whether students can correctly identify the increase in share prices and calculate the percentage change over time. Just over half of all students correctly answered both parts of the question in the field trial, by circling "True" for the first statement and "False" for the second.

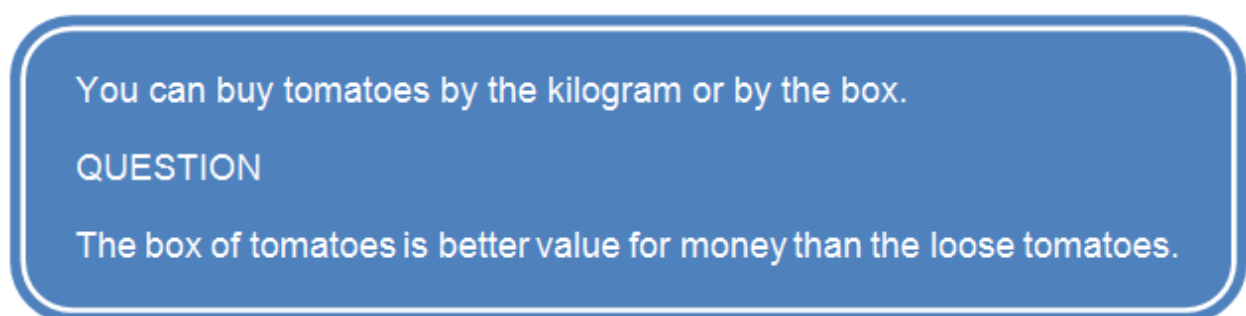
c. Evaluate financial issues

In this process the focus is on recognising or constructing financial justifications and explanations, drawing on financial knowledge and understanding applied in specified contexts. It involves such cognitive activities as explaining, assessing and generalising. Critical thinking is brought into play in this process, when students must draw on knowledge, logic and plausible reasoning to make sense of and form a view about a finance-related problem. The information that is required to deal with such a problem may be partly provided in the stimulus of the task, but students will need to connect such information with their own prior financial knowledge and understandings. In the PISA context, any information that is required to understand the problem is intended to be within the expected range of experiences of a 15-year-old – either direct experiences or those that can be readily imagined and understood. For example, it is assumed that 15-year-olds are likely to be able identify with the experience of wanting something that is not essential (such as a music player or games console). A task based on this scenario could ask about the factors that might be considered in deciding on the relative financial merits of making a purchase or deferring it, given specified financial circumstances.

The next example task *AT THE MARKET*, based on the same stimulus as the illustrative PISA financial literacy item 1, asks students to evaluate information in a financial context by drawing on every day prior knowledge.

- Illustrative PISA Financial literacy item 7 – AT THE MARKET

Figure 10. Item for the unit *AT THE MARKET*



You can buy tomatoes by the kilogram or by the box.

QUESTION

The box of tomatoes is better value for money than the loose tomatoes.

➤ QUESTION

- Buying a box of tomatoes may be a bad financial decision for some people. Explain why.....

The purpose of this question is to recognise that buying in bulk may be wasteful for some people (such as those living in small households) or in some situations (such as during a short family holiday), or may be unaffordable in the short term. Students evaluate a financial issue in the situation presented and score full credit if they can explain that buying more tomatoes at a cheaper price may not always be a good financial decision. Answers such as the following, which refer to waste, received full credit:

The tomatoes might rot before you use them all.

Because you may not need 10 kg of tomatoes.

Another kind of answer that received full credit focused on individuals being unable to afford bulk buying:

- You have to spend 22 zeds (rather than 2.75 or 5.50 for 1 or 2 kg) and you might not have that amount to spend.
- You might have to go without something else that you need to pay for the box of tomatoes.

In the field trial, over 80% of students gained credit for this item, by referring to either waste or affordability.

NEW OFFER provides an example of a more demanding task that falls within the *evaluate financial issues* category.

- *Illustrative PISA Financial literacy item 8 – NEW OFFER*

Figure 11. Item for the unit *NEW OFFER*

Mrs Jones has a loan of 8 000 zeds with FirstZed Finance. The annual interest rate on the loan is 15%. Her repayments each month are 150 zeds.

After one year Mrs Jones still owes 7 400 zeds.

Another finance company called Zedbest will give Mrs Jones a loan of 10 000 zeds with an annual interest rate of 13%. Her repayments each month would also be 150 zeds.

➤ **QUESTION**

- What is one possible negative financial consequence for Mrs Jones if she agrees to the Zedbest loan?

NEW OFFER asks students to reflect on and evaluate the consequences of changing from one set of loan conditions to another – a context that is less likely to be familiar to 15-year-olds than the context provided in *AT THE MARKET*. In the case of *NEW OFFER*, all of the necessary information is provided in the question, but to gain credit students need to identify what is relevant and reflect on the consequences of taking a particular action. There are a number of kinds of responses that are awarded full credit. In the field trial, the most common credited response was that Mrs Jones would have more debt. Equally acceptable are responses that refer to specific conditions of the loan such as that the total interest paid (over the course of the loan) will be greater, that the length of the loan will be greater, and that there are possible fees associated with switching loan companies.

Just over 40% of students in the field trial gained credit for this item.

A third example of a task that fits within the *evaluate* process category is provided in the illustrative PISA financial literacy item 2, *SPENDING CHOICES* where students should draw on plausible reasoning in a financial context to assess which of the tasks listed require prompt attention.

d. Apply financial knowledge and understanding

The fourth process picks up a term from the definition of financial literacy: “to apply such [financial] knowledge and understanding”. It focuses on taking effective action in a financial setting by using knowledge of financial products and contexts, and understanding of financial concepts. This process is reflected in tasks that involve performing calculations and solving problems, often taking into account multiple conditions. An example of this kind of task is calculating the interest on a loan over two years. This process is also reflected in tasks that require recognition of the relevance of prior knowledge in a specific context. For example, a task might require the student to work out whether purchasing power will decline or increase over time when prices are changing at a given rate. In this case, knowledge about inflation needs to be applied.

The following example, *TRAVEL MONEY*, falls into the process category *apply financial knowledge and understanding*.

- Illustrative PISA Financial literacy item 9 – TRAVEL MONEY

Figure 12. Item for the unit *TRAVEL MONEY*

Natasha works in a restaurant 3 evenings each week.
She works for 4 hours each evening and she earns 10 zeds per hour.
Natasha also earns 80 zeds each week in tips.
Natasha saves exactly half of the total amount of money she earns each week.

QUESTION

- Natasha wants to save 600 zeds for a holiday.
How many weeks will it take Natasha to save 600 zeds?

This task requires students to consider a set of conditions and constraints, while planning ahead to pay for future expenses – working out how long it will take to save for a holiday, given a fixed amount of savings each week. The correct answer is “6 weeks”. Fewer than half of the field trial sample gained credit for this item.

C. Contexts

In building a framework, and developing and selecting assessment items based on this framework, attention is given to the breadth of contexts in which the domain literacy is exercised. Decisions about financial issues are often dependent on the contexts or situations in which they are presented. By situating tasks in a variety of contexts the assessment offers the possibility of connecting with the broadest possible range of individual interests across a variety of situations in which individuals need to function in the 21st century.

Certain situations will be more familiar to 15-year-olds than others. In PISA, assessment tasks are framed in situations of general life, which may include but are not confined to school contexts. The focus may be on the individual, family or peer group, on the wider community, or even more widely on a global scale.

As a starting point, the FEG looked at the contexts used in the Programme for the International Assessment of Adult Competencies (PIAAC) literacy framework: *education and work*, *home and family*, *leisure and recreation*, and *community and citizenship* (OECD, 2009b). For the purposes of the financial literacy domain, the heading *leisure and recreation* was replaced by *individual* to reflect the fact that many of the financial interactions that young people have are related to themselves as individual consumers. Such interactions may include leisure and recreation, but are not limited to these. It was further decided to replace *community and citizenship* with *societal*. While *community and citizenship* captures the idea of a perspective wider than the personal, it was felt that the term *community* was not wide enough.

Societal, by contrast, implicitly encompasses national and global situations as well as the more local, thus better fitting the potential reach of financial literacy. The contexts identified for the PISA financial literacy assessment are, then, *education and work*, *home and family*, *individual* and *societal*.

e. Education and work

The context of *education and work* is of great importance to young people. Virtually all 15-year-olds will be starting to think about financial matters related to both education and work, whether they are spending existing earnings, considering future education options or planning their working life.

The educational context is obviously relevant to PISA students, since they are by definition a sample of the school-based population; indeed, many of them will continue in education or training for some time. However, many 15-year-old students are also already engaged in some form of paid work outside school hours making the work context equally valid. Furthermore, many will move from education into some form of employment (including self-employment) before reaching their twenties.

Typical tasks within this context could include understanding payslips, planning to save for tertiary education, investigating the benefits and risks of taking out a student loan, and participating in workplace savings schemes.

Item 5, in *PAY SLIP*, and item 9, in *TRAVEL MONEY*, illustrate the kind of task designed to reflect the *education and work* context category. Specifically, they are examples of the *work* context, asking students to address financial problems related to earned income; the first asking them to identify information on a payslip, and the second using the context of income to make a savings plan.

f. Home and family

Home and family includes financial issues relating to the costs involved in running a household. Family is the most likely household circumstance for 15-year-olds; however, this category also encompasses households that are not based on family relationships, such as the kind of shared accommodation that young people often use shortly after leaving the family home. Tasks within this context may include buying household items or family groceries, keeping records of family spending and making plans for family events. Decisions about budgeting and prioritising spending may also be framed within this context.

The two items from the *AT THE MARKET* unit, items 1 and 7 are categorised as home and family, since grocery shopping is usually done for a household. Item 2, in *SPENDING CHOICES*, is also in this context category: the setting is shared accommodation, and the choices to be made will affect the household (in this case, friends rather than a family sharing accommodation).

g. Individual

The context of the *individual* is important within personal finance since there are many decisions that a person takes entirely for personal benefit or gratification, and many risks and responsibilities that must be borne by individuals. These decisions span essential personal needs, as well as leisure and recreation. They include choosing personal products and services such as clothing, toiletries or haircuts, or buying consumer goods such as electronic or sports equipment, as well as commitments such as season tickets or a gym membership. They also cover the process of making personal decisions and the importance of ensuring individual financial security such as keeping personal information safe and being cautious about unfamiliar products.

Although the decisions made by an individual may be influenced by the family and society (and may impact society), when it comes to opening a bank account, buying shares or getting a loan it is typically the individual who has the legal responsibility and ownership. The context *individual* therefore includes contractual issues around events such as opening a bank account, purchasing consumer goods, paying for recreational activities, and dealing with relevant financial services that are often associated with larger consumption items, such as credit and insurance.

NEW BANK CARD is an example of an item from the *individual* context category.

- Illustrative PISA Financial literacy item 10 – *NEW BANK CARD*

Figure 13. Item for the unit *NEW BANK CARD*

Lisa lives in Zedland. She receives this new bank

➤ **QUESTION**

- The following day, Lisa receives the Personal Identification Number (PIN) for the bank card.
What should Lisa do with the PIN?
 - A. Write the PIN on notepaper and keep this in her wallet.
 - B. Tell the PIN to her friends.
 - C. Write the PIN on the back of the card.
 - D. Memorise the PIN.

This task assesses students' understanding of the individual's responsibility in maintaining security when accessing and using electronic banking. This question asks students to evaluate which of the four presented options is best practice when using a bank card. Over 90% of students in the field trial chose the correct option of memorising the PIN (D).

Other items from the *individual* context category, shown earlier, include item 3, in *MOTORBIKE INSURANCE*, item 6, in *SHARES*, and item 8, in *NEW OFFER*, all of which illustrate decisions that impact on the individual (renewing insurance, buying shares, refinancing a loan).

a. Societal

The environment young people are living in is characterised by change, complexity and interdependence. Globalisation is creating new forms of interdependence where actions are subject to economic influences and consequences that stretch well beyond the individual and the local community. While the core of the financial literacy domain is focused on personal finances, the *societal* context recognises that individual financial well-being cannot be entirely separated from the rest of society. Personal financial well-being affects and is affected by the local community, the nation and even global activities. Financial literacy within this context includes matters such as being informed about consumer rights and responsibilities, understanding the purpose of taxes and local government charges, being aware of business interests, and taking into account the role of consumer purchasing power. It extends also to considering financial choices such as donating to non-profit organisations and charities.

The task *BANK ERROR* (item 4, shown earlier) is categorised as falling within the *societal* context, since it relates to fraudulent behaviour targeted across society.

b. Non-cognitive factors

The PISA working definition of financial literacy includes the non-cognitive terms *motivation* and *confidence*, attitudes which, according to some, have an influence on money management behaviour (Johnson and Staten, 2010). PISA conceives of both financial attitudes and behaviour as aspects of financial literacy in their own right. Attitudes and behaviour are also of interest in terms of their interactions with the cognitive elements of financial literacy. Information collected about the financial attitudes and behaviour of 15-year-olds will also potentially constitute useful baseline data for any longitudinal investigation of the financial literacy of adults, including their financial behaviours.

The FEG identified four non-cognitive factors for inclusion in the framework: *access to information and education*, *access to money and financial products*, *attitudes towards and confidence about financial matters*, and *spending and saving behaviour*.

c. Access to information and education

There are various sources of financial information and education that may be available to students, including informal discussion with friends, parents or other family members, information from the financial sector, as well as formal school education. The literature in this area often refers to the process of 'financial socialisation', which can be seen as the process of acquiring financial literacy. Parents have a major role in the financial socialisation of children but, as discussed above, they may not have experienced all financial contexts and decisions that their children face (Gudmondson and Danes, 2011; Otto, 2013). Copying and discussing financial behaviours with friends is another important source of socialisation, this also may vary in terms of quality and reliability (Centiq, 2008). Moreover, the amount and quality of formal education and training about money and personal finance received by students varies within and across countries (OECD, 2014b, 2014c).

It is possible to ask about students' access to financial information and education through both the student questionnaire and the questionnaire for school principals. In the student questionnaire, students can be asked about the typical sources of information that they access in order to analyse the extent to which each source is correlated with financial literacy. This is intended to provide a description of students' main sources of financial socialisation, rather than assessing whether they understand the importance of using appropriate sources of information or advice, which is covered in the cognitive assessment. Students can also be asked about the types of tasks that they face and the financial concepts they are exposed to during curricular classes. In addition, the school questionnaire will ask principals about the availability and quality of financial education in their schools. Evidence about the extent to which there is a link between levels of financial literacy and financial education inside and outside schools is likely to be particularly useful in shaping education programmes for improving financial literacy.

d. Access to money and financial products

The results of the 2012 PISA financial literacy exercise showed that in the Flemish Community of Belgium, Estonia, New Zealand, and Slovenia students with a bank account scored higher in financial literacy than students with similar socio-economic status who did not hold a bank account (OECD, 2014c). Whilst this does not indicate a causal relationship, it is plausible to assume that real-life experiences of financial products may influence young people's financial literacy and vice versa. Personal experience may come, for example, from using financial products such as payment cards, from dealing with the banking system, or from occasional

working activities outside of school hours. In order to further understand the potential role of learning through experience, in addition to information on access to financial products, the 2018 non-cognitive student questionnaire will collect evidence on a range of practical financial experiences, such as visiting a bank branch, paying in cash or by card, reading a bank statement or making online purchases.

Students who have had more personal experience of dealing with financial matters from earning money or receiving an allowance might also be expected to perform better on the cognitive assessment than those without such experience; however, a recent review suggests that the key factor may not be experience, but the extent to which parents are involved in the spending decisions made by young people, with higher financial literacy associated with more involved parents (Drever et al., 2015). The 2018 framework therefore recognises the importance of knowing whether students have access to money, through which channels, and how spending and savings decisions are made (see 'spending and saving behaviour' below).

e. Attitudes towards and confidence about financial matters

The PISA definition of financial literacy highlights the important role of attitudes. Individual preferences can determine financial behaviour and impact on the ways in which financial knowledge is used. PISA 2012 showed that students' perseverance and openness to problem solving were strongly associated to their financial literacy scores (OECD, 2014c). In addition to this, the extent to which students believe that they are in control of their future, and their preference for current consumption may influence their financial decisions, their independence, and their propensity to learn how to make plans for their own financial security (Golsteyn et al., 2013; Lee and Mortimer, 2009; Meier and Sprenger, 2013). Moreover, confidence in one's own ability to make a financial decision may also be a key driver in explaining who will work through complex financial problems or make choices across several possible products. At the same time, however, confidence may turn into over-confidence, leading to a tendency to mistakes and overly risky decisions. The 2018 framework therefore recognises the importance of investigating students' perception of their own financial knowledge and skills and analysing whether this is associated with higher or lower levels of financial literacy.

f. Spending and saving behaviour

While items on the cognitive assessment test students' ability to make particular spending and savings decisions, it is also useful to have some measure of what their actual (reported) behaviour is: that is, how students save and spend in practice. PISA financial literacy provides the opportunity to look at the potential relationship between 15-year-olds' spending and saving behaviour and their results on the cognitive financial literacy assessment. It provides the opportunity to explore how students typically use their money and whether decisions are made alone, or with the guidance or recommendation of a trusted adult.

3. ASSESSING FINANCIAL LITERACY

The previous section has outlined the conceptual framework for financial literacy. The concepts in the framework must in turn be represented in tasks and questions in order to collect evidence of students' proficiency in financial literacy. In this section we discuss the structure of the assessment, the distribution of tasks across the framework variables, and the choice of response formats. This is followed by a short discussion of the impact of knowledge and skills from other domains on financial literacy and the implications for the assessment. To conclude, we describe the method by which data about financial behaviours and experience will be collected.

The conceptual framework is concerned with mapping the domain. It lays out the definition and the major variables that are addressed in the assessment instrument. The key ideas have been elaborated through lists of sub-topics and examples in the preceding section. These details should not be interpreted as a checklist of tasks included in the 2018 assessment. Given that only one hour of financial literacy assessment material is being administered in PISA, there is not enough space to cover every detail of each variable.

A. The structure of the assessment

In 2012, the PISA financial literacy assessment was developed as a one hour paper and pen exercise, to be completed alongside one hour of material from other cognitive domains. The financial literacy assessment was made of 40 items divided into two clusters, chosen from 75 initial tasks administered in the field trial. The choice between items was made based on their psychometric properties, such as ensuring that each item discriminated between high and low scoring students.

In 2015, items were transferred to a computer based delivery platform, and additional items were developed for this form of delivery in order to replace items that had been released in the report of the 2012 results. The 2015 financial literacy was developed as a one hour exercise, comprising 43 items divided into two clusters.

For 2018, further new items will be developed. For the first time, these will incorporate specially developed interactive elements, in order to provide additional reality and interest for the students and to seek to test the students' decision making processes. These will be used alongside items developed for the 2012 and 2015 assessments in order to ensure that the overall set of items continues to provide the necessary psychometric links across waves of data collection, and to provide the necessary breadth of coverage across the framework.

Interactive items will not be limited to a specific part of the framework, but will be designed to ensure the overall balance of items according to the distribution of score points discussed below. They will include items that are expected to be at different levels of difficulty.

The items will incorporate interactive elements in various ways, but always as a way of encouraging the student to interact with the screen in order to answer a question. Some items may require the student to actively seek more information by clicking links or asking questions, rather than relying on the information presented on the first screen. Others may include graphs or charts that can be manipulated to see a variety of potential outcomes. Such items may allow the student to test different scenarios and explain why certain outcomes occur. Simple drag and drop items could allow students to sort images to prioritise purchases, or to apportion a budget according to a family's needs. A realistic fraudulent website could be created on the interactive platform to test the students ability to spot the risks of such a site and illustrate a strategy for

checking its authenticity, for example, or students could be shown a recent news article and then indicate on a graph what is likely to have happened to the price of a particular share or government bond.

As with other PISA assessment domains, paper-based and computer-based financial literacy items are grouped in units comprising one or two items based around a common stimulus. The selection includes financially-focused stimulus material in diverse formats, including prose, diagrams, tables, charts and illustrations. All financial literacy assessments comprise a broad sample of items covering a range of difficulty that will enable the strengths and weaknesses of students and key subgroups to be measured and described.

B. Response formats and coding

Some PISA items require short descriptive responses; others require more direct responses of one or two sentences or a calculation, whilst some can be answered by checking a box. Decisions about the form in which the data are collected – the response formats of the items – are based on what is considered appropriate given the kind of evidence that is being collected, and also on technical and pragmatic considerations. In the financial literacy assessment as in other PISA assessments, two broad types of items are used: constructed-response items and selected-response items.

Constructed-response items require students to generate their own answers. The format of the answer may be a single word or figure, or may be longer: a few sentences or a worked calculation. Constructed-response items that require a more extended answer are ideal for collecting information about students' capacity to explain decisions or demonstrate a process of analysis. Item 9, in *TRAVEL MONEY*, illustrates a constructed response item that calls for a single figure, where there is a very restricted range of credit-worthy responses. Items 1 and 7, both from the unit *AT THE MARKET*, and item 8, in *NEW OFFER*, are typical of tasks that require more extended responses, where many different kinds of answers may gain full credit.

In the scoring of the *NEW OFFER* task, for example, four different kinds of full credit responses are identified:

- i. Answers that refer to Mrs Jones having more debt if she takes on the new loan offer:
- ii. She will owe more money.
- iii. She will be unable to control her spending.
- iv. She is going deeper into debt.
- v. Answers that refer to Mrs Jones having to pay more interest:
- vi. 13% of 10 000 is greater than 15% of 8 000.
- vii. Answers that refer to the increased length of time over which Mrs Jones will have the debt:
- viii. It might take longer to repay because the loan is bigger and the payments are the same.
- ix. Answers that refer to the possibility that Mrs Jones will have to pay a cancellation fee if she cancels her initial loan agreement with FirstZed Finance:

She may have a penalty fee for paying the FirstZed loan early.

The second broad type of item in terms of format and coding, is selected response. This kind of item requires students to choose one or more alternatives from a given set of options. The most common type in this category is the simple multiple-choice item, which requires the selection of one from a set of (usually) four options: see item 5, in *PAY SLIP*, and item 10, in *NEW BANK CARD*. A second type of selected-response item is complex multiple choice, in which students respond to a series of “Yes/No”-type questions. Item 2, in *SPENDING CHOICES*, illustrates a “Yes/No” set of selections. For this task students need to make three independent correct selections to gain credit. Item 3, in *MOTORBIKE INSURANCE*, has a similar format, in that three independent correct selections must be made to gain credit; but in this case each selection is from three options: “Increases cost”, “Reduces cost” and “Has no effect on cost”. Selected-response items are typically regarded as most suitable for assessing items associated with identifying and recognising information, but they are also a useful way of measuring students’ understanding of higher-order concepts that they themselves may not easily be able to express.

Although particular item formats lend themselves to specific types of questions, care needs to be taken that the format of the item does not affect the interpretation of the results. Research suggests that different groups (for example, boys and girls, and students in different countries) respond differentially to the various item formats. Several research studies on response format effects based on PISA data suggest that there are strong arguments for retaining a mixture of multiple-choice and constructed-response items. In their study of PISA reading literacy compared with the IEA Reading Literacy Study (IEARLS), Lafontaine and Monseur (2006) found that response format had a significant impact on gender performance. In another study, countries were found to show differential equivalence of item difficulties in PISA reading on items in different formats (Grisay and Monseur, 2007). This finding may relate to the fact that students in different countries are more or less familiar with the particular formats. In summary, the PISA financial literacy option includes items in a variety of formats to minimise the possibility that the item format influences student performance. Such an influence would be extrinsic to the intended object of measurement: in this case, financial literacy.

When considering the distribution of item formats, the question of resources must be weighed as well as the equity issues discussed in the preceding paragraphs. All except the most simple of constructed-response items are coded by expert judges who must be trained and monitored. Selected response and very short “closed” constructed response items do not require expert coding and therefore demand fewer resources.

The proportions of constructed- and selected-response items are determined taking account of all these considerations. The majority of the items selected for the PISA 2012 main survey do not require expert judgement.

Most items are coded dichotomously (full credit or no credit), but where appropriate an item’s coding scheme allows for partial credit. Partial credit makes possible more nuanced scoring of items. Some answers, even though incomplete, are better than others. If incomplete answers for a particular question indicate a higher level of financial literacy than inaccurate or incorrect answers, a scoring scheme has been devised that allows partial credit for that question. Such “partial credit” items yield more than one score point.

C. Distribution of score points

In this section we outline the distribution of score points across the categories of the three main framework characteristics discussed previously. The term “score points” is used in preference to “items”, as some partial credit items are included. The distributions are expressed in terms of ranges, indicating the approximate weighting of the various categories. The distribution will contain a mix of original items, developed for the 2012 assessment and those items developed for the 2015 and 2018 assessments. In particular, care will be taken to ensure that interactive items are included across different elements of the assessment.

While each PISA financial literacy item is categorised according to a single content, a single process and a single context category it is recognised that, since PISA aims to reflect real-life situations and problems, often elements of more than one category are present in a task. In such cases, the item is identified with the category judged most integral to responding successfully to the task.

The target distribution of score points according to financial literacy content areas is shown in Table 5. It is anticipated that interactive items will be developed for each of the content areas and that they are likely to be distributed across the four context areas. There are likely to primarily focus on three of the four processes, with less emphasis on ‘Identify financial information’.

Table 5. Approximate distribution of score points in financial literacy, by content

Money and transactions	Planning and managing finances	Risk and reward	Financial landscape	Total
30% - 40%	25% - 35%	15% - 25%	10% - 20%	100%

The distribution reflects that *money and transactions* is considered to be to the most immediately relevant content area for 15-year-olds.

Table 6 shows the target distribution of score points by the four processes.

Table 6. Approximate distribution of score points in financial literacy, by processes

Identify financial information	Analyse information in a financial context	Evaluate financial issues	Apply financial knowledge and understanding	Total
15% - 25%	15% - 25%	25% - 35%	25% - 35%	100%

The weighting shows that greater importance was attributed to evaluating financial issues and applying financial knowledge and understanding.

Table 7 shows the target distribution of score points by the four contexts.

Table 7. Approximate distribution of score points in financial literacy, by contexts

Education and work	Home and family	Individual	Societal	Total
10% - 20%	30% - 40%	35% - 45%	5% -15%	100%

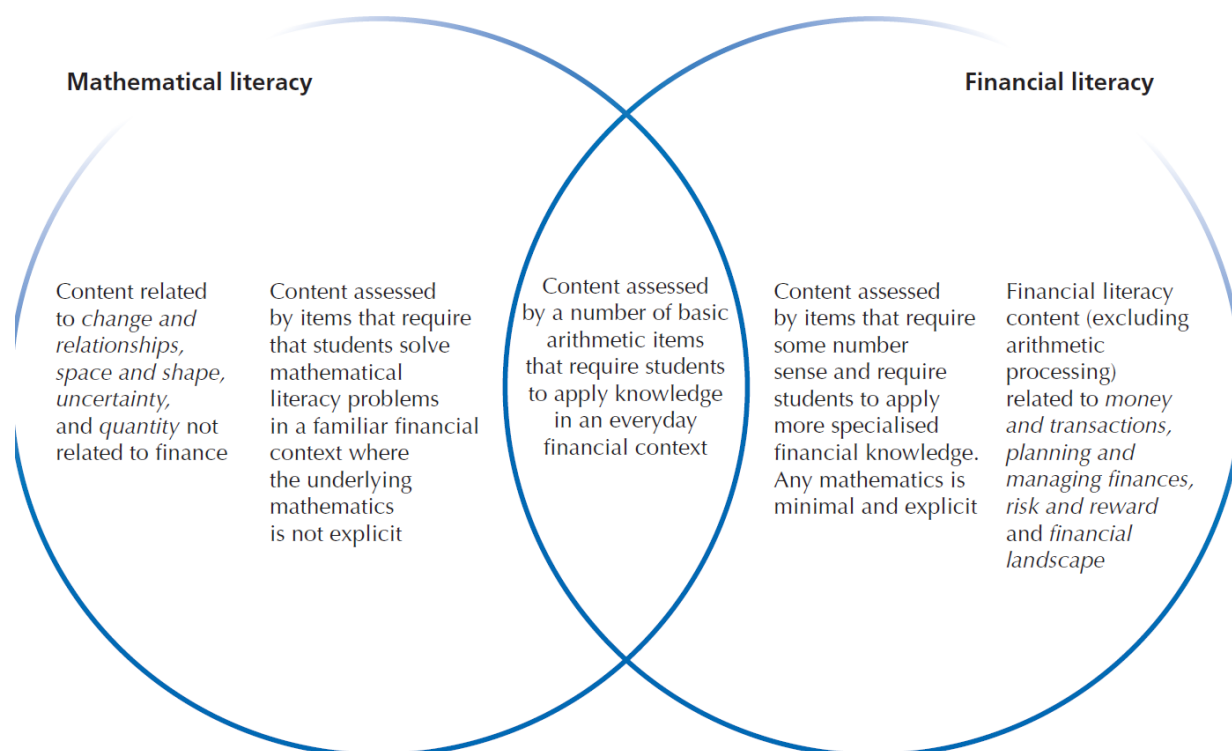
Consistent with an assessment of personal financial literacy of 15-year-olds, there is a clear emphasis on *individual*, but also a weighting towards the financial interests of the household or family unit. *Education and work* and *societal* contexts are given less emphasis, but included in the scheme as they are important elements of financial experience.

D. The impact of other domain knowledge and skills on financial literacy

A certain level of numeracy (or mathematical literacy) is regarded as a necessary condition of financial literacy. Huston (2010) argues that “if an individual struggles with arithmetic skills, this will certainly impact his/her financial literacy. However, available tools (e.g. calculators) can compensate for these deficiencies; thus, information directly related to successfully navigating personal finances is a more appropriate focus than numeracy skills for a financial literacy measure”. Mathematically-related proficiencies such as number sense, familiarity with multiple representations of numbers, and skills in mental calculation, estimation, and the assessment of reasonableness of results are intrinsic to some aspects of financial literacy.

On the other hand there are large areas where the content of mathematical literacy and financial literacy do not intersect. As defined in the PISA 2012 mathematics literacy framework, mathematical literacy incorporates four content areas: *change and relationships*, *space and shape*, *quantity* and *uncertainty*. Of these, only *quantity* directly intersects with the content of the PISA financial literacy assessment. Unlike the mathematical literacy content area *uncertainty*, which requires students to apply probability measures and statistics, in the PISA assessment the financial literacy content area *risk and reward* requires an understanding of the features of a particular situation or product that indicate a that there will be a risk of losing money and (sometimes) a possibility of gains. This is a non-numeric appreciation of the way financial well-being can be affected by chance and an awareness of the related products and actions to protect against loss. In the financial literacy assessment, the quantity-related proficiencies listed previously are applied to problems requiring more financial knowledge than can be expected in the mathematical literacy assessment. Similarly, knowledge about financial matters and capability in applying such knowledge and reasoning in financial contexts (in the absence of any specifically mathematical content) characterise much of all four content areas of financial literacy: *money and transactions*, *planning and managing finances*, *risk and reward* and *financial landscape*. Figure 14 represents the relationship between the content of mathematical literacy and financial literacy in PISA.

Figure 14. Relationship between the content of financial literacy and mathematical literacy in PISA



Operationally, there are few items populating the portion of the diagram where the two circles intersect. In the financial literacy assessment, the nature of the mathematical literacy expected is basic arithmetic: the four operations (addition, subtraction, multiplication and division) with whole numbers, decimals and common percentages. Such arithmetic occurs as an intrinsic part of the financial literacy context and enables financial literacy knowledge to be applied and demonstrated. Item 1, from the unit *AT THE MARKET*, illustrates an item that requires such arithmetic skills: the mathematics involved (dividing by a factor of ten) is very basic and is of a level well within the reach of most 15-year-olds. Use of financial formulae (requiring capability with algebra) is not considered appropriate. Dependence on calculation is minimised in the assessment; tasks are framed in such a way as to avoid the need for substantial or repetitive calculation. The calculators used by students in their classrooms and on the PISA mathematics assessment will also be available in the financial literacy assessment, but success in the items will not depend on calculator use.

A similar reasoning holds for reading skills. It is assumed that all students taking part in the financial literacy assessment will have some basic reading proficiency, even while it is known from previous PISA surveys that reading skill varies widely both within and across countries (OECD, 2010b). To minimise the level of reading literacy required, stimulus material and task statements are generally designed to be as clear, simple and brief as possible. In some cases, however, stimulus may deliberately present complex or somewhat technical language: the capacity to read and interpret the language of financial documents or pseudo financial documents is regarded as part of financial literacy. In item 4, in *BANK ERROR*, attentive reading of the e-mail message is required to detect the likelihood that the message is part of a scam.

Highly technical terminology relating to financial matters is avoided. The FEG has advised on

terms that it judges reasonable to expect 15-year-olds to understand. Some of these terms may be the focus of assessment tasks. For example, *PAY SLIP* (item 5), assesses whether students are able to read a simple pay slip and recognise (or infer the meaning of) the terms “gross” and “net”.

In practice, the results of the 2012 PISA financial literacy assessment gave a more precise measure of students’ performance in financial literacy in comparison with reading and mathematics performance. The results indicated that around 25% of the financial literacy score reflected factors that are uniquely captured by the financial literacy assessment, while the remaining 75% of the financial literacy score reflected skills measured in the mathematics and/or reading assessments. The association between financial literacy and other domains indicates that, in general, students who perform at higher levels in mathematics and/or reading also perform well in financial literacy. There were, however, wide variations in financial literacy performance for any given level of performance in mathematics and reading, meaning that the skills measured by the financial literacy assessment went beyond or fell short of the ability to use the knowledge that students acquired from subjects taught in compulsory education. For instance, in Australia, the Flemish Community of Belgium, the Czech Republic, Estonia, New Zealand and the Russian Federation, students performed better in financial literacy than students in other countries with similar performance in mathematics and reading, while on the contrary in France, Italy and Slovenia, students’ performance in financial literacy was lower, on average, when compared to that of students in the other participating countries and economies who displayed the same level of proficiency in reading and mathematics (OECD, 2014c).

4. REPORTING FINANCIAL LITERACY

The data from the 2012 financial literacy assessment is held in a database separate from the main PISA database. In 2015, the data from all domains will be presented together. The databases include, for the sampled students, their financial literacy, mathematics and reading cognitive results, the behaviour data from the short questionnaire on financial literacy, and data from the general student questionnaire and school questionnaire.

In each wave, it is possible to report on financial literacy as an independent result, and on financial literacy in relation to performance in other domains, financial behaviour, and in relation to some background variables such as socioeconomic status and immigrant status. The data also allow the development of further work under the aegis of the OECD Project on Financial Education.

The financial literacy cognitive data is scaled in a similar way to the other PISA data. A comprehensive description of the modelling technique used for scaling can be found in the *PISA 2012 Technical Report* (OECD, 2014d).

Each item is associated with a particular point on the PISA financial literacy scale that indicates its difficulty, and each student’s performance is associated with a particular point on the same scale that indicates the student’s estimated proficiency.

As with the other PISA domains, the relative difficulty of tasks in a test is estimated by considering the proportion of test takers getting each question correct. The relative proficiency of students taking a particular test is estimated by considering the proportion of test items that they answer correctly. A single continuous scale showing the relationship between the difficulty of items and the proficiency of students is constructed.

Starting from the 2012 assessment, the scale was divided into levels, according to a set of statistical principles, and then descriptions were generated based on the tasks located within each level, to encapsulate the kinds of skills and knowledge needed to successfully complete those tasks. The scale and set of descriptions are known as a described proficiency scale.

By calibrating the difficulty of each item, it is possible to locate the degree of financial literacy that the item represents. By showing the proficiency of each student on the same scale, it is possible to describe the degree of financial literacy that the student possesses. The described proficiency scale helps in interpreting what students' financial literacy scores mean in substantive terms.

Following PISA practice, a scale is constructed (based on participating OECD countries) having a mean of 500 and a standard deviation of 100. Five levels of proficiency in financial literacy were described in the 2012 assessment, as a first step in reporting how competency in financial literacy develops, and to enable comparisons of student performance between and within participating countries and economies (see OECD, 2014c, Chapter 2).

NOTES

1. The Joint Ministerial Statement from the 2012 APEC Finance Ministerial Meeting is available at http://www.apec.org/Meeting-Papers/Ministerial-Statements/Finance/2012_finance.aspx.

REFERENCES

- Atkinson A., F-A. Messy, L. Rabinovich and J. Yoong (2015, forthcoming), "Research and Literature on Financial Education for Long-term Savings and Investments" OECD Working Papers on Finance, Insurance and Private Pensions, No. 39, OECD Publishing.
- Atkinson, A. and F. Messy (2012), "Measuring Financial Literacy: Results of the OECD / International Network on Financial Education (INFE) Pilot Study", *OECD Working Papers on Finance, Insurance and Private Pensions*, No. 15, OECD Publishing.
- Barrett, A., I. Mosca and B. J. Whelan (2013), "(Lack of) Pension Knowledge". IZA Discussion Paper No. 7596 August 2013.
- Barsky, R.B., F.T. Juster, M.S. Kimball and M.D. Shapiro (1997), "Preference parameters and behavioural heterogeneity: An experimental approach in the health and retirement study", *Quarterly Journal of Economics*, No. 11, pp. 537-539.
- Bernheim, D., D. Garrett and D. Maki (2001), "Education and saving: The long-term effects of high school financial curriculum mandates", *Journal of Public Economics*, No. 85, pp. 435-565.
- Bloom, B.S. (ed.) (1956), *Taxonomy of Educational Objectives: The Classification of Educational Goals*, David McKay, New York.
- Bradley, L. (2012), *Young People and Savings*. Institute for Public Policy Research, London
- Bruhn, M., L. de Souza Leao, A. Legovini, R. Marchetti, and B. Zia (2013). *The impact of high school financial education : experimental evidence from Brazil*. Policy Research working paper ; no. WPS 6723; Impact Evaluation series ; no. IE 109. Washington, DC: World Bank Group. <http://documents.worldbank.org/curated/en/2013/12/18640673/impact-high-school-financial-education-experimental-evidence-brazil>
- CentiQ (2008), Summary: *Financial Understanding and Behaviour of 8- to 18-year-olds in the Netherlands*, CentiQ, www.wijzeringeldzaken.nl/media/13194/summary_financial_understanding_%208to18_netherlands.pdf.
- Cole, S., T. Sampson and B. Zia (2011), "Prices or Knowledge? What Drives Demand for Financial Services in Emerging Markets?", *The Journal of Finance*, Vol. 66, No. 6, pp. 1933-1967.
- Colombo, F., et al. (2011), Help Wanted?: Providing and Paying for Long-Term Care, OECD Health Policy Studies, OECD Publishing, Paris. DOI: <http://dx.doi.org/10.1787/9789264097759-en>
- Demircuc-Kunt, A., L. Klapper, D. Singer, and P. Van Oudheusden (2015). "The Global Findex Database 2014: Measuring Financial Inclusion around the World." Policy Research Working Paper 7255, World Bank, Washington, DC.
- Drever, A.I.; E. Odders-White; C.W. Kalish, N.M. Else-Quest; E.M. Hoagland and E.N. Nelms (2012) "Foundation of financial well-being: insights into the role of executive function, financial socialization and experience-based learning in childhood and youth". *Journal of Consumer Affairs*, Vol 49, No. 1, Spring 2015 13-38
- G20 (2012), *G20 Leaders Declaration*, Los Cabos, www.g20mexico.org/images/stories/docs/g20/conclu/G20_Leaders_Declaration_2012.pdf.
- G20 (2013), *G20 Leaders Declaration, Saint Petersburg*, https://g20.org/wp-content/uploads/2014/12/Saint_Petersburg_Declaration_ENG_0.pdf
- Gerardi, K., L. Goette, and S. Meier (2010), "Financial Literacy and Subprime Mortgage Delinquency: Evidence from a Survey Matched to Administrative Data", Federal Reserve Bank of Atlanta, Working Paper Series No. 2010-10.
- Golsteyn, B. B.H. Grönqvist and L. Lindahl (2013), "Time Preferences and Lifetime Outcomes", IZA Discussion Paper No. 7165
- Grisay, A. and C. Monseur (2007), "Measuring the equivalence of item difficulty in the various versions of an international test", *Studies in Educational Evaluation*, Vol. 33, No. 1, pp. 69-86.
- Gudmondson, C.G. and S.M. Danes (2011), "Family Financial Socialization: Theory and Critical Review", *Journal of Family and Economic Issues*, 32, pp. 644-667.
- Habschick, M., B. Seidl and J. Evers (2007), *Survey of Financial Literacy Schemes in the EU27*, Hamburg.

- Hastings, J. and L. Tejada-Ashton (2008), Financial Literacy, Information, and Demand Elasticity: Survey and Experimental Evidence from Mexico, NBER Working Paper, 14538, Cambridge, Massachusetts.
- Hilgert, M.A., J.M. Hogarth and S.G. Beverly (2003), Household Financial Management: The Connection Between Knowledge and Behavior, *Federal Reserve Bulletin*, Vol. 89, No. 7, pp. 309-322.
- Holt, C. and S. Laury (2002), "Risk aversion and incentive effects", *American Economic Review*, Vol. 92, No. 5, pp.1644-1655.
- Huston, S.J. (2010), "Measuring financial literacy", *The Journal of Consumer Affairs*, Vol. 44, No. 2, pp. 296-316.
- Johnson, C. and M. Staten (2010), *Do inter-temporal preferences trump financial education courses in driving borrowing and payment behaviour?*, Paper presented at the 1st Annual Boulder Conference on Consumer Financial Decision Making, June 27-29, 2010, Boulder, Colorado.
- Kempson, E., S. Collard and N. Moore (2005), *Measuring financial capability: an exploratory study*, Financial Services Authority, London.
- Lafontaine, D. and C. Monseur (2006), *Impact of Test Characteristics on Gender Equity Indicators in the Assessment of Reading Comprehension*, University of Liège.
- Lee, J. C., and J. T. Mortimer, (2009). "Family Socialization, Economic Self-Efficacy, and the Attainment of Financial Independence in Early Adulthood". *Longitudinal and Life Course Studies*, 1(1), 45–62.
- Lusardi, A. (2009), U.S. Household Savings Behavior: The Role of Financial Literacy, Information and Financial Education Programs, in C. Foote, L. Goette and S. Meier (eds.), *Policymaking Insights from Behavioral Economics*, Federal Reserve Bank of Boston, pp. 109-149.
- Lusardi, A. and O.S. Mitchell (2008), "Planning and Financial Literacy: How Do Women Fare? ", *American Economic Review*, Vol. 98, No. 2, pp. 413-417.
- Lusardi, A. and O.S. Mitchell (2011), Financial Literacy and Planning: Implications for Retirement Wellbeing, in A. Lusardi and O. S. Mitchell (eds.), *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, Oxford University Press.
- Lusardi, A. and P. Tufano (2009a), "Debt Literacy, Financial Experiences, and Overindebtedness", NBER Working Paper No. 14808, Cambridge, Massachusetts.
- Lusardi, A. and P. Tufano (2009b), Teach Workers about the Perils of Debt, *Harvard Business Review* (November), pp. 22-24.
- Lusardi, A., O.S. Mitchell and V. Curto (2010), "Financial literacy among the young", *The Journal of Consumer Affairs*, Vol. 44, No. 2, pp. 358-380.
- Meier, S. and C.D. Sprenger (2013), "Discounting financial literacy: Time preferences and participation in financial education programs", *Journal of Economic Behavior & Organization*, 95, November 2013, pp. 159-174 <http://dx.doi.org/10.1016/j.jebo.2012.02.024>.
- Messy F.A. and C. Monticone, (2015 forthcoming), Trends and recent developments on financial education in Europe. OECD Working Papers on Finance, Insurance and Private Pensions
- Messy F.A. and C. Monticone, (2016 forthcoming), Financial education policies in Asia and Pacific. OECD Working Papers on Finance, Insurance and Private Pensions
- Miller, M., J. Reichelstein, C. Salas, and B. Zia (2014). *Can you help someone become financially capable ? a meta-analysis of the literature*. Policy Research working paper ; no. WPS 6745. Washington, DC: World Bank Group. <http://documents.worldbank.org/curated/en/2014/01/18807418/can-help-someone-financially-capable-meta-analysis-literature>
- Ministerial Council for Education Early Childhood Development and Youth Affairs (2011), National Consumer and Financial Literacy Framework, MCEETYA Consumer and Financial Literacy Working Party, Melbourne.
- Money and Pensions Panel (2013), *Basic pension: the default option for labour-market pensions*. Report by the Committee of the Money and Pension Panel. February 2013.
- Moore, D. (2003), *Survey of Financial Literacy in Washington State: Knowledge, Behavior, Attitudes, and Experiences*, Social and Economic Sciences Research Center, Washington State University.
- OECD (2005a), *Recommendation on Principles and Good Practices for Financial Education*, OECD Publishing.
- OECD (2005b), *Improving Financial Literacy: Analysis of Issues and Policies*, OECD Publishing.

- OECD (2008), *Improving Financial Education and Awareness on Insurance and Private Pensions*, OECD Publishing.
- OECD (2009a), *Financial Literacy and Consumer Protection: Overlooked Aspects of the Crisis*, OECD Publishing.
- OECD (2009b), "PIAAC Literacy: A Conceptual Framework", *OECD Education Working Paper No. 34*, OECD Publishing.
- OECD (2009c), *PISA 2006 Technical Report*, OECD Publishing.
- OECD (2010a), *PISA 2009 Framework: Key Competencies in Reading, Mathematics and Science*, OECD Publishing.
- OECD (2010b), *PISA 2009 Results: What Students Know and Can Do*, Vol. I, OECD Publishing.
- OECD (2011), *Education at a Glance 2011: OECD Indicators*, OECD Publishing.
- OECD (2013), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, OECD Publishing
- OECD (2014a), *Education at a Glance 2014: OECD Indicators*, OECD Publishing. <http://dx.doi.org/10.1787/eag-2014-en>
- OECD (2014b), *Financial Education for Youth: The Role of Schools*, OECD Publishing.
- OECD (2014c), *PISA 2012 Results: Students and Money: Financial Literacy Skills for the 21st Century (Volume VI)*, PISA, OECD Publishing. <http://dx.doi.org/10.1787/9789264208094-en>
- OECD (2014d), *PISA 2012 Technical Report*, PISA, OECD Publishing, Paris.
- OECD/INFE (2009), *Financial Education and the Crisis: Policy Paper and Guidance*, OECD Publishing.
- OECD/INFE (2011), *Measuring Financial Literacy: Core Questionnaire in Measuring Financial Literacy: Questionnaire and Guidance Notes for conducting an Internationally Comparable Survey of Financial literacy*, OECD Publishing.
- OECD/INFE (2012), *OECD/INFE High-Level Principles on National Strategies for Financial Education*, OECD Publishing.
- OECD/INFE (2013), *Financial Literacy and Inclusion: Results of the OECD/INFE Survey Across Countries and By Gender*.
http://www.oecd.org/daf/fin/financial-education/TrustFund2013_OECD_INFE_Fin_Lit_and_Incl_SurveyResults_by_Country_and_Gender.pdf
- OECD/INFE (2015), *2015 OECD/INFE Toolkit for Measuring Financial Literacy and Financial Inclusion*, OECD Publishing
- Ratcliffe C. and S. McKernan (2013), *Forever in Your Debt: Who Has Student Loan Debt, and Who's Worried*. The Urban Institute and FINRA Investor Education Foundation
- Otto, A. (2013), "Saving in Childhood and Adolescence: Insights from Developmental Psychology", *Economics of Education Review*, 33, pp. 8-18.
- Schuchardt, J., S.D. Hanna, T.K. Hira, A.C. Lyons, L. Palmer and J.J. Xiao (2009), "Financial Literacy and Education Research Priorities", *Journal of Financial Counseling and Planning*, Vol. 20, No. 1, pp. 84-95.
- Smithers, R. (2010), University students expect to graduate with debts in excess of £15,000, *The Guardian*, 18 March 2010, www.guardian.co.uk/money/2010/mar/18/university-students-graduate-mounting-debts.
- Stango, V. and J. Zinman (2009), "Exponential Growth Bias and Household Finance", *Journal of Finance*, Vol. 64, No. 6, pp. 2807-2849.
- Van Rooij, M.A., A. Lusardi and R. Alessie (2011), "Financial Literacy and Stock Market Participation", *Journal of Financial Economics*, Vol. 101, No. 2, pp.449-472.
- Xu, L. and B. Zia (2012), "Financial Literacy around the World An Overview of the Evidence with Practical Suggestions for the Way Forward". World Bank Policy Research Working Paper 6107.
- Yoong, J. (2011), Financial Illiteracy and Stock Market Participation: Evidence from the RAND American Life Panel, in A. Lusardi and O. S. Mitchell (eds.), *Financial Literacy: Implications for Retirement Security and the Financial Marketplace*, Oxford University Press.

III. PISA 2018 – DRAFT QUESTIONNAIRE FRAMEWORK

INTRODUCTION

A. Aims of the Programme for International Student Assessment

The main features of the Programme for International Student Assessment (PISA) are as follows:

- PISA is a *system-level assessment*, representing a commitment by governments to monitor the outcomes of education systems.
- PISA is *policy-oriented* by linking data on students' learning outcomes with data on key factors that shape learning in and out of school.
- PISA is *carried out regularly* to enable countries to monitor their progress in meeting key learning objectives.
- PISA *assesses both subject matter content knowledge*, on the one hand, and *the capacity of individuals to apply that knowledge creatively*, including in unfamiliar contexts, on the other.

PISA focuses on *knowledge and skills towards the end of compulsory schooling*. In most countries the end of compulsory education is around the age of 15, where students are supposed to have mastered the basic skills and knowledge to continue on to higher education or in the workforce.

- PISA is designed to *provide comparable data across a wide range of countries*. Considerable efforts are devoted to achieving cultural and linguistic breadth and balance in assessment materials.
- PISA is a *collaborative effort* involving multiple parties.

PISA continues to yield indicators on effectiveness, equity, and efficiency of educational systems, setting benchmarks for international comparison and monitoring trends over time. PISA also builds a sustainable database that allows researchers world-wide to study basic as well as policy-oriented questions on education, including those related to society and economy. The OECD and the PISA Governing Board (PGB) continue to look for ways to increase the scientific quality and policy relevance of the PISA questionnaire to meet these needs.

The PISA study delivers a set of measures derived from assessment of the cognitive domains and contextual background information. It has been repeatedly shown that cognitive skills learned in school are related to students' backgrounds and learning contexts (OECD, 2013a, p. 168). The PISA study follows the literature (e.g. Heckman, 2006) in acknowledging the increasingly important role of non-cognitive aspects for achieving success in school and at work. Non-cognitive aspects refer to attitudes, orientations and strategies that are taken to underpin success in school and at work, such as motivation, positive self-beliefs, perseverance and self-control (Gutman & Schoon, 2013). Creating stronger links between the cognitive and non-cognitive instruments increases the analytical power and insights that can be gained from the study; the questionnaire framework developed here is aimed at continuing to enhance this aspect of the PISA study.

B. PISA 2018

PISA 2018 will be the seventh cycle of the programme, and in that time the information gathered by the questionnaires has become both highly visible and important with the scope of issues to be covered gradually expanding since PISA 2000. In 2018 there is a need to have a balanced approach that involves continuing trend constructs (i.e. constructs that were also assessed in at least one previous cycle of PISA) and measuring new constructs. Educational policies and practices are monitored through various indicators, together with student achievement, individual background, learning opportunities and non-cognitive outcomes. More recently, data analysis and reporting have become more complex, allowing for more in-depth reporting. In addition to providing tables of indicators, the patterns of input and process and outcome variables are identified within and across countries; trends are reported, relations are explored and impact is estimated. These analyses require more sophisticated statistical modelling approaches and detailed data on contextual factors regarding students, teachers, schools and educational systems.

After 18 years of implementation, the key characteristic of PISA is the availability of trend data about education systems. PISA not only allows for a description of change in a country's performance level over time, but also for a description of changes in non-cognitive outcomes, living conditions of adolescents and their families and the professional practices and organisational structures of schooling. As PISA involves repeated cycles of measurement, much can be learned from examining the stability and variability of conditions, processes, outcomes and their relations: (a) policy makers use trend data for constant diagnosis and feedback as shown in the recent *Ed Policy Outlook 2015* (OECD, 2015); (b) the explanatory power of the study will increase because changes in performance can be interpreted and explained more substantively, taking changes in input and process into account (Gustafsson, 2008; Hanushek & Woessmann, 2010); (c) the use of repeated cycles increases insight into the issues that challenge the comparability of scores across countries. In the early stages of PISA, policy makers and researchers may have been reluctant to interpret "soft" constructs, such as school climate, students' motivation, teachers' job satisfaction or parents' self-reported engagement, fearing their insufficient cross-cultural comparability. However, now that trends are available, the focus can be changed within countries, where the cross-sectional design of the study also enables cross-national comparisons of education and its outcomes.

The addition of global competence as a new domain in PISA 2018 has required and will continue to require thoughtful integration of the activities of the questionnaire expert group, reading expert group, and global competence expert group. This integration is achieved in two ways: first, one of the members of the questionnaire expert group is also a member of the reading expert group whilst another member of the questionnaire expert group is also a member of the global competence expert group; second, meetings were and will continue to be held between the chairs and leads of all three expert groups. The framework reflects the outcomes of these coordination activities with the expert groups.

C. Outline of the Framework

This document provides the background for the development of the background questionnaire. The framework defines all major domains and constructs (this term refers to what is measured by one or more items of a scale) that will be assessed in the background questionnaires of PISA 2018, referring to student background constructs, schooling constructs and non-cognitive and metacognitive constructs. The framework describes current literature on the constructs covered in the background questionnaire. The focus of the framework is much more on the why and how of assessing constructs than on their relations. The educational literature has a wide variety of theoretical models that have been proposed to link systemic or curriculum variables (e.g. instructional approaches and educational expenditure) to student achievement. Like most frameworks that were developed in the past, the present framework does not provide a “theory of PISA” that explains how all measured constructs are supposedly linked, although it is clear that many previous publications of PISA data address these links. A “yield study” such as PISA in which a set of tests is administered to a group of students only once has a limited scope for causal analysis. Therefore, the emphasis in the present framework is more on identifying constructs that are informed both by previous PISA cycles and current literature than on specifying their links.

The document is organised into two main parts: (1) defining the core content of the PISA questionnaires and elaborating the modular structure and (2) explaining the broad coverage of policy issues. Detailed references to current research are provided throughout the document.

The first part of this document links the current framework to the overarching (cross-cycle) structure of the PISA context assessment set out in the PISA 2012 and 2015 frameworks (Klieme et al., 2013; Klieme & Kuger, 2014; OECD, 2013a, p. 168 ff.). The constructs that need to be covered for monitoring trends in education are revisited, with reference to the general background of educational effectiveness research. These are measures that have been used previously for initial reporting, for international indicators (published in “Education at a Glance”) and for secondary analyses.

The second part of this document explores the in-depth policy issues that will be covered in PISA 2018, which will be organized by modules. A *module* comprises one or more related constructs (assessed by items or scales); for example, the module on domain-general student attitudes and behaviours is composed of various scales, such as self-related beliefs and attitudes towards school, well-being and utilisation of information and communication technology (ICT). Additionally, this part of the document explains how the modules will be implemented in the field trial for PISA 2018.

Some newly developed questions related to various modules will be tested in the 2018 field trial, providing a broad set of measures that can be used in the PISA 2018 main study and/or in later cycles. Based on careful analysis of the field trial data and thoughtful discussion among experts and policy makers (including the PGB), modules, constructs, questions and items will be selected for inclusion in the PISA 2018 main study.

1. DEFINING THE QUESTIONNAIRE CORE IN PISA 2018

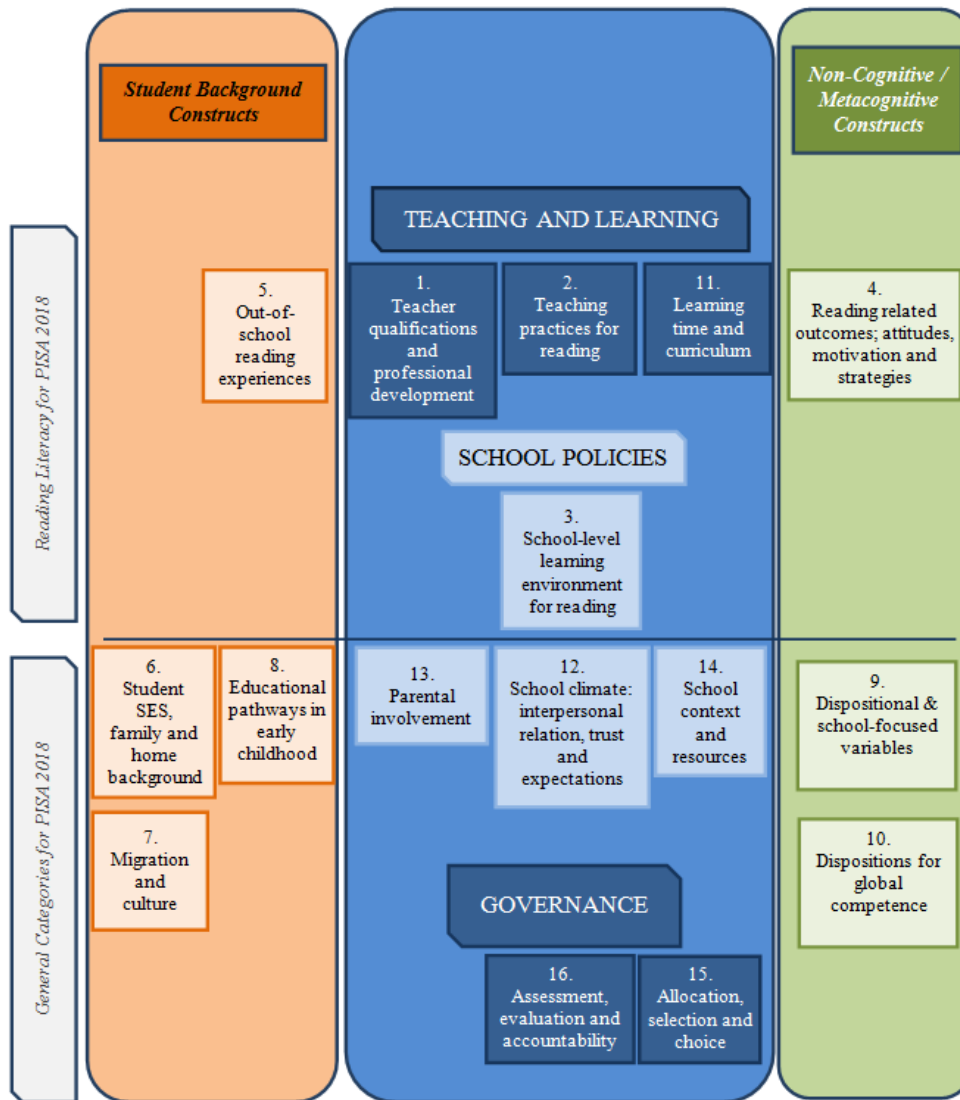
Choosing from the many measures that might be incorporated into the PISA study is informed by current educational research and findings of previous waves. One of the major forces driving the PISA design is the cyclical change of focus in the cognitive assessment: Reading literacy was the major domain of assessment in PISA 2000 and 2009 and is so again in PISA 2018, whilst mathematics was the focus of PISA 2003 and 2012; science took the lead in PISA 2006 and 2015. The major domain of cognitive assessment is the focus of domain-specific context assessment in the associated PISA questionnaire (e.g. various reading-related constructs assessed in the questionnaire for 2018 since reading is the major domain). However, there is also a need for stability in measures as identical measures administered in different waves can be used to gauge and understand trends in education. Stability has to be considered at two levels: on the one hand, there is the issue of stability across waves of three years (various questions in the questionnaires tend to recur in every cycle); on the other hand, stability can also be desirable in subject-specific constructs across waves of nine years (so, constructs assessed in the 2009 wave dealing with reading could be taken up again in 2018).

The questionnaire framework first established for PISA 2012 and continued for PISA 2015 has an overarching structure that delineated core questionnaire content that should be kept comparable across cycles (OECD, 2013a, p. 189 ff.) to allow for continuous monitoring of educational systems. The overarching framework refers to domain-specific as well as domain-general measures assessing conditions, processes and outcomes of education, both for students and for schools. Finding an appropriate balance between these domain specific and domain general measures and between processes and outcomes is crucial for the long-term success of the PISA programme. In order to establish valid and reliable trends at the country level, it is important to implement a stable set of variables that will be used as major reporting variables across PISA cycles.

This overarching framework is taken up in the following, along with specifying the constructs and measures in more detail and providing arguments that support the choice of core content for PISA 2018.

A. Outline of content covered in questionnaires:
(a) general framework of the 2018 study and modular approach, (b) constructs to be covered, and (c) global competence

Figure 15. PISA 2018 Questionnaire Modules



Columns one and two of Figure 15 (student background constructs and schooling constructs; modules 5 through 8) summarise student background characteristics related to their family and the education they received. The items associated with these first two columns shaded orange are typically asked of students or parents. The constructs in the middle column, which are coloured blue (modules 1 through 3 and 11 through 16), refer to educational processes on different levels (system governance, school policies, and teaching and learning). The questions associated with these columns are asked mainly of schools but may have some questions that students or parents answer. The last column on the right in the green colour (modules 4, 9 and 10) lists various non-cognitive and metacognitive constructs of students (referring to awareness of strategies), and these are asked mainly of students. The lower half of the figure deals with domain-general topics, whereas the upper part includes modules that mainly deal with domain-specific (in this case: reading-related) topics. Thus, the figure illustrates an approach using the combination of domain-general and domain-specific topics that is typical for all PISA cycles, with reading, mathematics or science being the major focus of assessment.

Every module represents a focus of policy making. Thus, the set of 16 modules covers a wide array of policy issues that are relevant across countries. This set is quite comprehensive, as can be seen by comparing the modular structure with current literature on educational policy. For example, most topics treated by Sykes, Schneider, and Plank (2009) and by OECD (2015) in their state-of-the-art review of educational policy research are covered here.

The organisation of this document uses a combination of columns and rows in Figure 15 to group topics into coherent and organised sections rather than simply following a numbering schema. This grouping was also done for the 2015 PISA questionnaire framework. Non-cognitive and metacognitive constructs, which are contained in the green column, are covered first and contain modules 4, 9 and 10. Student background constructs, which are the orange column, are covered next and contain modules 5, 6, 7 and 8. Teaching and learning constructs, which are at the top of the middle blue columns, are addressed next. These modules include 1, 2 and 11. Lastly, school policies and governance constructs are discussed last and include modules 3, 12, 13, 14, 15 and 16.

PISA treats the core questionnaires (school questionnaire and student questionnaire) separate from optional questionnaires that countries may choose to implement. PISA 2018 will still keep them separate from an operational point of view, but the questionnaire expert group emphasises connections among the questionnaires in its framework and intends to make as transparent as possible the connections (such as complementarities) between standard questionnaires and options in the questionnaire development stage of item selection or writing so that it remains clear what is conceptually covered in each questionnaire and it is clear which constructs are available at the individual and school level, and which constructs are assessed in which part of the total sample.

- Reading as the major domain

For PISA 2018, a new reading framework has been developed. Even if this new framework shares a lot of communalities with the 2000 and 2009 ones, it is conceptualised in order to address the main difference between print and online reading (Afflerbach & Cho, 2010) – the fact that in online reading, the text is not given; the reader has to build his/her own text, choosing which paths to follow, and which ones to dismiss, in a context where the reader is offered many more options and opportunities in which to get lost. The PISA 2018 framework for reading literacy aims at addressing the additional complexities linked to online reading comprehension as defined by Coiro and Dobler (2007), such as additional sources of prior knowledge (knowledge about

search engines, website structures), higher incidence of multilevel forward inferential reasoning (predicting what is behind the link) and new dimensions of self-regulation reading (integration of physical actions such as clicking, scrolling down with cognitive processes such as predicting, assessing and evaluating the quality of information). Moreover, the new reading framework highlights the importance of task management processes, such as setting goals and plans as well as monitoring and regulation, which are important metacognitive constituents.

In adherence with these new developments linked to online reading in the PISA 2018 reading framework, all the reading related non-cognitive, teaching, learning and curriculum constructs and questions should cover not only print, but also online reading (e.g. online reading practices, motivational aspects, metacognition and teaching practices related to online reading).

- *Global Competence as a new domain*

For 2018 the new domain global competence is being introduced. This domain is seen as critically important because our learning, working and living environments are becoming more global, interconnected and interdependent. Global competence is required for many people in the world, at present, and will be for all young people going forward. Young people will encounter, actively engage with and help shape those environments during their lifetime no matter where they are born educated, work or live. Given this context, young people need to leave school equipped with the necessary knowledge, skills and attitudes that will enable them to learn, work and live in a globalised world and that they can develop further as they move through life. They need to leave equipped with the following: a knowledge of and interest in engaging with the world around them; a growing confidence and a spirit of curiosity, adventure, flexibility and resilience; and the communication and interaction skills necessary to make the most of the opportunities and challenges that fast-changing, interconnected and interdependent environments bring. Students need to experience classrooms and schools that foster the value of, and embrace the diversity of, peoples, languages and cultures. The schools need to encourage intercultural sensitivity and help students move beyond tolerance to acceptance, respect and appreciation and away from ethnocentric world views. Our students also need to engage in experiences that facilitate international and intercultural relations, exchanges and conversations and encourage reflection upon, and understanding of, the learning outcomes from such experiences (Bennett, 1993; Sinicrope, Norris, & Watanabe, 2007).

Several authors have signalled that because of the increasing speed of development in society (information and communications technology [ICT] and cross-border working), schools need to adapt their curricula to account more precisely for what students will need in their future lives (e.g. Fisch, McLeod, & Brennen, 2009). For example, Anderson (2008) points out that the knowledge and skills required to prosper in 21st century society go far beyond the traditional literacies. He identifies knowledge construction, adaptability, finding, organising and retrieving information, information management, critical thinking and teamwork as the skills demanded by modern societies. Meanwhile, Binkley et al. (2012), in their framework for 21st century skills, maintain that achieving competence requires specific knowledge, skills, attitudes, values and ethics.

- *Constructs to be covered in the questionnaires*

Since the beginning of PISA in 2000, the background questionnaires have served two interrelated purposes. The first purpose has been to provide a context through which to interpret the cognitive assessment scores (both within and between education systems). This context has been most often situated in a general form of the inputs-processes-outputs model of education systems. The second purpose of the background questionnaire has been to provide reliable and valid non-cognitive outcomes, which can inform policy and research in their own right. Over the six cycles of PISA to date, new non-cognitive outcomes have emerged for both domain-specific and domain-general features of education. Enhancements to the background questionnaire have also tracked developments in psychometric theory and survey research methodology for the purpose of providing increasingly reliable and valid measures of non-cognitive constructs that are not sensitive to cultural differences in response style. These developments have occurred while maintaining the goal of revealing trends across PISA cycles.

PISA 2018 is the seventh cycle of PISA and the third full cycle where reading is the major domain of assessment. In addition, PISA 2018 will also introduce a new construct – global competence – while diminishing the distinctions among the major and minor domains and maintaining trend items for reporting purposes. The goals of PISA 2018, therefore, require new thinking with respect to the development and design of the background questionnaire. Specifically, the goal of diminishing the distinctions among the major and minor domains, along with the addition of the global competence construct, will require additional questionnaire delivery time.

This section of the questionnaire framework presents the constructs for PISA 2018 and is organised around: (1) non-cognitive and metacognitive constructs, (2) student background constructs, (3) teaching and learning constructs, and (4) school policies and governance constructs.

a. Non-cognitive and metacognitive constructs

The main challenge of PISA concerns measuring and documenting the outcomes of education that have been reached up to the age of 15 years. At its essence, educating a person means *fostering* his or her *individual development* as a unique, self-determined, knowledgeable person who gradually gains in ability to participate in society. As each PISA assessment is a cross-sectional study, PISA does not capture developmental processes in the same way that longitudinal studies can, but PISA serves as a snapshot of developmental status at the age of 15. Though this snapshot includes an assessment of achievement in reading literacy, mathematics and science, other outcomes are important. Success in school – and in life – also depends on being committed to learning, sharing values and beliefs, respecting and understanding others, being motivated to learn and being able to regulate one's own learning behaviour. These constructs can be perceived as prerequisites of learning, but they may themselves also be judged as goals of education, as elaborated in the OECD project *Defining and Selecting Key Competencies* (DeSeCo) (Rychen & Salganik, 2003). Educational research has shown that non-cognitive factors are very important for individual development as well as for success in life and well-being, and thus have an impact on individuals and society alike (Almlund, Duckworth, Heckman, & Kauth, 2011; Heckman, Stixrud, & Urzua, 2006).

Given the increasing importance of non-cognitive outcomes, PISA complements the assessment of cognitive, learning-related behaviour (e.g. self-regulation, strategies and invested time) with non-cognitive and metacognitive outcomes, such as attitudes, beliefs, motivation and aspirations. These outcomes are measured mainly in the student questionnaire, but also in the school questionnaire. These outcomes may be of a general nature, such as the achievement motivation and well-being of students and the drop-out rates of schools, or related to the domains of the cognitive assessment, such as reading engagement, interest in mathematics, or enjoyment of science. Domain-specific non-cognitive outcomes are also mentioned in the respective definitions of literacy, so this array of constructs serves as a link between the cognitive frameworks and the questionnaire framework.

b. Student background

In order to understand learning outcomes, educational careers and equity issues within and across countries, family background variables, such as socio-economic status (SES) and ethnic background, must be taken into account. The distribution of educational opportunities and outcomes depending on these background variables enables a comparison of countries in terms of opportunities provided to students.

PISA has become well known for its detailed, theory-based assessment of family background, SES and immigration background. Much effort has gone into the definition and operationalisation of individual student background indicators, finally leading to the establishment of a powerful, integrated indicator for economic social and cultural status, also known as ESCS (Willms, 2006). The components of this indicator need to be assessed in as stable a way as possible across the PISA cycles. In addition, information on parental support fosters understanding of how formal education and family background interact in promoting student learning.

Furthermore, PISA gathers retrospective and prospective information about educational pathways and careers across the lifespan. In recent years, researchers as well as public debate in many countries have stressed the importance of early childhood education (Blau & Curie, 2006; Cunha, Heckman, Lochner, & Masterov, 2006). Therefore, PISA intends to collect at least some information on primary and pre-primary education participation, bearing in mind that, for the most part, this would be solicited from 15-year-olds, which could challenge the validity of the reports.

Beyond individual student background, the social, ethnic and academic composition of the school has an impact on students' learning processes and outcomes. Therefore, PISA uses aggregated student data to characterise background factors on school level in addition to structural factors, such as the location and size of a school.

c. Teaching and learning

School-based instruction is the core process of formal, systematic education. Therefore, policy makers need information on teaching, learning and the organisation of schools. To increase the explanatory power of the study, assessment of teaching and learning will focus on the *major domain of assessment*, which in 2018 is reading, as well as the experimental domain for 2018, global competence. The knowledge base of educational effectiveness research allows for the identification of core factors with an expected bearing on reading: teachers' qualifications, teaching practices, classroom climate, learning time and learning opportunities provided both within and outside of school (Creemers & Kyriakides, 2008; Scheerens & Bosker, 1997). For teaching processes, the focus should be on three basic dimensions (Klieme, Pauli, & Reusser,

2009): (i) structure and classroom management; (ii) teacher support; and (iii) cognitive challenge. Addressing teacher and teaching-related factors in PISA is difficult because sampling is by age rather than by grade or class. Another challenge is linked to the reading domain itself. When students are 15 years old, reading is no longer taught as a standalone subject in the same way that mathematics and science are. However, reading literacy growth is still stimulated by teaching practices, and reading strategies are taught or learned through test language lessons and language arts and literature courses, but also through foreign language lessons and social and natural science courses, known in its entirety as “content literacy” (McKenna & Robinson, 1990; Shanahan & Shanahan, 2008). While for mathematics and science, it seems obvious that the questions about teaching and learning should be related to mathematics and science lessons only, there is clear evidence that in order to get rich and valuable information about reading (especially online reading) investigating only test language instruction lessons would result in an impoverished information. One of the most striking differences between countries in the reading curriculum nowadays is the emphasis on and time dedicated to content literacy (i.e. teaching reading in other subjects) (Lafontaine, Dupont, Hindrycks, & Schillings, 2015). Consequently, if a teacher questionnaire would be implemented in PISA 2018, it would be more relevant to administer it to a sample of teachers across domains than to test language teachers only.

d. School policies and governance

As policy makers have limited direct impact on teaching and learning processes, it is important to gather information on school-level factors that help improve schools, and thus indirectly improve student learning. As with teacher and teaching variables (Barile et al., 2012), school effectiveness research has built a strong knowledge base showing that “essential supports” promote school effectiveness (Bryk, Sebring, Allensworth, Luppescu, & Easton, 2010; Chapman et al., 2011). These essential supports comprise professional capacity and a focus on professional development; a well-organised curriculum; leadership and school management; parental involvement; a challenging but nurturing school climate (clear norms and shared values, high achievement expectations, truthful and mutually supportive interactions between stakeholders); and the use of assessment and evaluation for improvement. These factors will be addressed within the PISA questionnaires as domain-general processes on the school level. In addition, school-level support for teaching the major domain will be covered, such as the provision of libraries, ICT equipment, and a school curriculum for reading literacy, including multimodal aspects of reading in a digital era.

The PISA 2018 questionnaires also need to address issues related to governance at the system level (Hanushek & Woessmann 2011; Woessmann, Lüdemann, Schütz, & West, 2007). Allocation, selection and choice as well as assessment and evaluation are the basic processes that policy makers and/or school administrators use to control school quality to monitor and foster school improvement. Some of this information can be gained from other sources (as documented in OECD’s *Education at a Glance*), while other information can be assessed through the PISA school questionnaire.

B. Previous use of PISA context data: Measures that have been and will be important for analysis and reporting

a. Comparability and trend items

In previous cycles statistical analyses were conducted to address the question of whether the administered scales measured identical underlying constructs in all participating countries and whether scores could be compared across countries. Exploratory and confirmatory factor analyses have been used in previous rounds to test this comparability and to establish whether constructs and scores can be viewed as invariant across countries (e.g. OECD, 2012, 2014b). The questionnaire expert group expects that addressing invariance issues will continue to play an important role in the analyses of questionnaire data and that state-of-the-art procedures will be employed to examine comparability across countries.

An important asset of the PISA study is the use of trend items. There are a fair number of questionnaire items that have been used in at least one previous round. The questionnaire expert group intends to retain trend items in the 2018 questionnaire as much as possible to ensure that data can be compared with previous rounds and trend analyses can be conducted. After the questionnaire framework has been established, a detailed overview of items that have been used in one or more previous cycles will be prepared. This overview will then provide important input for decisions on the selection of items to include in the 2018 questionnaire.

b. The PISA 2009 report

This section illustrates how PISA data from both the cognitive instrument and the questionnaires has been used by the OECD in the past to create various volumes as discussed below as well as thematic reports. It was used to frame the thinking of what variables should be included in the 2018 study. This section focuses on reading – the major domain of 2018 – when it was last the major domain in 2009. This section describes which background variables were used in the PISA 2009 Report. The PISA 2009 report made use of questionnaire materials, especially for the diverse and sophisticated analyses presented in Volumes III and IV. In addition to student achievement, non-cognitive outcomes, such as student engagement, cognitive strategies, and metacognitive strategies, were studied in detail, and the impact of background variables, individual support factors, school level input, processes and policies, as well as system-level factors was reported – all assessed in the student and school questionnaires. In more detail, the six volumes describing PISA 2009 results discussed the following questionnaire data:

- *Volume I: Student Performance in Reading, Mathematics, and Science*
 - Student background: gender
- *Volume II: Overcoming Social Background: Equity in Learning Opportunities and Outcome*
 - Student background: economic, social and cultural status (ESCS), gender, immigration status, language spoken at home, age of arrival, country of origin
 - Individual support assessed through parent questionnaire: parental support (at beginning of primary education/at age 15), pre-primary education (attendance, quality)

- *Volume III: Learning to Learn: Student Engagement, Strategies, and Practice*
 - Student background: ESCS, gender, immigration status, language spoken at home
 - Outcomes: enjoyment of reading, time and material used for reading, metacognition (awareness of strategies), self-reported use of reading strategies (memorisation, elaboration, control)
- *Volume IV: What Makes a School Successful? Resources, Policies, and Practices*
 - Student background: socio-economic status (SES)
 - Student-reported processes: learning time (previous education, learning time at school, enrichment/remedial education, after-school lessons), teacher-student relations, disciplinary climate, teacher's stimulation of reading engagement
 - School input, policies and processes (reported by the principal): type of school (public/private), number of programmes, class size, educational resources (e.g. ICT, library), school responsibility for assessment and curriculum/for resource allocation, extra-curricular activities provided, age of school entry, grade repetition, school admittance/grouping/transfer policies, assessment practices/purposes, use of achievement data, school accountability, methods for monitoring teachers, teacher and student behaviour, parent involvement and expectations, leadership, school climate
- *Volume V: PISA 2009 Results. Learning Trends: Changes in student performance since 2000*
 - Trends in equity: socio-economic status (SES), migrant status and language spoken at home
 - Trends in non-cognitive outcomes: reading attitudes and practices (reading for pleasure, diversity of texts read, reading engagement by SES, reading fiction), and school climate indicators (teachers-student relations and disciplinary climate)
- *Volume VI: Students On Line. Digital technologies and performance.*
 - ICT familiarity (optional questionnaire): access to ICT at home and at school, use of ICT at home and at school, students' attitudes towards and self-confidence in using computers, self-confidence in doing ICT tasks and activities; navigation indices extracted from log-file data (number of pages visited, number of relevant pages visited)

In PISA 2000, in addition to the main international report, an in-depth thematic report was dedicated to reading (Kirsch, De Jong, Lafontaine, McQueen, Mendelovits, & Monseur, 2002).

As will be outlined below in more detail, most measures that were described in the PISA 2009 Report are prioritised in the 2018 instruments, thereby ensuring continuity and the opportunity to compare reported findings between 2009 and 2018.

c. Education at a glance and research publications

Several “PISA in focus” publications explore issues related to reading non-cognitive outcomes: *Do students read for pleasure? (n° 8)* *Are boys and girls ready for the digital age? (n° 12)* *Could learning strategies reduce the performance gap between advantaged and disadvantaged students? (n° 30)*.

Many additional examples of scientifically productive analyses can be found in the literature. For example, the ERIC international database on educational research currently lists numerous peer reviewed journal articles that are using PISA 2009 data to study reading education. The majority of these papers discuss non-cognitive, domain-specific outcomes: re-scaling the questionnaire items, studying their structure within and across countries, analysing patterns of outcomes, looking at the impact of student and family background and identifying and explaining school effects.

2. COVERAGE OF POLICY ISSUES FOR 2018

For PISA 2018 a balance must be struck between the need for trend and the need for new or changed constructs. Where possible and sensible, constructs and modules will be carried forward intact or with only minor changes. If measures are outdated, redundant, or do not comply with psychometric criteria (such as low internal consistency), they will be recommended for deletion. Lastly, and importantly, there will be additions based on two issues: (1) constructs central to the research literature that have not been previously covered to date will be included and (2) constructs relevant to the new domain of global competence will be introduced. Existing constructs will be extended to cover the new domain of global competence or the new dimensions in reading covering online reading. In the future, some new scales measuring existing constructs will be developed, in order to improve the measurement and try to avoid existing cross-cultural biases linked to the use of Likert scales. Additionally, not every construct needs to be tested during every cycle.

This section is divided into four subsections that group the modules into the larger constructs of: (1) assessing non-cognitive and metacognitive constructs; (2) assessing student background; (3) assessing teaching and learning processes; and (4) assessing school policies and governance. The subsection on assessing non-cognitive and metacognitive constructs has the largest number of changes for 2018. In addition to a discussion of reading-specific variables and some significant changes in the current literature, the new domain of global competence is presented in this subsection.

A. Assessing non-cognitive and metacognitive constructs

This subsection summarises the conceptual foundations for modules 4 (reading-related outcomes: attitudes, motivation, attitudes & strategies), 9 (dispositional & school-focused variables), and 10 (dispositions for global competence) (see Figure 15).

Traditionally, PISA has assessed student outcomes in terms of achievement tests. Students' motivations, attitudes, beliefs and behaviours were seen as important precursors and predictors of scholastic performance, educational attainment and labour market success. However, from the standpoint of both educational policy and labour market policy, non-cognitive outcomes, which are separate from academic achievement, are seen as increasingly important in their own right, because they are instrumental for personal growth, individual success, long-term achievement and the society as a whole. Research findings have shown the considerable power of non-

cognitive outcomes for success in secondary education, higher education and the workforce (e.g. Heckman, Stixrud, & Urzua, 2006; Lindqvist & Vestman, 2011; Poropat, 2009; Richardson et al., 2012; Roberts et al., 2007). Additionally, professional and public debates often question the purely achievement-based approach that student assessments have mostly taken in the past; it is widely acknowledged that there is more to education than knowledge and cognitive skills. Therefore, non-cognitive outcomes become increasingly interesting as standalone outcomes (Marsh et al., 2006). Non-cognitive outcomes are important constructs, and they often function as moderators and mediators for relations with other constructs in the assessment. PISA offers a unique possibility of investigating complex relations between non-cognitive outcomes and achievement at the individual, school and country levels.

Previous PISA cycles have focused on domain-specific student attitudes and behaviours, such as interest and motivation towards reading and mathematics, mathematics self-concept or maths anxiety, or knowledge of reading strategies (metacognition). Most of these scales display robust relations with student proficiency scores. This tradition is continued in module 4 (reading-related outcomes) in PISA 2018. In addition, the current framework includes a broader set of non-cognitive student factors to extend the coverage of centrally important constructs, to increase the policy relevance of the PISA 2018 database, and to acknowledge the increased interest in non-cognitive assessments both in policy and in research.

The questions included cover achievement-relevant dispositions that students carry into school settings, school-focused variables, and reading-specific variables. As such, the assessment adopts a distinctly hierarchical approach, acknowledging that both cross-situational predispositions and school/domain-specific constructs must be attended to in order to fully understand and explain student engagement and behaviour in the achievement domain (Elliot & Thrash, 2001). Domain-specific questions will focus not only on reading, but also on global competence. Due to time constraints, only constructs that have a rich and longstanding history in the achievement literature and/or that are highly popular in contemporary research (even if relatively new) are included in the questionnaire.

a. Reading-related outcomes (module 4)

- Reading motivation, engagement and practices

Reading engagement, motivation and practices have been shown in a number of studies to be strongly linked with reading proficiency (Becker, McElvany, & Kortenbruck, 2010; Guthrie, Wigfield, Metsala, & Cox, 2004; Klaua, & Guthrie, 2014; Mol & Bus, 2011; Morgan, & Fuchs, 2007; Pfof, Dörfler, & Artelt, 2013; Schaffner, Philipp, & Schiefele, 2014; Schiefele, Schaffner, Möller, & Wiegfield, 2012). In PISA 2000, engagement in reading (i.e. interest, intrinsic motivation, avoidance and practices) had one of the strongest relationships with reading proficiency, stronger than the association between reading literacy and socio-economic status (OECD, 2002; 2010). It was shown that a high level of engagement compensated, to some extent, for poor socioeconomic background. Indeed, at least in some countries, students issued from a low socioeconomic background but highly engaged in reading, outperformed students from a high socioeconomic background but poorly engaged in reading. It was also found that if boys were equally engaged in reading as girls, the gender gap would be reduced by two thirds (OECD, 2002, 2010). In other studies, reading engagement has been shown to explain reading achievement more than any other variable besides previous reading achievement (Guthrie & Wigfield, 2000). Thus, motivation and engagement are powerful variables, and levers on which one can act in order to enhance reading proficiency and reduce gaps between groups of students.

Among motivational attributes of the reader, the instruments developed in PISA in the past mainly targeted interest and intrinsic motivation. When mathematics and science were the major domains, other motivational constructs were investigated, such as self-efficacy and self-concept. Typically, self-efficacy (Bandura, 1997; Ferla, Valcke, & Cai, 2009) refers to the individual's perceived capacity of doing specific tasks. Whereas self-concept is a general measure of the individual's own perceived abilities related to a domain (reading, mathematics or science) (Marsh & Craven, 1997). Positive self-concept and self-efficacy are highly related to motivation, learning behaviour, general expectations for the future and students' performance (OECD, 2007). Because self-efficacy and self-concept are important motivational attributes and proved to be strong correlates of reading achievement (Baker & Wigfield, 1999; Marsh & Craven, 2006; Solheim, 2011), and even predictors of reading growth in secondary schools (Morgan & Fuchs, 2007; Retelsdorf, Köller, & Möller, 2011), these constructs are measured in the PISA 2018 students' questionnaire. Following Chapman and Tunmer's recommendations (1995), the instruments should cover not only perceptions of *competence* in reading, but also perceptions of *difficulty* with reading. Klauda and Guthrie (2015) in a longitudinal study investigating relationships between several motivational and engagement constructs and reading achievement (fluency and general and information texts comprehension), have provided evidence that perceived difficulty is a stronger predictor of reading achievement than self-efficacy. Similarly, they confirmed that avoidance and devaluation negatively predict growth in engagement and motivation indicators among grade 7 pupils, independent of their positive counterparts – self-efficacy, engagement in reading tasks and value.

Identifying students who perceived themselves as struggling readers is of huge importance, especially because PISA 2018 emphasises the importance of basic components of reading such as fluency and because the reading scale has been extended at the lower end since PISA 2009.

- *Metacognition*

Like engagement, metacognition has both a significant correlation with reading proficiency and is responsive to teaching and learning. The prominent metacognitive reading strategies include setting reading goals, adapting reading depending on these goals, knowing how to summarise a text or remember essential information, monitoring comprehension and knowing how to repair comprehension problems. The new PISA 2018 reading literacy framework acknowledges the paramount importance played by these reading task management processes. They are now an integral part of the model of reading processing organizing the reading literacy framework (see figure 2).

A number of studies have found a strong association between reading proficiency and metacognition (Artelt, Schiefele, & Schneider, 2001; Brown, Palincsar, & Armbruster, 2004). Explicit or formal instruction of metacognitive strategies leads to an improvement in text understanding and information use. One main finding of the report of the U.S. National Reading Panel (2000) was that remediating poor reading literacy is possible through explicit teaching of metacognitive skills. That is, when readers are given cognitive and metacognitive strategy instruction, they make more significant gains on measures of reading comprehension than students only trained with conventional instruction procedures (Baker, & Carter-Beall, 2009; Dole, Nokes, & Drits, 2009; Pressley, Graham, & Harris, 2006; Pressley et al., 1989; Rosenshine & Meister, 1994; Rosenshine, Meister, & Chapman, 1996; Waters & Schneider, 2010).

In PISA 2009, students' metacognitive knowledge was assessed asking students about how useful they thought various reading strategies were in order to solve a reading task (summarising a text or understanding and remembering a text). A "benchmark criterion was applicable, because the judgments were evaluated relative to those of experts" (Artelt & Schneider, 2015, p. 6). Correlations of the two metacognition indices and reading achievement in PISA 2009 were robust: the median correlations in OECD countries were, respectively, 0.46 and 0.39.

The growing importance of digital reading literacy in PISA 2018 makes even more important the need to assess metacognition. Coiro and Dobler (2007) pointed out that in online reading, efficient and specific self-regulated strategies are crucial to facilitating reading goals and plans (select the most relevant links and pathways, avoid distracting information). A new metacognition scenario will focus on one of the most important processes of online reading, namely an assessment of the quality and the credibility of sources.

b. Dispositional and school-focused variables (module 9)

Complementing the reading-related outcomes are domain general dispositions toward competence and school-focused variables. Students enter school with broad ranging tendencies and schools encourage students to focus on their schoolwork in particular ways, some of which facilitate engagement and some of which undermine performance. Both dispositional variables and school-focused variables are needed to fully understand students' motivation and engagement. These variables have not been included in PISA questionnaires. Dispositional variables represent the personality-based context in which students approach or avoid learning; they emerge as a function of a lifetime of socialization from parents, teachers, coaches and cultural emphasis, and they capture how behaviour is energized over time. School-focused variables are context-specific and emerge not only out of dispositional tendencies, but also situational emphases; they are the proximal predictors of the aforementioned domain-specific variables, as well as achievement outcomes per se. Together, dispositional and school-focused variables provide important information regarding the degree to which a given student fits the portrait of an optimal learner and achiever.

The following paragraphs overview the target dispositional and school-focused variables, especially focusing on constructs that are new to PISA for 2018. These variables are domain neutral but focus on non-cognitive components important to learning. The dispositional variables include achievement motives of competitiveness, workmastery, and fear of failure; incremental mind-set; perseverance; subjective well-being; and ICT motivation and practices. The school-focused variables include learning beliefs and attitudes towards school and achievement goals.

- *Dispositional variables*

Achievement motives – Competitiveness, Workmastery, and Fear of Failure: The questionnaire expert group has replaced the construct of *achieving motivation* with the constructs of *competitiveness and workmastery*. Achieving motivation, as assessed in the 2015 field trial, represents a combination of competitiveness items and extrinsic motivation items. For the past four decades, achievement motivation theorists have conceptualised their central construct as comprising two fundamental components (Helmreich, Beane, Lucker, & Spence, 1978; see also Elliott & McGregor, 2001): Competitiveness (the dispositional desire to outperform others) and workmastery (the dispositional desire to work hard to master tasks). Research shows that these two components of approach-oriented achievement motivation are linked to different sets of antecedents and consequences; so, when assessing achievement motives, it is important to measure these constructs separately (Baranik, Barron, & Finney, 2007; Murayama & Elliot, 2012; Spence & Helmreich, 1983).

The questionnaire expert group has replaced the 2015 construct of general *test anxiety* with the construct of *fear of failure*. Like test anxiety, fear of failure is a dispositional tendency to worry about failure and its consequences. Continuing to assess this dispositional avoidance tendency can be accomplished by shifting the focal construct from test anxiety to fear of failure. Test anxiety is a multidimensional construct that is perhaps best assessed as worry at the task- or domain-specific level of analysis (Hembree, 1988). Fear of failure is the tendency to self-protectively avoid mistakes and failures because they are experienced as shameful, which may be more predictive of cognitive achievement in real life situations than test anxiety. This construct has been studied in the achievement literature for over forty years (Birney, Burdick, & Teevan, 1969). Research has shown that fear of failure leads students to be self-protective and to avoid challenging situations and opportunities that are essential for learning and development (Covington, 1992; Heckhausen, 1975; Kaye, Conroy, & Fifer, 2008).

The optimal learner is high in workmastery and low in fear of failure. Competitiveness alone can be problematic, but the confluence of high competitiveness and workmastery appears to be beneficial. The positive/negative implications of competitiveness is a “hot topic” in the achievement literature, and data on this variable both within and across countries should prove valuable and garner considerable attention.

Incremental mind-set: Incremental mind-set has been one of the most popular variables in the literature for the past fifteen years (Dweck, 1999). Students with an incremental (relative to entity) mind-set believe that ability is changeable rather than fixed, and this is another core characteristic of an optimal learner. Incremental mind-set is related to grit and is considered a key predictor of mastery-approach goals (increase) and performance-avoidance goals (decrease), and has also been shown to facilitate persistence upon failure and performance attainment (Aronson, Fried, & Goode, 2002; Blackwell, Trzesniewski, & Dweck, 2007; Dweck, 2007).

Perseverance: This construct was included in the 2012 student background questionnaire and was not included in the 2015 wave because of time constraints. The possibility of looking at trends in perseverance at a time interval of six years is valuable given that the perseverance construct is currently very popular in the literature because it has been shown to be an important predictor of achievement (Duckworth, Peterson, Matthews, & Kelly, 2007; Tutu & Constantin, 2012). Many different labels are used in the current literature for this construct, including “persistence” and “grit”. The optimal learner is high in perseverance.

Subjective well-being: Subjective well-being can be defined as “good mental states, including all of the various evaluations, positive and negative, that people make of their lives and the affective reactions of people to their experiences” (OECD, 2013a, p. 10; see also OECD, 2011b, 2013b). This definition encompasses three elements of subjective well-being: *life evaluation* – one’s reflective assessment of one’s life (including the single “general life satisfaction” question); *affect* – an emotional state, typically at a particular point of time; and *eudaemonia* – a sense of meaning and purpose in life, which can increase a sense of belonging. The recent growing interest from researchers and policy makers in this construct has resulted in recommendations to statistical agencies to “incorporate questions on subjective well-being in their standard surveys to capture people’s life evaluations, hedonic experiences and life priorities” (Stiglitz et al., 2009, p. 216). OECD (2013) has responded to this charge in providing guidelines on measuring subjective well-being. To date, 27 out of 34 OECD national statistical offices have committed to collecting at least the minimal information proposed by the OECD guidelines (the single “general life satisfaction” question). For PISA 2018, the QEG has included information on all three elements, *life evaluation*, *affect*, and *eudemonia*. The optimal learner has a positive life evaluation, frequent positive affect and infrequent negative affect, and strong eudemonia.

ICT motivation and practices: Module 9 covers one aspect of technology that is relevant in everyday lives and across all educational domains, namely ICT. ICT-related behavioural characteristics and motivational attributes can be regarded as domain-general student outcomes. Since ICT subsumes a broad range of devices, it may play a role across all educational domains. Following the OECD’s DeSeCo project and the 21st Century Skills Initiative, The optimal learner will exhibit general skills related to information, media, and technology above and beyond the traditional core subjects (OECD, 2005; Partnership for 21st Century Skills, 2008). PISA 2018 will assess students’ interest in ICT, practices of ICT use, perceived competence and autonomy in using ICT and a specific question on use of social media within the ICT familiarity questionnaire.

- *School-focused variables*

Learning beliefs and attitudes towards school: Beliefs about one’s own success or failure in school learning have been shown to be strong predictors for further effort and success, including test scores in student assessments (Opdenakker & Van Damme, 2000; Rumberger & Palardy, 2005). PISA 2018 continues with several 2012 assessments, including measure of students’ school self-efficacy beliefs, students’ evaluation of their experience in school, and students’ attitudes toward school. The optimal learner has strong school self-efficacy and a positive inclination toward school.

Achievement goals: Mastery-approach and performance-avoidance goals. Achievement goals have been one of the most popular variables in the achievement literature for the past three decades (Maehr & Nicholls, 1980). One important characteristic of optimal learners is that they are focused on improvement in the classroom (i.e. they pursue mastery-approach goals). Students adopting mastery-approach goals have been shown to engage in deep learning, to persist upon failure and to show high levels of intrinsic motivation (Hulleman, Schrager, Bodmann, & Harackiewicz, 2010; Kaplan & Maher, 2007; Middleton & Perks, 2014). An equally important characteristic of an optimal learner is that he/she is *not* striving to avoid performing worse than other students (i.e. he/she does not pursue performance-avoidance goals). Students adopting performance-avoidance goals have been shown to engage in shallow learning, to give up in the face of failure, and to evidence low levels of both performance and intrinsic motivation (Hulleman et al., 2010; Rawsthorne & Elliot, 1999; Van Yperen, Blaga, & Postmes, 2014). In short, achievement goals are considered key proximal predictors of the two central, gold standard outcomes indicative of sustainable student success – performance attainment (which

shows that short-term learning has taken place) and intrinsic motivation (which shows that the motivation for continued, long-term learning is in place). Focusing on both the presence of the positive (mastery-approach goal pursuit) and the absence of the negative (performance-avoidance goal pursuit) is important, as both are essential for optimal learning to take place (Elliot, 2005).

c. Dispositions for global competence (module 10)

Global competence is set to be the new domain for PISA 2018 and the assessment procedure is still in the process of being developed by the relevant expert group. This work is seen as critically important because learning, working and living environments are becoming more global, interconnected and interdependent. Young people will encounter, actively engage with and help shape those environments during their lifetime no matter where they are born, educated, work and live. It is important that students leave school equipped with a knowledge of, and interest in engaging with, the world around them; a growing confidence, spirit of curiosity and adventure, flexibility and resilience; and the communication and interaction skills necessary to make the most of the opportunities and challenges that fast-changing, interconnected and interdependent environments bring.

For PISA 2018, global competence is defined as the capability and disposition to act and interact appropriately and effectively, both individually and collaboratively, when participating in an interconnected, interdependent and diverse world. The domain global competence is made up of the following four dimensions:

- **Communication- and relationship-management** refers to the willingness and capability to adapt one's communication and behaviour in order to interact appropriately and effectively with others holding diverse perspectives and in different contexts.
- **Knowledge of⁵ and interest in global developments, challenges and trends** refers to a learner's interest in and knowledge of cultures, major issues, events and phenomena in the world, as well as the learner's ability to understand their global significance and their implications for adapting appropriately and effectively to learning, working, and living situations with others holding diverse perspectives and in different contexts.
- **Openness and flexibility** refers to being receptive to and understanding of new ideas, people and situations, as well as differing perspectives and practices. It also refers to the ability to seek out and understand new and differing perspectives and experiences and to appropriately and effectively adapt one's thinking, behaviours and actions to learning, working and living situations that involve others holding diverse perspectives and in different contexts.
- **Emotional strength and resilience** refers to the ability to deal appropriately with the ambiguity, changes, and challenges that different perspectives and experiences can present and to have the resilience to maintain one's identity and/or develop personally despite or as a result of encountering different perspectives and experiences.

⁵ Refers to implicit and explicit and procedural and declarative knowledge.

Generally, the questionnaire items related to global competence (construct-related and contextual information to inform the interpretation of outcomes) will focus on the two dimensions, *openness and flexibility*, and *emotional strength and resilience*. These two dimensions will have four scales: (1) openness; (2) flexibility; (3) emotional strength; and (4) resilience, with approximately 10 questions per scale. The questions for each scale will primarily be behaviour-based (e.g. “When I meet people who are different from me, I am interested in learning more about them. I like to eat in a variety of ethnic restaurants.”), will include context in a simple way (e.g. “When I travel abroad...”; “When I am at school...”) and will be presented in a consistent format whenever possible (e.g. Likert scale).

Topics for the global competence questionnaire items could include (most measure comprise both construct and contextual components):

- Languages:
 - o Own languages spoken/understood by the student/by his/her mother/father/siblings at home
 - o Taught in school; number of languages spoken/understood by the student
 - o Number of languages to which the student has access or with which the student has experience (actual, virtual)
- Global developments/challenges/trends:
 - o The student’s engagement with others about global events/issues (e.g. chat online);
 - o Extent of exposure to/awareness of global developments/challenges/trends (e.g. via news, other media)
- Migration/movement of student:
 - o The student’s own background; the student’s experience living abroad
 - o The possibility of the student working/studying abroad in the future
 - o Demographics in his/her community from other countries/cultures
- Student interaction with or exposure to people from other countries/cultures; student’s travel experience
- Student’s degree of curiosity/motivation to travel

Existing questionnaire items (i.e. from the school and teacher questionnaires) could also be used or refined to inform the measure of global competence. For example:

- School climate:
 - o Student demographics (at school level)
 - o Teacher demographics (from other cultures), whether teachers taught in another country, whether teachers hold a certification from another country
 - o School philosophy /values/ policies related to global competence
 - o Support given to student for language and/or culturally-based reasons (e.g. revised to focus on global competence-related challenges)

- Global competence in the curriculum
- Professional development: how/degree to which teachers are prepared to manage multiculturalism and/or facilitate global competence (e.g. global trends, international events)
- Reading and resources related to global competence

Finally, if a teacher questionnaire is administered, then the following items could be used to measure global competence:

- Teacher demographics/ background (e.g. cultures represented)
- Number of teachers who have taught in another country
- Number of teachers who hold a certification from another country
- Exposure/experience/engagement with international, global events
- ICT literacy
- awareness of and support for school policies that relate to/support global competence
- Teacher practice related to global competence, including cross-curricular practices

Most existing measures in the area of global competence have been developed for older learners (e.g. college students, adult employees) rather than for the 15-year-old students who will be assessed in PISA 2018. Nonetheless, various measures can be derived from extant literature (e.g. Deardorff, 2009) and is a task that will be done by the global competence expert group with recommendations made to the questionnaire expert group. At present the intention is to use the questionnaire as a means to measure two of the four dimensions of global competence (i.e. openness and flexibility and emotional strength and resilience) across four scales (i.e. openness, flexibility, emotional strength and resilience), and contextual information.

B. Assessing student background

This subsection covers module 6 (student SES, family and home background), module 7 (ethnicity and migration), and module 8 (educational pathways in early childhood). These topics require careful revisiting every cycle because they contain the basic information needed for calculation of the Index of ESCS.

Student SES, family and home background (module 6): Whilst we do not request ethnicity directly in PISA, we will keep measures of socio-economic status and other background variables basically unchanged in order to maintain trend related to social and ethnic indicators. However, some minor changes have become necessary. Due to extensive development in the ICT sector, for example, questions on technology equipment in the student's home are slightly outdated and may need editing. Thus, the measures of home possessions will be updated to ensure better coverage of within- and cross-country variation of home possessions. These changes are expected not to impact the important trend measures in this module.

Migration and culture (module 7): Questions to inform indicators of both migration and culture continue to be important measures. Linguistic and cultural diversity are basic facts of life in most regions of the world. Many nations are home to several subpopulations with different languages and cultures. International migration perpetuates this diversity. In OECD countries, first and second generation immigrant students currently compose 10 to 20% of the student population

(OECD, 2010a; 2010b). At the same time, students from ethnic minority groups and immigrant students often face particular challenges. In a number of education systems, immigrant students perform at significantly lower levels than their native peers in key school subjects (Stanat & Christensen, 2006), and both groups are often faced with overt or covert discrimination with potentially detrimental consequences for their psychological development and well-being. Thus, providing students from different linguistic and cultural backgrounds with equal opportunities is often considered one of the central challenges for education systems in the 21st century (e.g. OECD, 2010).

PISA 2015 put a special focus on diversity-related aspects of the school climate. A new question developed for the PISA 2015 field trial asked students about their membership in a group that they believe to be discriminated against in their country. If they identify themselves as belonging to such a minority group, they are asked whether they feel treated in a respectful and fair manner by their teachers and equal to their classmates – a factor that has been shown to be related to educational outcomes (e.g. Fisher, Wallace, & Fenton, 2000; Wong, Eccles, & Sameroff, 2003). Another new question, implemented in the optional parent questionnaire, assessed perceived barriers to parental involvement. Additionally, teachers and principals are asked about diversity-related assumptions among teachers in their school. The wording of the question is based on research on multiculturalism support (Van de Vijver, Breugelmans, & Schalk-Soekar, 2008). Additionally, PISA 2015 examined palpable aspects of multicultural education practices or the extent to which multicultural educational practices are implemented in different schools. Altogether, findings from this module may help researchers better understand educational inequalities and can stimulate several ways to address these inequalities: teacher training programmes; school development activities focusing on diversity management strategies; or policy interventions supporting parents from diverse populations to get more involved in their children's schooling. However, these items have proved problematic and will need to be reworked if they are to be successful in measuring this important concept in 2018. Given the relevance of this module for global competence, such a reworking (presumably with more emphasis on opportunities for intercultural encounters) is worthwhile.

Educational pathways in early childhood (module 8): When children enter primary school, they have varying levels of ability in their language, pre-reading and early numeracy skills, and these differences are often maintained later in life. Promoting school readiness and better adjustment to school is hypothesised to be an efficient means of raising the achievement levels of all children, but especially of those children who experience a lack of parental support or who grow up in disadvantaged circumstances. It has been argued that investing in early education programmes will have large long-term monetary and non-monetary benefits (Heckman, 2006). The importance of pre-school quality has been acknowledged and analysed in OECD reports as well.

According to UNESCO (2006), Early Childhood Care and Education (ECCE) programmes are defined as “programmes that, in addition to providing children with care, offer a structured and purposeful set of learning activities either in a formal institution (pre-primary or ISCED 0) or as part of a non-formal child development programme” (p. 348). The focus of the internationally comparable International Standard Classification of Education Level 0 (ISCED 0) is much narrower. Currently at least four strands of research support the relevance of applying a broader definition of ECCE than focussing on ISCED 0 alone: (1) brain research, (2) studies on domain-specific development and support, (3) evaluation studies of model programmes, and (4) longitudinal large-scale studies. The broader definition of ECCE is commonly employed in these studies. Thus, conclusions about the importance of early child care should be drawn with ECCE and not with ISCED 0 in mind.

However, when evaluating the body of research, it becomes obvious that a number of characteristics of *the kind of ECCE provided* seem to influence whether benefits can be observed or not and whether these benefits disappear or persist. Students' early childhood opportunities to learn are best assessed in terms of curriculum and the quantity and quality of early childhood learning experiences. For example, one of the best sources available, the British EPPE study, did find short-term effects showing that pre-school attendance was beneficial for cognitive and socio-emotional development, in particular for children from disadvantaged backgrounds. However, in the long term only those children who attended a high-quality pre-school centre showed persisting beneficial pre-school effects (e.g. Sammons et al. 2008; Sylva, Melhuish, et al., 2011; cf. also Valenti & Tracey, 2009). Also, a certain degree of intensity in terms of hours per week/months seems to be a precondition for beneficial effects of ECCE attendance (Logan et al., 2011; Sylva, Stein, et al., 2011).

Thus, asking about early education experience in PISA only makes sense if specific aspects of dosage, quality and curriculum can be retrieved retrospectively, which is highly unlikely (Fivush & Hamond, 1990; Markowitsch & Welzer, 2009). As a consequence, PISA 2018, whilst keeping a short question on ISCED 0 attendance in the student questionnaire, will continue to administer a series of questions in the parent questionnaire, expecting parents to be the more reliable source of information. Those countries applying the optional parent questionnaire will gain information on basic characteristics of the early childhood education and care arrangements of PISA participants and reasons for attending or not attending early childhood education and care.

C. Assessing teaching and learning processes

This subsection summarises the conceptual foundations for modules, including module 1 (teacher qualifications and professional knowledge), module 2 (reading teaching practices), module 5 (out-of-school reading experience), and module 11 (learning time & curriculum).

Teaching and learning are at the heart of schooling. Most cognitive and non-cognitive, curricular and cross-curricular goals of school education are achieved – or impeded – by the way students and teachers interact in classrooms. Whilst teaching is the core process in schools, the curriculum determines its content, and professional teachers are the force who implement the curriculum, orchestrate learning activities and thus arrange for quality learning time.

PISA has been designed as a yield study that assesses skills and broad areas of literacy rather than curricular domains, and it samples a birth cohort rather than a grade level or intact classrooms. Thus, it might be questioned why this programme should address teaching and learning processes at all. However, there is ample evidence that teaching and learning activities are very good predictors of student competencies, whatever their character might be. So, if PISA is to inform educational policy making at the system and the school level, it must cover this important area. Clearly, the PISA study should focus on more general and internationally comparable constructs, rather than fine-grained content matter that considerably varies across countries. Therefore, module 2 describes reading education by broad lists of teaching and learning activities, drawn from several theories, such as self-determination theory. In addition, general dimensions of teaching quality, such as well-structured instruction, classroom management, support and cognitive activation, are applied to reading education. Furthermore, module 11 covers learning time – including non-mandatory, additional instruction within and out of school – as well as the coherence, focus and rigour of the reading curriculum. Finally, the teaching force will be described in terms of initial education, beliefs and professional development (module 1).

a. Reading teaching practices (module 2) and learning time and curriculum (module 11)

- Teaching practices and classroom support for reading growth and engagement in PISA

Research on reading shows that classroom practices, such as direct teaching of reading strategies (Pressley, 2000; Rosenshine, & Meister, 1997; Waters, & Schneider, 2010), on the one hand, teachers' scaffolding and support for autonomy, competence and ownership (Guthrie, Ho, & Kluda, 2013; Guthrie, Wigfield, & You, 2012), on the other hand, are powerful ways of improving students' reading proficiency, awareness of strategies (metacognition) and engagement (motivation) in reading.

As PISA 2018 is intended to inform educational policy, assessing variables in the 2018 Questionnaire that are responsive to policy decisions, among which reading literacy teaching practices is important. Two broad sets of theories or models inform the selection of the most relevant constructs:

- *Practices supporting reading engagement and motivation* are based on self-determination theory (Deci & Ryan, 1985; Reeve, 2012; Vansteenkiste, Lens, & Deci, 2006). The pathway to self-determination in reading for students depends on support from significant others. A favourable context assures competence in reading and expands autonomy in directing one's own reading activities. When the family and school context give the individual a sense of confidence (perceived competence/self-efficacy) and autonomy (being in charge of one's self) in reading, the individual grows towards intrinsically motivated and self-determined reading. The teacher is a "significant other" for reading literacy. A host of studies show that teachers who improve students' sense of ownership and competence enable students to become active readers who are high achievers in reading. By contrast, teachers who neglect these instructional practices undermine students' efforts to become autonomous, resulting in students who are disengaged from reading and fail to progress in reading achievement (Guthrie, 2008).
- *Practices enhancing reading skills and metacognitive strategies* are based on direct instruction of reading strategies (Pressley, 2000), and also match with the "cognitive activation" facet of the tridimensional model of quality of teaching (Klieme, Pauli, & Reusser, 2009), namely the model on which the selection of teaching constructs for PISA 2012 and 2015 has been built. According to this model, classroom management processes, teacher support and cognitive activation independently predict mathematics, science and language skills growth, whereas teacher support mainly predicts motivation growth, and cognitive activation predicts cognitive gains at different grades (Klieme, Steinert, & Hochweber, 2010).

However, when students are 15 years old, reading is no longer taught as a subject matter in the same way that mathematics and science are. It is therefore a challenge to tailor questions capturing the classroom practices and opportunity-to-learn students are exposed to. Reading instruction is indeed not only given in test language lessons, but also in other courses (such as social science, natural science, foreign languages, civic education and ICT courses). It is even more of a challenge for the new skills and processes related to digital reading, which might be learned mostly in non-formal contexts, taught in specific courses in some countries, considered as "transversal" in other ones or not taught at all.

There is an ongoing debate in the field of reading research between scholars supporting the view “every teacher is a reading teacher” (“generic” or “intermediate” reading literacy) and advocates of the “content” or “disciplinary” literacy view, who argue that reading texts in mathematics, chemistry, and history require specific reading skills linked to the domain that should be taught by the content matter teachers (Shanahan & Shanahan, 2008). Taking into account the time limitations and the fact that teaching practices were not considered as a priority for PISA 2018, it is proposed to ask the students a limited number of questions about their experience and exposure to “generic” or “intermediate” literacy teaching practices in their classes in general, as it was already done in PISA 2009 (Lafontaine, Baye, Vieluf, & Monseur, 2015).

b. Out-of-school experience (module 5)

In previous PISA cycles (2000 and 2009), reading practices were measured as *self-reported frequencies of reading activities with diverse content in various media*. The initial list included fiction and non-fiction books, comics, newspapers, magazines and e-mails. In 2009, new items about online reading practices were included, some of them focused on social online practices (blogs, forums, emails), others on searching information (reading news online, searching information to learn about a topic, searching practical information). While the online searching information practices explain a significant, yet small proportion of variance of the online digital reading, the social practices online have no clear relationship with digital reading in most of the countries (OECD, 2010, vol. VI; Naumann, 2015)).

For PISA 2018, the list of online reading practices scales will be extended in order to take into account the recent and emerging practices (e.g. e-books, social networks). Despite the growing importance of reading on digital devices, it should be kept in mind that traditional forms of reading (especially reading books) is still the most influential practice for reading comprehension and vocabulary growth, as was shown in a recent longitudinal study by Pfof, Dörfler, and Artelt (2013).

As for the reading cognitive assessment, it is important to keep at least some of the items measuring reading practices the same as in 2009 for trend purposes. For instance, the comparison between 2000 and 2009 of reading practices showed that 15-year-olds read less in 2009 than in 2000, and the decline was on average stronger for males. This is important information. However, this trend purpose has to be carefully balanced with other challenges. Notably, questions kept exactly alike might take a different meaning because reading itself has changed: the term “book” for instance might refer to printed books or to e-books. The amount of reading practices reported by students could drastically change depending on how questions are worded.

Due to the well-known limitations of self-reported measures based on Likert scales (Allen, Cipelewski & Stanovitch, 1992; Pfof, Dörfler, & Artelt, 2013), alternative ways of capturing reading practices (forced-choice, more behavioural or situational questions) will be experimented during the field trial. If some of these latter would prove to be more stable, culturally invariant and better correlates of reading, they could replace or complement some of the trend items, starting a new trend for the reading practices.

c. Teacher qualification and professional development (module 1)

Many studies have demonstrated a clear influence of teacher-related factors on student learning and outcomes. The continued growth of focus on teacher-related policies within the OECD was exemplified by the fourth International Summit on the Teaching Profession in 2014 (ISTP) (cf. Schleicher, 2014). In addition to teachers' professional behaviour within the classroom (see section 2.3.1 above), the composition of the teaching force in terms of age and educational level, their initial education and qualification, their individual beliefs and competencies, as well as professional practices on the school level (– such as collaboration and professional development) – have been core topics in educational policy. Basic information on these topics will be available from the PISA 2018 school and student questionnaires as well as the optional teacher questionnaires.

D. Assessing school policies and governance

This subsection summarises the conceptual foundations for module 3 (School-level learning environment for reading) and modules 12-18.

a. School policies and approaches to educational governance

During the last two decades, research on educational effectiveness has largely been concerned with the impact of school-level factors on students' learning. Studies show that school qualities have effects on student progress, with variation in schools appearing to affect students' behaviour. It has been asserted that the environment at the school level can influence the behaviour of teachers and students and thus – mostly indirectly – their consequent success in teaching and learning. Both “soft” factors, such as school climate and parental involvement, and “hard” factors, such as school management activities and allocation policies, vary and are related to student outcomes within and across countries.

School learning environment for reading (module 3): Conceptually, this module overlaps to a considerable degree with other modules dealing with school-level factors, such as module 11, learning time and curriculum, module 14, school leadership and management, and module 18, assessment, evaluation, and accountability (cf. above). In addition to those, the questionnaire expert group recommends that some questions in the school questionnaire directly focus on the status of reading education in the school and available resources. Accordingly, a question has been developed to investigate the overall value of reading within the school (Is reading education a shared priority for stakeholders?), along with questions on resources particularly available for reading education: size of teaching staff; resources such as libraries and digital learning devices; and cooperation with external partners.

School climate (module 12): School climate encompasses shared norms and values, the quality of relationships and the general atmosphere of a school. An academic focus – a general consensus about the mission of the school and the value of education, shared by school leaders, staff, and parents – impacts the norms in student peer groups and facilitates learning (Opdenakker & Van Damme, 2000; Rumberger & Palardy, 2005). In addition, an orderly learning atmosphere maximises the use of learning time. By contrast, disrespectfulness and an unruly environment are counterproductive for teachers and students alike and distract from the school's actual mission. As in previous PISA assessments, school climate will be assessed in the student questionnaire (e.g., “student-teacher relationship”, “achievement pressure”) and the school questionnaire (e.g., “teacher morale”, “behaviour affecting school climate”).

However, a number of items were added to the teacher-student-relationship scale for 2015 in order to cover different aspects of supportive learning environments that have been identified by Caldwell and Bradley (1984) in their “home inventory”: emotional warmth; guidance and boundaries; stimulation/scaffolding and stability. The complete question should be implemented again in the PISA 2018 field trial if data prove successful from 2015: students report on their relationship with teachers and parents, whilst school principals and parents (in the optional questionnaires) will be asked to answer parallel items. All in all, an interesting picture of social relationships between students, schools and parents could emerge in countries that administer both the parent and the school questionnaires. Finally, two scales should be continued in the student questionnaire that cover more problematic, often hidden aspects of school climate, that nevertheless are highly important from a pedagogical as well as a policy point of view: bullying by peers and unfair treatment by teachers. If the field trial results provide evidence of trustworthy results, the PISA description of school contexts may become much richer in the future.

Parental involvement (module 13): Over the past years, the involvement of parents in educational processes has gained importance in the educational debate and relevance for educational policy. Parents are not only an important audience, but are also powerful stakeholders in education. Thus information on parents’ opinions and engagement is highly valuable for large-scale assessments like PISA. Parental involvement in education has been part of PISA since 2006 when the parent questionnaire was administered for the first time, directly addressing the parents of the PISA students. For PISA 2015, specific aspects of parental involvement were added to all questionnaires, focusing on (a) parent-school communication and collaboration and (b) specific aspects of parental support for learning; these scales should be continued for 2018. Depending on the addressee of the questionnaires (i.e. teachers, students, school principals, or parents), specific aspects of parental engagement are highlighted. Aspects of parental involvement can also be found in scales of other modules. Nevertheless, the majority of items and topics regarding parental involvement are included in the parent questionnaire that will be administered as an international option.

Specific questions related to reading literacy will be addressed in the parent questionnaire, especially the parental support to the development of their child’s early literacy skills (at International Standard Classification of Education (ISCED 0), and parents’ interest and motivation for reading. There is extensive evidence that parental support before formal instruction of reading, like joint reading of books, or playing language games, is critical to the development of early or “emergent” literacy (for a synthesis, see Neuman & Dickinson, 2002). Regarding parental support for adolescents’ reading motivation and practices, Klauda (2009), in an extensive review, underlined that “there is clearly still much to be learned in this area, and previous research has certain limitations” (Klauda, 2009, p. 352).

School context and resources (module 14): Information on school type (public vs. private) and class size has always been included in the school questionnaire. In addition to these trend questions, PISA 2015 field trial expanded the module which allows discrimination between types of private schools (religious/denominational, not-for-profit, for-profit) and provides more advanced information on ICT use. All PISA cycles so far have included a question on the degree of problems a school experiences due to missing resources. The different approaches over time were systematised and implemented in one coherent question in the school questionnaire.

Allocation, selection, choice and grade repetition (module 15): The way students are channelled into educational pathways, schools, tracks or courses is a core issue of educational governance (“stratification”, “streaming” or “tracking”). On the school level, selection and allocation procedures are important aspects of school organisation. Highly selective schools

provide a learning environment that may differ from the environment offered by more comprehensive schools. For all of these reasons, appropriate trend questions answered by school administrators and parents have been retained.

As a new topic, PISA 2015 added information on grade repetition (percentage of repeaters was asked at school level), which is a very important aspect of vertical differentiation, as shown in the PISA 2009 report. Many longitudinal studies have demonstrated grade retention to have a negative impact on individual careers and outcomes (e.g. Griffith et al. 2010; Ou & Reynolds, 2010), student behaviour and well-being (e.g. Crothers et al., 2010). Grade repetition is less common in secondary schools compared to primary schools, but late retention seems to have larger (negative) effects (Ou & Reynolds, 2010). Greene and Winter (2009) showed that once a test-based retention policy had been installed, those who are exempted from the policy did worse. Babcock and Bedard (2011) even showed that a large number of students being retained could have a positive (!) effect on the cohort (i.e. all students, including those who are promoted). Kloosterman and De Graaf (2010) argued that in highly tracked systems, such as in some European countries, grade repetition might serve as a preferred alternative to moving into a lower track; indeed, they found evidence that this strategy is preferred for students with higher SES. Thus, changing grade repetition policies might be a viable option regarding low-cost interventions (Binder, 2009). Therefore, it is worthwhile to take a closer, comparative look at grade retention policies and their (differential) effects on students who do and do not repeat a grade. PISA 2018 will explore one related question in the school.

b. Assessment, evaluation, and accountability (module 16)

Assessing students and evaluating schools is a common practice in most countries (Ozga, 2012). Since the 1980s, policy instruments, such as performance standards, standard-based assessment, annual reports on student progress and school inspectorates, have been promoted and implemented across continents. Reporting and sharing data from assessments and evaluations with different stakeholders provides multiple opportunities for monitoring, feedback and improvement. In recent years, there has been a growing interest in the use of assessment and evaluation results through feedback to students, parents, teachers and schools as one of the most powerful tools for quality management and improvement (OECD, 2010, p. 76); formative assessment also known as assessment for learning has been one of the dominant movements (Baird et al., 2014; Black, 2015; Hattie, 2009,). Accountability systems based on these instruments are increasingly common in OECD countries (Rosenkvist, 2010, Scheerens, 2002, p. 36).

Previous PISA cycles have covered aspects of assessment, evaluation and accountability in the school questionnaire by identifying a variety of purposes for the assessment of students. School leaders have been asked whether they use test results to make comparisons with other schools at the district or national level, as well as to improve teacher instructions, by asking students for written feedback on lessons, teachers or resources. In the following, relevant research on school evaluation and student assessment is summarised to provide the rationale for questionnaire development in PISA 2018.

Evaluation: The evaluation of schools is used as a means of assuring transparency, deciding and making judgements about systems, programmes, educational resources and processes and also to guide school development (Faubert, 2009). In PISA 2018 the term evaluation will be used for processes on school and system level, as was done in PISA 2015, to make comparison possible. Evaluation criteria need to be defined and applied from the viewpoints of different stakeholders (Sanders & Davidson, 2003).

Evaluation can be either external or internal (Berkenmeyer & Müller, 2010). It is called external evaluation if the process is controlled and headed by an external body and the school does not define the areas that are judged. An evaluation is called internal if it is part of a process controlled by the school and in which the school defines the areas that are judged; the evaluation may be conducted by members of the school (self-evaluation) or by persons/institutions commissioned by the school. Different evaluation practices generally coexist and benefit from each other (Ryan, Chandler, & Samuels, 2007). External evaluation can expand the scope of internal evaluation and also validate results and implement standards or goals. Internal evaluation can improve the interpretation and increase the utilisation of external evaluation results. However, improvement of schools seems to be more likely when an internal evaluation is applied, compared to external evaluation. Therefore, processes and outcomes of evaluation might differ between internal and external evaluation. Moreover, country and school-specific context factors may influence the implementation of evaluations as well as conclusions and effects for schools. In many countries, individual evaluation of teachers and principals, separate from school-wide evaluation, is also common (Faubert, 2009; Santiago & Benavides, 2009); they will be treated here as separate types of evaluation.

Assessment: The past decade a number of countries have implemented national standards to assess students learning outcome. Together with formative assessment practices, the summative assessment systems influence the way teachers teach and students learn. In particular, formative assessment practices can enhance students' achievement (Black and Wiliam 1998). However, there is a large variation in the implementation of formative assessment practices, which has also been reported in recent studies in the US, Canada, Sweden, Scotland, Singapore and Norway among others (Wylie & Lyon, 2015, DeLuca et al. 2015, Jonsson et al 2015, Hayward, 2015, Ratnman- Lim & Tan, 2015, Hopfenbeck et al 2015). PISA 2018 therefore aims to assess the formative and summative aspects of students learning by including questions in both the student and school leader questionnaire.

The 2015 field trial suggested that the administration time of the questionnaire was very long, requiring a reduction of the constructs covered in the questionnaire eventually administered in PISA 2015. A final choice of constructs in 2018 can be based on experiences that will be gained in the 2015 data administration, possibly complemented by the 2018 field trial.

E. Dealing with response bias

The analysis of bias has played an important role in the analysis of PISA data. Two types of approaches have been implemented: adapted instruments (using novel constructs or response formats) and advanced statistical modelling. The field trial will enable some experimenting with different question formats – e.g. using interactive features of the computer-based administration system – and with new content – e.g. measures of students' physical and emotional well-being. The field trial will assess psychometric properties, but will also check whether new measures are accepted by stakeholders in different cultures. Based on the field trial findings, a subset of measures will be selected for the main study.

Well-known examples of design measures are the use of overclaiming, anchoring vignettes and cognitive interviews to examine response styles. Overclaiming is a procedure in which the student is asked about his or her knowledge about a number of mathematical concepts (the procedure has been applied as part of the field trial of PISA 2012 in which mathematics was assessed), some of which are non-existing. Students indicating to have a high knowledge of these non-existing concepts are supposed to show strong response styles, which is somewhat akin to social desirability. Statistical correction for overclaiming had an impact on cross-national

differences in scores. As a test of the adequacy of bias-correction procedures, correlations between motivation and performance are often computed at individual and country level. These correlations are usually positive and significant, yet small at individual level, but strong and negative at country level (Marsh & Hau, 2004). Cross-national differences in response styles have been argued to be responsible for this change of correlations after aggregation. Correction for overclaiming led to a sizable reduction of the negative correlation between motivation and achievement at country level (Kyllonen & Bertling, 2013).

Anchoring vignettes follow a different procedure (King, Murray, Salomon, & Tandon, 2004). Before assessing a target construct, such as reading motivation, descriptions of hypothetical persons are presented, usually with very high, medium and very low levels of motivation. The students are asked how he or she would rate the motivation of these hypothetical persons. The response on the target item, how the students would rate their own reading motivation, is then re-scaled using the anchor scores of the three hypothetical persons. The results of the use of anchoring vignettes have been mixed. Very promising results were obtained in the 2012 field trial where the country-level correlations between motivation and achievement were rather close to the individual-level correlations. However, the 2015 field trial yielded a more complex and less supportive set of results with regard to the use of anchoring vignettes. A problem of the use of anchoring vignettes is the additional testing time they require and the relatively high reading load of the items.

Various statistical procedures to correct for cross-national bias have been proposed. An example is the correction for response styles, such as acquiescence, extremity and social desirability, in analyses of covariance, using response style indexes as covariates (He & Van de Vijver, in press; Van de Vijver & He, 2014). Such procedures do not tend to have much impact on the negative correlation between motivation and achievement at country level. Another example of such a procedure is the use of propensity score matching (Buckley, 2009). The idea behind this approach is to increase the comparability of samples obtained in different countries by matching the samples on relevant background characteristics, such as socioeconomic status. Although still frequently used, there are indications that the negative correlation between motivation and achievement at country level is not strongly affected by propensity matching procedures. A final example is the statistical modelling of a response style factor in a confirmatory factor analysis (Billiet & McClendon, 2000). Again, not much experience has been gained using this procedure in a PISA context, but the applicability of the procedure may be limited given that it can only be used in balanced scales (in which items are formulated in a positive direction, and other items are formulated in the opposite direction, such as in measures of extroversions where some items assess extroversion and some introversion). PISA background scales do not use such a balanced approach.

It can be concluded that there is no simple way that would allow us to eliminate cross-cultural bias in the background questionnaire of PISA 2018. Building on previous experience, we propose to combine different approaches. Therefore, it would be advisable to use cognitive interviewing to avoid response styles and other problematic aspects of items as much as possible, and to use data of the field trial to further investigate the feasibility of different statistical approaches (including the standard tests of invariance using confirmatory factor analysis). However, it is important to recognise that the questionnaire design and statistical analyses proposed may not be sufficient to eradicate all bias sources in the background questionnaire of PISA 2018.

In summary, it is proposed to use the field trial to test the psychometric properties of the new scales and question formats. Given the inconclusiveness about how to deal with cross-cultural bias and response styles, we propose not to use the field trial for further development of these issues, but to employ procedures that were used before (e.g. confirmatory factor analysis and IRT modelling of response styles).

REFERENCES

- Afflerbach, P., & Cho, B.-Y. (2010). Determining and describing reading strategies. Internet and traditional forms of reading. In H. S. Waters & W. Schneider (Eds.), *Metacognition, Use and Instruction* (pp. 201-225). New York: Guilford Press.
- Almlund, M., Duckworth, A. L., Heckman, J. J., & Kautz, T. (2011). *Personality psychology and economics* (IZA Discussion Papers 5500). Bonn: Institute for the Study of Labour (IZA).
- Allen, L., Cipielewski, J. & Stanovitch, K. E. (1992). Multiple indicators of children's reading habits and attitudes: Construct validity and cognitive correlates. *Journal of Educational Psychology, 84*, 489-503.
- Anderson, R. (2008). Implications of the information and knowledge society for education. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 5-22). New York: Springer.
- Aronson, J., Fried, C. B., & Good, C. (2002). Reducing the effects of stereotype threat on African American college students by shaping theories of intelligence. *Journal of Experimental Social Psychology, 38*, 113-125.
- Artelt, C., Schiefele, U., & Schneider, W. (2001). Predictors of reading literacy. *European Journal of Psychology of Education, 16*, 363-383.
- Artelt, C. & Schneider, W. (2015 online). Cross-country generalizability of the role of metacognitive knowledge for students' strategy use and reading competence. *Teachers College Record, 117* (1), 010304, 1-32.
- Assor, A., Kaplan, A., & Roth, G. (2002). Choice is good, but relevance is excellent: Autonomy-enhancing and suppressing teacher behaviours predicting students' engagement in schoolwork. *British Journal of Educational Psychology, 72*, 261-278.
- Babcock, P., & Bedard, K. (2011). The wages of failure: New evidence on school retention and long-run outcomes. *Education Finance and Policy, 6*, 293-322.
- Baker, L., & Carter-Beall, L. (2009). Metacognitive processes and reading comprehension. In S.E. Israel & G.G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 272-288). New York: Routledge.
- Baker, L., & Wigfield, A. (1999). Dimensions of children's motivation for reading and their relations to reading activity and reading achievement. *Reading Research Quarterly, 34*, 452-477.
- Baird, J., Hopfenbeck, T. N., Newton, P., Stobart, G., & Steen-Utheim, A. T. (2014). *State of the Field Review: Assessment and Learning*. Report for the Norwegian Knowledge Centre for Education.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: W.H. Freeman.
- Baranik, L. E., Barron, K. E., & Finney, S. J. (2007). Measuring goal orientation in a work domain: Construct validity evidence for the 2 x 2 framework. *Educational and Psychological Measurement, 67*, 697-718.
- Barile, J. P., Donohue, D. K., Anthony, E. R., Baker, A. M., Weaver, S. C., Henrich, C. (2012). Teacher-student relationship climate and school outcomes: Implications for educational policy initiatives. *Journal of Youth and Adolescence, 41*, 256-267.
- Becker, M., McElvany, N., & Kortenbruck, M. (2010). Intrinsic and extrinsic reading motivation as predictors of reading literacy: a longitudinal study. *Journal of Educational psychology, 102*, 773-785.
- Bennett, M. (1993). Towards ethno-relativism: A developmental model of intercultural sensitivity. In M. Paige (Ed.), *Education for the intercultural experience* (pp. 21-71). Yarmouth, ME: Intercultural Press.
- Berkenmeyer, N., & Müller, S. (2010). Schulinterne Evaluation - nur ein Instrument zur Selbststeuerung von Schulen? In H. Altrichter & K. Maag Merki (Eds.), *Handbuch neue Steuerung im Schulsystem* (pp. 195-218). Wiesbaden: VS Verlag für Sozialwissenschaften.

- Billiet, J. B., & McClendon, M. J. (2000). Modeling acquiescence in measurement models for two balanced sets of items. *Structural Equation Modeling*, 7, 608-628.
- Binder, M. (2009). Why are some low-income countries better at providing secondary education? *Comparative Education Review*, 53, 513–534.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., & Rumble, M. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, & E. Care (Eds.) *Assessment and teaching of 21st century skills* (pp. 17-66). Dordrecht, Heidelberg, London, New York: Springer.
- Birney, R., Burdick, H., & Teevan, R. (1969). *Fear of failure*. New York: Van Norstrand-Reinhold Co.
- Black, P. (2015). Formative assessment – an optimistic but incomplete vision, *Assessment in Education: Principles, Policy and Practice*, 22: 1, 161 – 177.
- Blackwell, L. S., Trzesniewski, K. H., & Dweck, C. S. (2007). Implicit theories of intelligence predict achievement across an adolescent transition: A longitudinal study and an intervention. *Child Development*, 78, 246-263.
- Blau D., & Curie, J. (2006). Pre-school, day care, and after-school care: Who's minding the kids? In *Handbook of the economics of education* (pp. 1164-1278). Amsterdam: Elsevier.
- Bonita, R., De Counter, M., Dwyer, T., Jamrozi, K., & Winkelmann, R. (2001). *Surveillance of risk factors for noncommunicable disease: The WHO STEPwise approach*, World Health Organization, Geneva.
- Bourdieu, P. (1986). The forms of capital. In J. Richardson (Ed.), *Handbook of theory and research for the sociology of education* (pp. 241-258). New York: Greenwood.
- Brown, A. L., Palincsar, A. S., & Armbruster, B. B. (2004). Instructing comprehension-fostering activities in interactive learning situations. In R. B. Ruddell & N. J. Unrau (Eds.), *Theoretical Models and Processes of Reading International Reading Association* (fifth ed., pp. 780-809). Newark.
- Bryk, A. S., & Driscoll, M. E. (1988). *The high school as community: Contextual influences and consequences for students and teachers*. Madison: National Center on Effective Secondary Schools, University of Wisconsin.
- Bryk, A. S., Sebring, P. B., Allensworth, E., Luppescu, S., & Easton, J. Q. (2010). *Organizing schools for improvement: Lessons from Chicago*. Chicago: University of Chicago Press.
- Buckley, J. (2009). *Cross-national response styles in international educational assessments: Evidence from PISA 2006*. Retrieved from https://edsurveys.rti.org/PISA/documents/Buckley_PISAresponsestyle.pdf
- Caldwell, B. M., & Bradley, R. H. (1984). *Home observation for measurement of the Environment: Administration manual, Revised Edition*. Unpublished manuscript. Little Rock: University of Arkansas.
- Chapman, C. P., Armstrong, P., Harris, A., Muijs, D. R., Reynolds, D., & Sammons, P. (Eds.) (2011). *School effectiveness and improvement research, policy, and practice: Challenging orthodoxy*. Abingdon: Routledge.
- Chapman, J. W., & Tunmer, W. E. (1995). Development of young children's reading self-concepts: An examination of emerging subcomponents and their relationship with reading achievement. *Journal of Educational Psychology*, 87, 154–167.
- Coiro, J., & Dobler, E. (2007). Exploring the online reading comprehension strategies used by sixth-grade skilled readers to search for and locate information on the Internet. *Reading Research Quarterly*, 42 (2)234-257.
- Covington, M. V. (1992). *Making the grade: A self-worth perspective on motivation and school reform*. New York: Cambridge University Press.
- Creemers, B. P. M., & Kyriakides, L. (2008). *The dynamics of educational effectiveness: A contribution to policy, practice and theory in contemporary schools*. London: Routledge.
- Crothers, L. M., Schreiber, J. B., Schmitt, A. J., Bell, G. B., Blasik, J., Comstock, L. A., Greisler, M. J., Keener, D., King, J. M., & Lipinski, J. (2010). A preliminary study of bully and victim behavior in old-for-grade students:

- Another potential hidden cost of grade retention or delayed school entry. *Journal of Applied School Psychology*, 26, 327-338.
- Cunha, F., Heckman, J. J., Lochner, L. J., & Masterov, D. V. (2006). Interpreting the evidence on life cycle skill formation. In E. Hanushek & F. Welch (Eds.), *Handbook of the economics of education* (pp. 697-812). Amsterdam: Elsevier.
- Deardorff, D.K. (2009). *The SAGE handbook of intercultural competence*. Thousand Oaks: Sage.
- Deci, E. L., & Ryan, R. M. (1985). *Intrinsic motivation and self-determination in human behavior*. New York: Plenum Press.
- DeLuca, C. Klinger, D., Pyper, J., & Woods, J. (2015). Instructional rounds as a professional learning model for systemic implementation of assessment for learning. *Assessment in Education: Principles, Policy and Practice*, 22:1, 122-139.
- Dole, J. A., Nokes, J. D., & Dritis, D. (2009). Cognitive strategy instruction. In S. E. Israel, & G. G. Duffy (Eds.), *Handbook of research on reading comprehension* (pp. 347-372). New York: Routledge.
- Duckworth, A. L., Peterson, C., Matthews, M. D., & Kelly, D. R. (2007). Grit: Perseverance and passion for long-term goals. *Journal of Personality and Social Psychology*, 92, 1087-1101.
- Dweck, C. S. (1999). *Self theories: Their role in motivation, personality, and development*. Philadelphia: Taylor and Francis.
- Dweck, C. S. (2007). *Mindset: The new psychology of success*. New York: Ballantine Books.
- Elliot, A. J. (2005). A conceptual history of the achievement goal construct. In A. Elliot & C. Dweck (Eds.), *Handbook of competence and motivation* (pp. 52-72). New York: Guilford Press.
- Elliot, A. J., & McGregor, H. A. (2001). A 2 × 2 achievement goal framework. *Journal of Personality and Social Psychology*, 80, 501-519.
- Elliot, A. J., & Thrash, T. M. (2001). Achievement goals and the hierarchical model of achievement motivation. *Educational Psychology Review*, 12, 139-156.
- Faubert, V. (2009). *School evaluation: Current practices in OECD countries and a literature review*. Paris: OECD Education Working Papers.
- Ferla, J., Valcke, M., & Cai, Y. (2009). Academic self-efficacy and academic self-concept: reconsidering structural relationships. *Learning and Individual Differences*, 19(4), 499–505.
- Fisch, K., & McLeod, S. (2009). *Did you know? 3.0 – 2009 Edition*. Retrieved 15 May 2009 from http://www.youtube.com/watch?v=PHmwZ96_Gos.
- Fisher, C. B., Wallace, S. A., & Fenton, R. E. (2000). Discrimination distress during adolescence. *Journal of Youth and Adolescence*, 29, 679-695.
- Fivush, R., & Hamond, N. R. (1990). *Knowing and remembering in young children*. New York: Cambridge University Press.
- Goodman, R. (1997). The strengths and difficulties questionnaire: A research note. *Journal of Child Psychology and Psychiatry*, 38, 581-586.
- Greene, J. P., & Winters, M. A. (2009). The effects of exemptions to Florida's test-based promotion policy. Who is retained? Who benefits academically? *Economics of Education Review*, 28:1, 135-142.
- Gustafsson, J. E. (2008). Effects of international comparative studies on educational quality on the quality of educational research. *European Educational Research Journal*, 7, 1-17.
- Guthrie, J. T. (2008). Reading motivation and engagement in middle and high school: Appraisal and intervention. In J. T. Guthrie (Ed.), *Engaging adolescents in reading* (pp. 1-16). Thousand Oaks: Corwin.

- Guthrie, J. T., Ho, A. N., & Klauda, S. L. (2013). Modeling the relationships among reading instruction, motivation, engagement, and achievement for adolescents. *Reading Research Quarterly*, 48, 9-26.
- Guthrie, J. T., & Wigfield, A. (2000). Engagement and Motivation in Reading. In M. L. Kamil & P. B. Mosenthal (Eds.), *Handbook of reading research* (Vol. 3, pp. 403-422). Mahwah, NJ: Erlbaum.
- Guthrie, J. T., Wigfield, A., & You, W. (2012). Instructional contexts for engagement and achievement in reading. In S.L. Christensen, A.L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 601-634). New York: Springer Science.
- Gutman, L. M., & Schoon, I (2013). *The impact of non-cognitive skills on outcomes for young people. Literature review*. London: Institute of Education, University of London.
- Johnon, A., Lundahl, C., & A. Holmgren (2015). Evaluating a large-scale implantation of Assessment for Learning in Sweden. *Assessment in Education: Principles, Policy and Practice*, 22:1, 104-121
- Hanushek, E. A., & Woessmann, L. (2010). *The high cost of low educational performance: The long-run economic impact of improving PISA outcomes*. Paris: OECD.
- Hanushek, E. A., & Woessmann, L. (2011). The economics of international differences in educational achievement. In E. A. Hanushek, S. Machin, & L. Woessmann (Eds.), *Handbook of the economics of education* (Vol. 3, pp. 89-200). Amsterdam: Elsevier.
- Hattie, J. (2009). *Visible learning. A synthesis of over 800 meta-analyses relating to achievement*. London & New York: Routledge.
- Hayward, L. (2015). Assessment is learning: The preposition vanishes. *Assessment in Education: Principles, Policy and Practice*, 22:1, 27-43.
- Hopfenbeck, T., Florez Petour, M. T., & Tolo, A. (2015). Balancing tensions in educational policy reforms: large –scale implementation of Assessment for Learning in Norway. *Assessment in Education: Principles, Policy and Practice*, 22:1, 44-60.
- He, J., & Van de Vijver, F. J. R. (in press). *The motivation-achievement paradox in international educational achievement tests: Towards a better understanding*. Festschrift for David Watkins.
- Heckhausen, H. (1975). Fear of failure as a self-reinforcing system. In I. Sarason & C. Spielberger (Eds.), *Stress and anxiety* (Vol. 2, pp. 117-128). Washington, DC: Hemisphere.
- Heckman, J. J. (2006). Skill formation and the economics of investing in disadvantaged children. *Science*, 312, 1900-1902.
- Heckman J. J., Stixrud, J., & Urzua, S. (2006). The effects of cognitive and noncognitive abilities on labor market outcomes and social behavior. *Journal of Labor Economics*, 24, 411-482.
- Helmreich, R. L., Beane, W., Lucker, G. W., & Spence, J. T. (1978). Achievement motivation and scientific attainment. *Personality and Social Psychology Bulletin*, 4, 222-226.
- Hembree, R. (1988). Correlates, causes, effects, and treatment of test anxiety. *Review of Educational Research*, 58, 47-77.
- Hopfenbeck, T. N., & Stobart, G. (2015). Large-scale implementation of Assessment for Learning. *Assessment in Education: Principles, Policy & Practice*, 22, 1-2.
- Hulleman, C. S., Schrager, S. M., Bodman, S. M., & Harackiewicz, J. M. (2010). A meta-analytic review of achievement goal measures: Different labels for the same constructs or different constructs with similar labels? *Psychological Bulletin*, 136, 422-449.
- Kaplan, A., & Maehr, M. L. (2007). The contributions and prospects of goal orientation theory. *Educational Psychology Review*, 19, 141-184.

- Kaye, M. P., Conroy, D. E., & Fifer, A. (2008). Individual differences in incompetence avoidance: A comparison of multiple dimensions of perfectionism and fear of failure. *Journal of Sport and Exercise Psychology*, 30, 110-132.
- King, G., Murray, C. J., Salomon, J. A., & Tandon, A. (2004). Enhancing the validity and cross-cultural comparability of measurement in survey research. *American Political Science Review*, 98, 191-207.
- Kirsch, I., De Jong, J., Lafontaine, D., McQueen, J., Mendelovits, J., & Monseur, C. (2002). *Reading for a change. Performance and engagement across countries. Results from PISA 2000*. Paris: OECD.
- Klauda, S. L. (2009). The role of parents in adolescents' reading motivation and activity. *Educational Psychology Review*, 21, 325-363.
- Klauda, S. L., & Guthrie, J. T. (2014). Comparing relations of motivation, engagement, and achievement among struggling and advanced adolescent readers. *Reading and Writing*, 28, 239-269.
- Klieme, E., Backhoff, E., Blum, W., Buckley, J., Hong, Y., Kaplan, D., Levin, H., Scheerens, J., Schmidt, W., Van de Vijver, F. J. R., & Vieluf, S. (2013). PISA 2012 context questionnaires framework. In Organisation for Economic Co-operation and Development (Eds.), *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy* (pp. 167-258). Paris: OECD Publishing.
- Klieme, E., & Kuger, S. (Eds.) (2014). *PISA 2015 Draft Questionnaire Framework*. Paris: OECD. Retrieved from: <http://www.oecd.org/pisa/pisaproducts/PISA-2015-draft-questionnaire-framework.pdf>.
- Klieme, E., Pauli, C., & Reusser, K. (2009). The Pythagoras study - Investigating effects of teaching and learning in Swiss and German mathematics classrooms. In T. Janik & T. Seidel (Eds.), *The power of video studies in investigating teaching and learning in the classroom* (pp. 137-160). Münster: Waxmann
- Klieme, E., Steinert, B., & Hochweber, J. (2010). Zur Bedeutung der Schulqualität für Unterricht und Lernergebnisse. In W. Bos, E. Klieme, & O. Köller (Eds.), *Schulische Lerngelegenheiten und Kompetenzentwicklung. Festschrift für Jürgen Baumert* (pp. 231-256). Münster: Waxmann.
- Kloosterman, R., & De Graaf, P. M. (2010). Non-promotion or enrolment in a lower track? The influence of social background on choices in secondary education for three cohorts of Dutch pupils. *Oxford Review of Education*, 36, 363-384.
- Kristjansson, B., Petticrew, M., MacDonald, B., Krasevec, J., Janzen, L., Greenhalgh, T., et al. (2009). School feeding for improving the physical and psychosocial health of disadvantaged students. *Cochrane Collaboration*, 2009(3). Retrieved from http://r4d.dfid.gov.uk/PDF/Articles/SR_SchoolFeeding.pdf.
- Kyllonen, P., & Bertling, J. (2013). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, & D. Rutkowski, (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*. Boca Raton, FL: CRC Press.
- Lafontaine, D., Baye, A., Vieluf, S., & Monseur, C. (2015). Equity in opportunity-to-learn and achievement in reading: A secondary analysis of PISA 2009 data. *Studies in Educational Evaluation*, 47, 1-11.
- Lafontaine, D., Dupont V., Hindryckx, G., & Schillings P. (2015). Pratiques d'enseignement et compétences en lecture des élèves : Qu'évaluent les enquêtes internationales et que peuvent en apprendre les enseignants? In M. Bianco et L. Lima (eds). *Enseigner à comprendre les textes*, Editions Haier.
- Lindqvist, E., & Vestman, R. (2011). The labor market returns to cognitive and noncognitive ability: Evidence from the Swedish enlistment. *American Economic Journal: Applied Economics*, 3, 101-128.
- Logan, J., Piasta, S., Justice, L., Schatschneider, C., & Petrill, S. (2011). Children's attendance rates and quality of teacher-child interactions in at-risk preschool classrooms: Contribution to children's expressive language growth. *Child and Youth Care Forum*, 40, 457-477.
- Maehr, M. L., & Nicholls, J. G. (1980). Culture and achievement motivation: A second look. In N. Warren (Ed.), *Studies in cross cultural psychology* (Vol. 3, pp. 221-267). New York: Academic Press.

- Markowitsch, H. J., & Welzer, H. (2009). *The development of autobiographical memory*. London: Psychology Press.
- Marsh, H. W., & Craven, R. (1997). Academic self-concept: Beyond the dustbowl. In G. Phye (Ed.), *Handbook of classroom assessment: Learning, achievement, and adjustment* (pp. 131-198). Orlando, FL: 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*, 133-163.
- Marsh, H. W., Hau, K.-T., Artelt, C., Baumert, J., & Peschar, J. L. (2006). OECD's brief self-report measure of educational psychology's most useful affective constructs: cross-cultural, psychometric comparisons across 25 countries. *International Journal of Testing, 6*, 311–360.
- Marsh, H. W., & Hau, K.T. (2004). Explaining paradoxical relations between academic self-concepts and achievements: Cross-cultural generalizability of the internal/external frame of reference predictions across 26 countries. *Journal of Educational Psychology, 96*, 56-67.
- McKenna, M. C., & Robinson, R. D. (1990). Content literacy: A definition and implications. *Journal of Reading, 34*, 184-186.
- Middleton, M., & Perks, K. (2014). *Motivation to learn*. Thousand Oaks: Sage Publishing.
- Mol, S. E., & Bus, A. G. (2011). To read or not to read: A meta-analysis of print exposure from infancy to early adulthood. *Psychological Bulletin, 137*, 267–296.
- Morgan, P., & Fuchs, D. (2007). Is there a bidirectional relationship between children's reading skills and reading motivation? *Exceptional Children, 73*, 165-183.
- Murayama, K., & Elliot, A. J. (2012). The competition-performance relation: A meta-analysis and a test of the opposing processes model of competition and performance. *Psychological Bulletin, 138*, 1035-1070.
- National Reading Panel. (2000). *Report of the National Reading Panel: Teaching children to read*. Washington, DC: US Government Printing Office.
- Naumann, J. (2015). A model of online reading engagement: linking engagement, navigation and performance in digital reading. *Computers in Human Behavior, 53*, 263-277.
- Neuman, S. B., & Dickinson, D. K. (Eds) (2002). *Handbook of early literacy research*. New York: Guilford Press.
- OECD. (2004). *Learning for tomorrow's world: First results from PISA 2003*. Paris: OECD Publishing.
- OECD. (2005). *PISA 2003 technical report*. Paris: OECD.
- OECD. (2005a). *Formative assessment: Improving learning in secondary education* (p. 21). Paris: OECD.
- OECD. (2005b). *The definition and selection of key competencies: Executive summary*. Retrieved from: <http://www.oecd.org/dataoecd/47/61/35070367.pdf>.
- OECD. (2006). *PISA – The OECD Programme for International Student Assessment*. Paris: OECD.
- OECD (2012). *PISA 2009 Technical Report*. Paris: OECD.
- OECD. (2010). *PISA 2009 results: What students can do? Student performance in reading, mathematics and science, Vol. 1*. Paris: OECD Publication.
- OECD. (2011a). *PISA 2009 Results: Students On Line: Digital Technologies and Performance, Vol. VI*. Paris: OECD Publication.
- OECD (2011b) *How's life? Measuring Well-Being*. Paris: OECD Publication. Retrieved from <http://dx.doi.org/10.1787/9789264121164-en>.

- OECD. (2013a). *PISA 2012 assessment and analytical framework: Mathematics, reading, science, problem solving and financial literacy*. Paris: OECD Publishing. Retrieved from: <http://www.oecd.org/pisa/pisaproducts/PISA-2015-draft-questionnaire-framework.pdf>.
- OECD. (2013b). Guidelines on measuring subjective well-being. Paris: OECD Publishing. Retrieved from <http://dx.doi.org/10.1787/9789264191655-en>.
- OECD. (2014a). *PISA 2015 draft questionnaire framework*. Paris: OECD Publishing.
- OECD (2014b). *PISA 2012 Technical Report*. Paris: OECD.
- OECD. (2015). *Education Policy Outlook 2015: Making Reforms Happen*. Paris: OECD Publishing.
- Opdenakker, M. C., & Van Damme, J. (2000). Effects of schools, teaching staff and classes on achievement and well-being in secondary education: Similarities and differences between school outcomes. *School Effectiveness and School Improvement*, 11, 165-196.
- Ou, S.-R., & Reynolds, A. J. (2010). Grade retention, postsecondary education, and public aid receipt. *Educational Evaluation and Policy Analysis*, 32, 118-139.
- Ozga, J. (2012). Introduction. Assessing PISA, *European Educational Research Journal*, 11(2), 166-171.
- Partnership for 21st Century Skills. (2008). *21st century skills, education & competitiveness: A resource and policy guide*. Tucson, AZ: Partnership for 21st Century Skills. Retrieved from: http://www.p21.org/storage/documents/21st_century_skills_education_and_competitiveness_guide.pdf
- Pfost, M., Dörfler, T., & Artelt, C. (2013). Students' extracurricular reading behavior and the development of vocabulary and reading comprehension. *Learning and Individual Differences*, 26, 89-102.
- Poropat, A. E. (2009). A meta-analysis of the five factor model of personality and academic performance. *Psychological Bulletin*, 135, 322-338.
- Pressley, M., Johnson, C. J., Symons, S., McGoldrick, J. A., & Kurita, J. A. (1989). Strategies that improve children's memory and comprehension of text. *Elementary School Journal*, 90(1), 3-32.
- Pressley, M. (2000). What should comprehension instruction be the instruction of? In M. L. Kamil, P. Mosenthal, P. D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 545–561). Mahwah: Erlbaum.
- Pressley, M., Graham, S., & Harris, K. (2006). The state of educational intervention research as viewed through the lens of literacy intervention. *British Journal of Educational Psychology*, 76, 1-19.
- Purves, A. C. (1987). I.E.A. an agenda for the future. *International Review of Education*, 33, 103-107.
- Ratnman-Lim, C. T. L & Tan, L.H. Kiat (2015). Large-scale implementation of formative assessment practices in an examination oriented culture, *Assessment in Education: Principles, Policy and Practice*, 22:1, 61 – 78..
- Rawsthorne, L. J., & Elliot, A. J. (1999). Achievement goals and intrinsic motivation: A meta-analytic review. *Personality and Social Psychology Review*, 3, 326-344.
- Reeve, J. (2012). A self-determination theory perspective on student engagement. In S. L. Christenson, A. L. Reschly, & C. Wylie (Eds.), *Handbook of research on student engagement* (pp. 149–172). New York, NY: Springer.
- Retelsdorf, J., Köller, O., & Möller, J. (2011). On the effects of motivation on reading performance growth in secondary school. *Learning and Instruction*, 21, 550-559.
- Richardson, M., Abraham, C., & Bond, R. (2012). Psychological correlates of university students' academic performance: A systematic review and meta-analysis. *Psychological Bulletin*, 138, 353-387.
- Roberts, B. W., Kuncel, N. R., Shiner, R., Caspi, A., & Goldberg, L. R. (2007). The power of personality: The comparative validity of personality traits, socioeconomic status, and cognitive ability for predicting important life outcomes. *Perspectives on Psychological Science*, 2, 313-345.

- Rosenkvist, M.A. (2010). "Using Student Test Results for Accountability and Improvement: A Literature Review". OECD Education Working Papers, No. 54, OECD Publishing: Paris.
- Rosenshine, B., & Meister, C. (1994). Reciprocal teaching: A review of the research. *Review of Educational Research*, 64(4), 479-530.
- Rosenshine, B., & Meister, C. (1997). Cognitive strategy instruction in reading. In A. Stahl & A. Hayes (Eds.), *Instructional models in reading* (pp. 85-107). Mahwah, NJ: Erlbaum.
- Rosenshine, B., Meister, C., & Chapman, S. (1996). Teaching students to generate questions: A review of the intervention studies. *Review of Educational Research*, 66(2), 181-221.
- Rumberger, R. W., & Palardy, G. J. (2005). Test scores, dropout rates, and transfer rates as alternative indicators of high school performance. *American Educational Research Journal*, 42, 3-42.
- Ryan, K. E., Chandler, M., & Samuels, M. (2007). What should school-based evaluation look like? *Studies in Educational Evaluation*, 33, 197-212.
- Rychen, D. S., & Salganik, L. (2003). *Key competencies for a successful life and a well-functioning society*. Goettingen: Hogrefe & Huber.
- Salmon, J., Ball, K., Crawford, D., Booth, M., Telford, A., & Hume, C. (2005). Reducing sedentary behaviour and increasing physical activity among 10-year-old children: Overview and process evaluation of the "Switch-Play" intervention. *Health Promotion International*, 20, 7-17.
- Sammons, P., Anders, Y., Sylva, K., Melhuish, E., Siraj-Blatchford, I., Taggart, B., & Barreau, S. (2008). Children's cognitive attainment and progress in English primary schools during Key Stage 2. Investigating the potential continuing influences of pre-school education. *Zeitschrift für Erziehungswissenschaft, Sonderheft*, 11, 179-198.
- Sanders, J. R., & Davidson, E. J. (2003). A model for school evaluation. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (Kluwer International Handbooks of Education, 9, pp. 806-826). Dordrecht: Kluwer Academic Publishers
- Santiago, P., & Benavides, F. (2009). *Teacher evaluation: A conceptual framework and examples of country practices*. Paris: OECD. Retrieved from <http://www.oecd.org/education/school/44568106.pdf>
- Schaffner, E., Philipp, M., & Schiefele, U. (2014). Reciprocal effects between intrinsic reading motivation and reading competence? A cross-lagged panel model for academic track and non-academic track students. *Journal of Research on Reading*, 00 (00), 1-18. DOI:10.1111/1467-9817.12027.
- Scheerens, J. (2002). School self-evaluation. Origins, definition, approaches, methods and implementation. In D. Nevo (Ed.), *School-based evaluation. An international perspective* (35-69). Amsterdam: JAI Press.
- Scheerens, J., & Bosker, R. J. (1997). *The foundations of educational effectiveness*. Oxford: Pergamon Press.
- Schiefele, U., Schaffner, E., Möller, J., & Wigfield, A. (2012). Dimensions of reading motivation and their relation to reading behavior and competence. *Reading Research Quarterly*, 47, 427-463.
- Schlagmüller, M., & Schneider, W. (2006). *WLST 7-12. Würzburger Lesestrategie Wissenstest für die Klassen 7 bis 12*. Goettingen: Hogrefe.
- Scriven, M. (2003). Evaluation theory and metatheory. In T. Kellaghan, D. L. Stufflebeam, & L. A. Wingate (Eds.), *International handbook of educational evaluation* (Kluwer International Handbooks of Education, 9, pp. 15-31). Dordrecht: Kluwer Acad. Publ.
- Shanahan, T., & Shanahan, C. (2008). Teaching disciplinary literacy to adolescents: Rethinking content-area literacy. *Harvard Educational Review*, 78(1), 40-59.

- Scheerens, Jaap (2002). School self-evaluation: origins, definition, approaches, methods and implementation. In: David Nevo (Ed.), *School-based evaluation: an international perspective. Advances in program evaluation*, 8. JAI, Amsterdam, the Netherlands, pp. 35-69. ISBN 9780762308613
- Sinicrope, C., Norris, J., Watanabe, Y. (2007). Understanding and assessing intercultural competence: A summary of theory, research, and practice. *Second Language Studies*, 26 (1), 1-58.
- Solheim, O.J. (2011). The impact of reading self-efficacy and task value on reading comprehension scores in different item formats. *Reading Psychology*, 32, 1-27.
- Spence, J., & Helmreich, R. (1983). Achievement-related motives and behaviors. In J. Spence (Ed.), *Achievement and achievement motives: Psychological and sociological approaches* (pp. 10-74). San Francisco: Freeman.
- Spurrier, N., Sawyer, M., Clark, J., & Baghurst, P. (2003). Socio-economic differentials in the health-related quality of life of Australian children: Results of a national study. *Australian and New Zealand Journal of Public Health*, 27, 27–33.
- Stanat, P., & Christensen, G. (2006). *Where immigrant students succeed -A comparative review of performance and engagement in PISA 2003*. Paris: OECD.
- Stiglitz J., Sen, A., & Fitoussi J.-P. (2009). *Report by the commission on the measurement of economic performance and social progress*. Retrieved from http://www.stiglitz-senfitoussi.fr/documents/rapport_anglais.pdf
- Sykes G., Schneider, B., & Plank, D. N. (Eds.). (2009). *Handbook of education policy research*. New York: Routledge.
- Sylva, K., Melhuish, E., Sammons, P., Siraj-Blatchford, I., & Taggart, B. (2011). Pre-school quality and educational outcomes at age 11: Low quality has little benefit. *Journal of Early Childhood Research*, 9, 109-124.
- Sylva, K., Stein, A., Leach, P., Barnes, J., Malmberg, L.-E., & the FCCC Team. (2011). Effects of early child-care on cognition, language, and task-related behaviours at 18 months: An English study. *British Journal of Developmental Psychology*, 29, 18-45.
- Taboada, A., Tonks, S. M., Wigfield, A., & Guthrie, J. T. (2013). Effects of motivational and cognitive variables on reading comprehension. In D. E. Alvermann, N. J. Unrau, & R. B. Ruddell (Eds.), *Theoretical models and processes of reading* (6th ed.). Newark: International Reading Association.
- Tutu, A., & Constantin, T. (2012). Understanding performance through persistence and job competency. *Procedia – Social and Behavioral Sciences*, 33, 612-616.
- UNESCO. (2006). *Education for all global monitoring report 2007: Strong foundations*, Paris: UNESCO.
- Valenti, J. E., & Tracey, D. H. (2009). Full-day, half-day, and no preschool effects on urban children's first-grade reading achievement. *Education and Urban Society*, 41, 695-711.
- Van de Vijver, F. J. R., Breugelmans, S. M., & Schalk-Soekar, S. R. G. (2008). Multiculturalism: Construct validity and stability. *International Journal of Intercultural Relations*, 32, 93-104.
- Van de Vijver, F. J. R., & He, J. (2014). Report on social desirability, midpoint and extreme responding in TALIS 2013. *OECD Education Working Papers*, No. 107. Paris, France: OECD Publishing. doi:10.1787/5jxswcftw76h-en
- Vansteenkiste, M., Lens, W., & Deci, E. L. (2006). Intrinsic versus extrinsic goal contents in self-determination theory: Another look at the quality of academic motivation. *Educational psychologist*, 41, 19-31.
- Van Yperen, N., Blaga, M., & Postmes, T. (2014). A meta-analysis of self-reported achievement goals and nonself-report performance across three achievement domains (work, sports, and education). *Plos One*, 9, e93594.
- Waters, H. S., & Schneider, W. (2010). (Eds.). *Metacognition, strategy use, and instruction*. New York, NY: Guilford Press.
- Willms, J. D. (2006). *Learning divides: Ten policy questions about the performance and equity of schools and schooling systems*. Montreal: UNESCO Institute for Statistics.

- Wilson, M. (Ed.). (2004). *Towards coherence between classroom assessment and accountability: 103rd yearbook of the National Society for the Study of Education*. Chicago: University of Chicago Press.
- Woesmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2007). *School accountability, autonomy, choice, and the level of student achievement: International evidence from PISA 2003* (OECD Education Working Paper, No. 13.). Paris. OECD.
- Woesmann, L., Lüdemann, E., Schütz, G., & West, M. R. (2009). *School accountability, autonomy and choice around the world*. Northampton: Edward Elgar Publishing.
- Wong, C. A., Eccles, J. S., & Sameroff, A. (2003). The influence of ethnic discrimination and ethnic identification on African American adolescents' school and socio-emotional adjustment. *Journal of Personality, 71*, 1197-1232.
- Wylie, C., & C. Lyon (2015). The fidelity of formative assessment implementation: issues of breadth and quality. *Assessment in Education: Principles, Policy and Practice, 22:1*, 140–160.

IV. FRAMEWORK DEVELOPED FOR PISA 2018: MEASURING ADOLESCENT WELL-BEING IN PISA

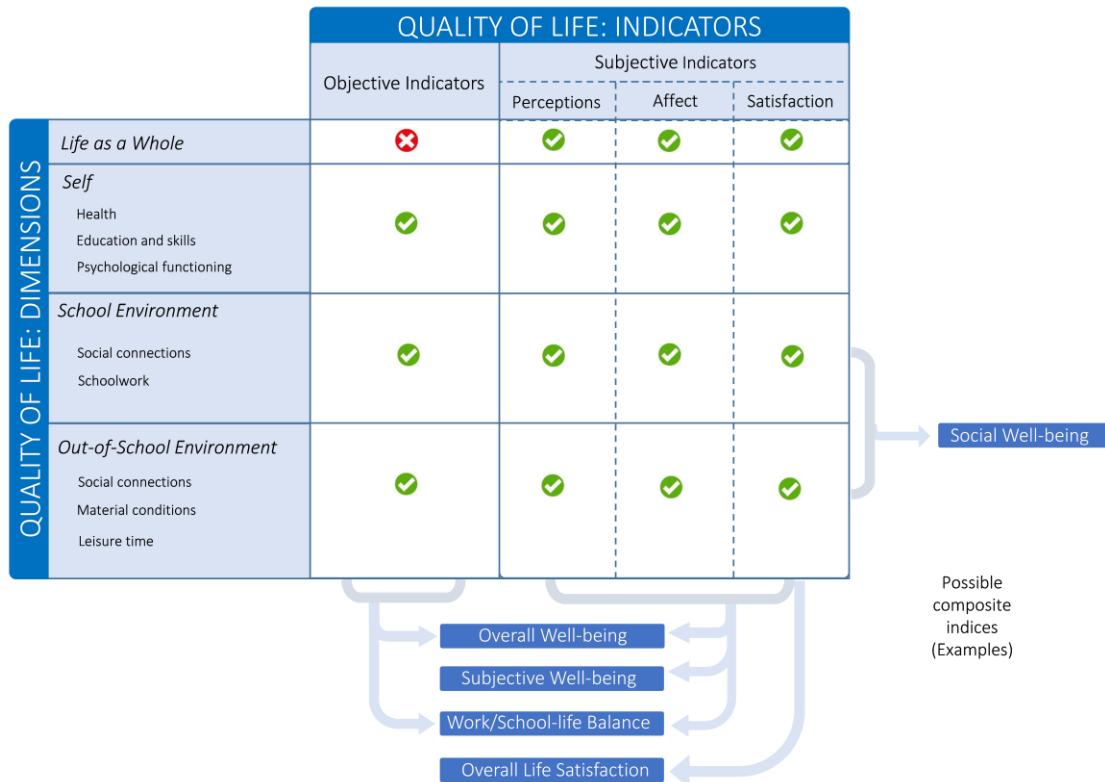
JONAS P. BERTLING, PH.D.

EXECUTIVE SUMMARY

WELL-BEING, defined as the quality of people's lives and their standard of living is of growing interest to policymakers and educators around the world. There seems to be consensus around the notion that well-being is a multi-dimensional construct that comprises both objective, material components and subjective, psychological facets. While there is a growing body of research on the topic, only few large-scale studies for adolescents have taken a comprehensive view on well-being including all of its components. No large-scale assessment directly links well-being to educational achievement of students. Claims regarding relationships between student learning and well-being are therefore limited.

Measuring well-being in PISA bears the chance of creating international benchmarks of student well-being across OECD and partner countries and creating a database of tremendous utility for educators, researchers, and policymakers. The field of well-being research is scattered in itself with different approaches taken in health research, educational research, psychological research, and economics research. The framework outlined here aims at integrating different perspectives on the issue and present a comprehensive model that offers a cross-classification of different well-being dimensions and a spectrum of indicators (including both objective to subjective indicators) available for each dimension.

Figure 16. Framework Overview



The proposed modular framework (see 16) distinguishes three main well-being dimensions in addition to a students' perceived quality of life as a whole:

- First, well-being in term of how fit and healthy students are and how they feel about themselves and their lives (*Self*);
- Second, well-being in terms of the school environment a student is exposed to (*School environment*); and
- Third, well-being in terms of the living environment and circumstances outside of school a student is experiencing (*Out-of-school environment*).

Under each of these broader dimensions, several sub-dimensions can be directly mapped to the quality of life dimensions proposed in other frameworks.

Possible measurement approaches are presented and specific indicators are outlined for all framework components. Recommendations are informed by a review of the relevant literature as well as pragmatic considerations of space in the questionnaire, student burden, and available survey methods for the PISA population. In doing so, special consideration is given to issues of cross-cultural comparability and age appropriateness of the proposed survey methods. In order to measure experienced well-being in a brief and efficient manner, innovative survey methods drawing on techniques of the day reconstruction method are outlined, thereby further extending the approaches successfully implemented in PISA 2015.

The framework is modular in nature in two ways. First, the framework can be broken into modules by area (i.e., self, school, and outside of school). Second, the framework can be broken into modules by evidence fragment within or across all dimensions (i.e., objective well-being indicators, subjective perceptions, affect, and satisfaction). Lastly, the framework defines a set of potential composite indicators drawing evidence fragments from different cells in the framework to create robust reporting elements in areas of key policy interest. Several broader composite indicators are proposed in addition to more specific construct measures. Composite indicators include, among others, indices of overall well-being, subjective well-being, social well-being and work/school-life-balance.

INTRODUCTION

Well-being can be defined as the quality of people's lives and their standard of living and is often quantified in terms of both objective measures, such as household income, educational resources, or health status and subjective indicators such as experienced affect, perceptions of quality of life and life satisfaction (e.g., Casas, 2011). The term subjective well-being has been proposed to refer to subjective measures of well-being (Andrews & Robinson, 1991; Diener, Lucas, & Oishi, 2005; Stiglitz, Sen, & Fitoussi, 2009).

Economists have started pointing out the shortcomings of relying only on gross domestic product (GDP) as indicators of nations' well-being. Several possible alternatives have been proposed (Diener & Seligman, 2004; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004; Stiglitz et al., 2009). For instance, in their 2009 paper, Stiglitz and coauthors recommended that, "Statistical offices should incorporate questions to capture people's life evaluations, hedonic experiences, and priorities in their own surveys" (Stiglitz et al., 2009, p. 58). Several countries have started collecting data and reporting more comprehensive well-being metrics, including measures of subjective well-being (e.g., Boarini, Kolev, & McGregor, 2014; Evans, Macrory, & Randall, 2015; Federal Interagency Forum on Child and Family Statistics, 2009; Foundation for Child Development, 2012; UNICEF, 2007; UNICEF Spain, 2012; Statham & Chase, 2010;

The Children's Society, 2015). Numerous studies have identified important determinants for adult subjective well-being (SWB), often defined as people's experiences of their lives as desirable, following the definition proposed by Diener (e.g., Diener, 1999). Among the most important are health, employment-related factors (income, unemployment) and social contacts (Dolan, Peasgood, & White, 2008; Sacks, Stevenson, & Wolfers, 2010; Winkelmann & Winkelmann, 1998; Helliwell, Layard, & Sachs, 2015). There is empirical evidence that SWB and health are related to important work-related outcomes, with healthy individuals being more productive and making less use of health care (e.g., Keyes & Grzywacz, 2005). In addition, longitudinal studies have shown that mental health is an important predictor of subsequent work performance (e.g., Wright, Bonett, & Sweeney, 1993). Diener and Chan (2011) reported that people who are happier tended to have better health status and higher life expectancy than individuals experiencing frequent anger, depression, or anxiety. Moreover, childhood happiness supports the development of social and self-regulatory skills that in turn may increase the chance of reaching important life outcomes in adulthood (Morgan et al., 2008). Findings pointing to the importance of well-being for general life outcomes and workforce success have increased interest in the well-being construct also for the business community (Beal, Rueda-Sabater, & Heng, 2015).

Policymakers now call for valid information on their citizens' and workforce's well-being in addition to indicators of knowledge and skills. Large international health surveys (e.g., World Health Survey; Health Behavior in School-aged Children; WHO-5, Topp, Oestergaard, Sondergaard, & Bech, 2015; KIDSCREEN, Ravens-Sieberer et al., 2014) and adult household surveys (e.g., Gallup World Poll; Boarini, Comola, Smith, Manchin, & De Keulenaer, 2012) already include measures of well-being. However, most international well-being assessments were so far focused heavily on adult populations. In that vein, the authors of the recently published 2015 Good Childhood Report state that, "People's subjective well-being has become a topic of widespread – and growing – interest. However, discussion of children's subjective well-being has been notable by its absence." (The Children's Society, 2015, p.9). This has started to change with more researchers starting to explicitly focus on the specific issues of adolescent and child well-being (e.g., Ben-Arieh, 2008; Cummins & Lau, 2005; Lippman, Moore, & McIntosh, 2011; Pollard & Lee, 2003; Huebner, 2001; Bradshaw, Keung, Rees, & Goswami, 2011; Gilman & Huebner 2003; Huebner & Dew 1996; Saha, Huebner, Suldo, & Valois, 2010). However, not all studies have focused on the general children and adolescent population but rather on specific subgroups. Based on a larger literature review, Casas (2011), for instance, concluded that, "most studies on child and adolescent well-being and quality of life in the English-language literature have been focused on very specific, small populations, such as children with particular health problems. Very few studies refer to large samples, and cross-national comparisons are scarce." (p. 562).

As the first international large-scale assessment, the *Programme for International Student Assessment (PISA)* added a few selected questions on students' subjective well-being to their 2015 student questionnaires (PISA 2015 Draft Assessment Framework, OECD, 2013b). Data on these questions will add substantial value to the international educational policy and research agenda since, for the first time, well-being indicators can be related directly to achievement on the student level across a large number of economies. However, the set of questions included in PISA 2015 is limited in scope and therefore limits conclusions that may be drawn based on resulting data. PISA 2018 bears the chance of changing that. A separate well-being questionnaire encompassing questions covering the entire well-being construct could constitute the building block for international benchmarks on adolescent well-being. OECD already established guidelines for the measurement of well-being with adults in the context of their *Better Life Initiative* (e.g., OECD, 2015). OECD now has a chance to establish measures of international benchmarks in adolescent well-being.

Monitoring well-being of adolescents is important as this population constitutes the future workforce and thereby determines future chances for countries to compete in an international economy. Media reports about extremely long school hours and raising suicide rates in some high performing countries, or findings that large proportions of students report disliking school and showing diminished school engagement (McGill, Hughes, Alicea, & Way, 2012), experiencing high levels of anxiety stress regarding school (Natsuaki, Biehl, & Ge, 2009), call into questions the trade-offs between different educational and societal objectives. As Helliwell et al. (2015) remarked in the recently published World Happiness Report 2015, "if schools do not measure the well-being of their children, but do measure their intellectual development, the latter will always take precedence." (p. 11).

Schleicher (2015) described three ways in which well-being is of direct policy relevance to PISA, first, as a construct of intrinsic importance representing governments' efforts in ensuring their citizens' and residents' well-being; second, as an important determinant of adult well-being; third, as a substantial driver of educational outcomes in the school system. Educators and policymakers are in deep need of valid and reliable information on student well-being as an international benchmark that is also essential for evaluating the efficacy of policy interventions targeting at child well-being, such as bullying prevention programs.

Since a majority of the research on well-being to date is based on research with adult populations, one crucial initial step in defining well-being measures for adolescent populations is evaluating whether the construct is the same for younger populations or whether aspects of the construct play different roles for these respondent groups. Some of the key components of well-being for adults, such as job satisfaction, earnings, or work-life balance are conceptually rooted in adult life and need to be adapted for younger populations. Also, the relatively different priorities and opportunities of adolescents to spend leisure time, spend time with friends, as well as their relationships with parents, teachers, and adults in general deserve special attention (The Children's Society, 2015). Especially peer relationships become more important in adolescence (Hardy, Bukowski, & Sippola, 2002; Way & Greene, 2006; McGill et al., 2012; Wang & Eccles, 2012; Way, Reddy, & Rhodes, 2007). The potentially greater importance of social connections and relationships for adolescent well-being over material conditions is nicely illustrated by findings that, when a large sample of 14 and 15-year olds were asked about what having a good life meant to them, the five most commonly used terms were "friends", "family", "bullying", "parents", and "school". (The Children's Society, 2015).

I am proposing an integrated framework for measuring well-being in PISA that is informed by other frameworks that have been proposed, for both children and adults, and that integrates so far often separately treated perspectives. The proposed framework aims especially to accomplish the following:

1. The framework explicitly takes into account that well-being is a multi-dimensional construct and that its measurement requires covering different domains, not just overall life satisfaction.
2. The framework incorporates the distinction between overall well-being and subjective well-being. Moreover, the framework distinguishes objective from subjective indicators for student well-being.
3. The explicit focus of the framework is on adolescent well-being, meaning that special emphasis is placed on the life environment of school-aged children. In addition to those well-being indicators used for adults that apply to this population as well, additional indicators are discussed to fully capture adolescent well-being.

4. The framework focuses especially on those areas that seem most promising for PISA for measuring adolescent well-being based on data collected through the main student questionnaire or a supplementary well-being questionnaire. That is, the focus is on *individual* well-being. Special emphasis is placed on those indicators that can be collected from individual respondents. Indicators that might be collected at the system level are briefly mentioned but not elaborated on in detail. This applies, for example, to aspects of environmental quality, crime or employment statistics, which might be inferred from available other data sources based on a school's location.
5. A separate section is devoted to measurement challenges and how these may be mastered in the context of PISA. Measurement challenges include age appropriateness of items and item formats, cross-cultural comparability, and respondent burden. Multi-method assessment strategies involving self-reported biodata and behaviors, subjective self-reports, and elements of the *day reconstruction method* or *event reconstruction method* (e.g., Kahneman & Krueger, 2006) are outlined as possible solutions.
6. The framework includes a detailed appendix with a proposed well-being assessment plan for PISA 2018 that also specifies which components of the framework are already covered by previous PISA questionnaires, and which components would need to be added.

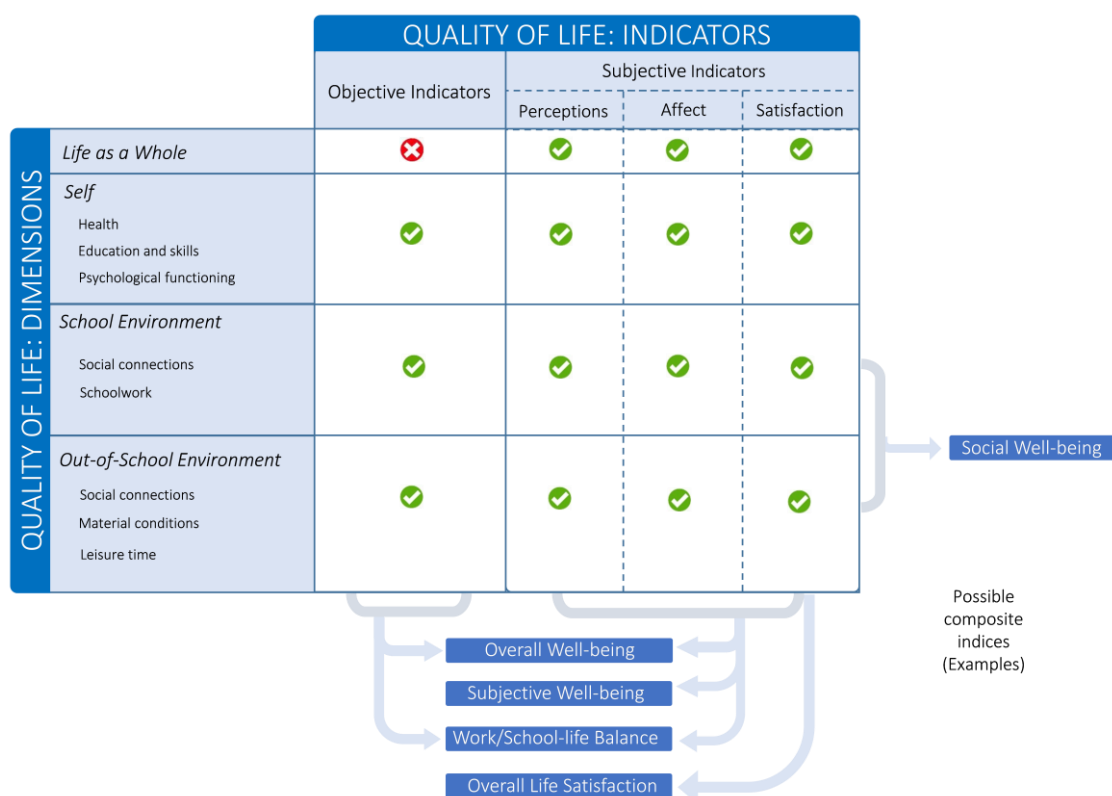
1. WELL-BEING AS A MULTI-DIMENSIONAL CONSTRUCT

Adolescent well-being, defined as the quality of students' lives and their standard of living, is of growing interest to policymakers and educators around the world. There seems to be consensus around the notion that well-being is a multi-dimensional construct that comprises both objective, material components and subjective, psychological facets. While there is a growing body of research on the topic, only few large-scale studies for adolescents have taken a comprehensive view on well-being including all of its components. Rather, some studies have focused mainly on material well-being and health outcomes (e.g., HBSC) and other studies focusing more heavily on subjective well-being (e.g., Children's Worlds, Gallup Student Poll). None of these studies directly links well-being to educational achievement of students.

Despite sometimes used interchangeably, it is important to differentiate between well-being and subjective well-being, the latter being a sub-component of the larger well-being construct. Well-being is a multifaceted construct that comprises different dimensions with each of them entailing both objective and subjective components. Subjective well-being can be defined as "people's evaluations of their lives—the degree to which their thoughtful appraisals and affective reactions indicate that their lives are desirable and proceeding well." (Diener, Oishi, & Lucas, 2015). Subjective well-being includes both affective (i. e., the experiences of pleasant emotions as well as negative emotions) and cognitive components (i.e., judgment of individuals' overall life satisfaction or satisfaction with specific domains). The framework outlined here aims at integrating different perspectives on the issue and present a comprehensive model that offers a cross-classification of different well-being dimensions and a spectrum of indicators (including both objective to subjective indicators) available for each dimension. That is, it combines the interest in an overall measure of well-being with a more nuanced multi-faceted assessment, acknowledging that context-specific assessments can yield more differentiated information (e.g., Casas et al., 2013).

Figure 17 presents a graphical depiction of the overall framework. Three main well-being dimensions are distinguished in addition to overall well-being: First, well-being in term of how fit and healthy students are and how they feel about themselves and their lives (*Self*). Second, well-being in terms of the school environment a student is exposed to (*School environment*).. Third, well-being in terms of the living environment and circumstances outside of school a student is experiencing (*Out-of-school environment*). Under each of these broader dimensions, several sub-dimensions (e.g., social connections, or health) can be directly mapped to the 11 quality of life dimensions proposed by the OECD better life initiative (OECD, 2013a), as well as to other key frameworks described in the literature.

Figure 17. Proposed PISA 2018 Well-being Framework



While not always labelled in the same way, the dimensions distinguished in the proposed model show high conceptual alignment with well-being dimensions described in other frameworks. For instance, Lippman et al. (2011) distinguished between indicators of child well-being related to the self (physical health, development, and safety; cognitive development & education; psychological/emotional development; social development and behavior), relationships on different levels (family, peers, school, community, macrosystem), and contexts (family, peers, school, community, macrosystem). Adamson (2007) suggested a classification of well-being into the following clusters: material well-being, health and safety, educational well-being, interpersonal relationships, behavior and risks, and subjective well-being. Land, Lamb, & Mustillo (2001) proposed a composite index of Child and Youth Well-being in the United States composed of the following components: material well-being (poverty, employment, income), health (mortality rate, personal health), social relationships, safety/behavioral concerns (e.g. smoking), educational attainment, place in community (includes enrolment and engagement), and emotional well-being. Bradshaw, Hoelscher, and Richardson (2007) presented a framework based on the following components: material situation (poverty, deprivation, and parental

worklessness), housing, health, subjective well-being (self-defined health, personal well-being, well-being at school), education, children's relationships, civic participation, and risk and safety. Moore et al. (2008) proposed a microdata child well-being index based on: child and health safety, child educational achievement and cognitive development, social and emotional child development, family processes, and family demographics. They further distinguish individual child well-being from contextual well-being, the former comprising physical health, psychological health, social health, and educational/intellectual, and the latter comprising family, community, and socio-demographic factors.

One key difference to most of the previously proposed frameworks is that I conceptualize subjective well-being as a cross-dimension that intersects with all other well-being dimensions rather than as an independent separate dimension. The components of the subjective well-being construct previously proposed in the literature (i.e., life satisfaction, affect, psychological functioning) are specified both in regards to well-being overall and in regards to each well-being dimension, including individuals' subjective perceptions of their material conditions. As Casas et al. (2011) pointed out, assessing quality of life needs to involve measuring material and non-material characteristics of life in large populations, and subjective measures should be utilized to add to objective measures rather than replace them. By distinguishing several aspects among subjective well-being indicators, the proposed framework also ties back to early work on the matter. Almost four decades ago, Campbell and colleagues (Campbell, Converse & Rogers, 1976) proposed conceptualizing non-material quality of life as peoples' "perceptions, evaluations, and aspirations concerning their own lives and life conditions".

The framework is modular in nature in several ways. First, the framework can be broken into modules by area (i.e., life as whole, self, school, and outside of school). Second, the framework can be broken into modules by evidence fragment within or across all dimensions (i.e., objective well-being indicators, subjective perceptions, affect, and satisfaction). Lastly, the framework defines a set of potential composite indicators drawing evidence fragments from different cells in the framework to create robust reporting elements in areas of key policy interest. In addition to the proposed composite indicators of overall well-being and subjective well-being, composite around emotional well-being, social well-being, life satisfaction and work-life-balance (school-life balance) are suggested for further consideration. A work-life balance index could, for instance, serve as a benchmark for how well students in different countries are able to integrate curricular demands and school life with time for personal activities, leisure, and maintenance of a healthy lifestyle. By proposing the creation of broader composite indices, the framework directly addresses the policy need for a smaller set of robust reporting elements following the model of similar fashion as PISA's index of *Economic, Social, and Cultural Status* (ESCS) that is now widely used as a benchmark measurement of socioeconomic status and often considered a gold standard. Several authors have emphasized the value of creating composite well-being indices, particularly to make measuring progress easier and facilitate comparisons between trends across sub-groups or regions (Ben-Arieh 2008, Fernandes, Mendes, & Teixeira, 2012; Land et al., 2007).

Before describing each of the framework components in detail, the next section is devoted to important measurement challenges and methodological considerations.

2. ADDRESSING MEASUREMENT CHALLENGES

Measuring psychological, subjective, or noncognitive constructs in PISA faces the challenges of how robust measurement approaches can be implemented while keeping student burden low, minimizing perceived intrusiveness of used questions, and maximizing cross-national and cross-cultural comparability of recorded responses. The success of collecting valid cross-national data on students' well-being will depend on the degree to which these measurement challenges can be addressed. In the following, I summarize five implementation recommendations, each of them directly addressing one measurement challenge in the PISA context that provided a basis for the selection of proposed measures as listed in the appendix.

A. Balance single-item measures with multi-item indices

Previous studies aiming to measure well-being have often relied on single-item indicators or a set of very few questions. While this approach is appealing from the administration perspective, it is dissatisfying from the conceptual as well as the practical validity and reporting perspective. In order to provide valid and reliable measurement of well-being across nations and economies it is crucial to rely on multiple indicators for the construct at hand. Providing policy makers with robust and actionable data therefore requires a multi-item measurement approach. Compiling country rankings based on a single question alone seems problematic in several ways: First, there is no sufficient consensus in the research literature what the best single question for measuring well-being is, and research evidence on the equivalence of approaches is insufficient. Second, using a single question or a small set of single-item indicators will overemphasize certain aspects of well-being while underrepresenting others and thereby fail to capture the construct in its entirety. Third, different questions might be to different degrees equivalent across countries and some well-being questions are likely to be more affected by cross-cultural norms and response styles than others. Creating a valid international well-being indicator requires sampling a larger number of questions and field-testing them in order to select the most appropriate questions for operational use. Fourth, reporting elements based on multiple questions can increase the reliability of well-being indicators used in cross-country comparisons. These recommendations are consistent with recommendations from Casas et al. (2012) who recommend that future research for cross-country comparability should collect data using more than one scale in a given area and that "we need much more data and from more countries to analyze in any real depth the qualities and possible weaknesses of each scale for the international comparison of adolescent populations." (p.26)

At the same time, it should be noted that some components of the overall well-being construct require fewer questions to ensure valid and reliable measurement than others. Specifically some module components that represent objective indicators might be captured directly as manifest variables. However, not all variables of interest can be directly measured calling for the use of one or several proxies for a manifest variable of interest. Creating multi-item indicators would be consistent with current practice for the PISA student questionnaires and is used by other large-scale testing programs as well. TIMSS and PIRLS are currently using a multi-item index approach and NAEP has recently moved to an index approach for more robust reporting (Bertling, 2014).

B. Use meaningful number of scale points and clearly distinguishable scale labels

Most of the established well-being instruments have been developed and validated for adult populations. This poses the challenge of adapting questionnaire items to the PISA age population. These instruments tend to use response formats with substantially more response scale points than what is the current PISA practice. For instance, a 0-10 or 1-10 scale where only the scale endpoints are verbally labelled is a widely used question type. One well-documented issue with such subjective well-being questions is that observed response category frequency distributions are highly skewed with large proportions of respondents way above the scale midpoint. The very sparse frequencies on the lower end of the response scale call into question whether all scale points are conceptually meaningful and practically useful. Moreover, it motivates the question whether scales with fewer scale points might be equally valid (or even more valid). This problem seems even more severe for younger respondents than for adults. Data from the 2015 Children's Worlds Survey show, for instance, that across 15 countries more than 80% of all student responses fell into the categories 9 and 10 on a zero to 10 scale. (Rees & Main, 2015). Because of the well-documented problems of partially labelled response scales, researchers have recommended using scales with fully labelled response options whenever possible (Dillman et al., 2014; Gehlbach, 2015; Krosnick & Fabrigar, 1997). Reducing the number of scale points below 11 points could potentially improve the validity of questions used in PISA.

While many of the well-being questions used with adults seems somewhat problematic due to too large numbers of scale points, the currently dominant question format in the PISA student questionnaires might be subject to the opposite problem. PISA currently uses mostly a 4-point Likert-type response format with the verbal anchors "Strongly disagree", "Disagree", "Agree", and "Strongly agree". Both the number of only four scale points and the nature of the verbal labels of agreement type have been criticized in the survey literature. Findings from leading researchers in the field (e.g., Gehlbach, 2015; Tourangeau et al.) point to problems with too few scale points and caveats of the agreement type questions that have motivated other national large-scale testing programs, such as NAEP, to develop alternatives to this format (Bertling, 2015).

Finding the right compromise between fewer versus more scale points and fully labelled versus unlabeled response scales is a balancing act. Following the best current practices recommended by survey researchers stands in contrast to aligning the PISA well-being questionnaire with established questions in order to provide means for comparisons with data from other surveys. The need for translatability across a large number of languages further poses a challenge for potentially extending responses scales beyond four scale points. It is recommended, if feasible, to compare alternative item versions in the PISA field trial, such as items with fully labeled versus incompletely labelled response options and items with fewer scale points. If such comparisons are not feasible in the field trial, questions should be adapted to a format that aligns with established best practices in PISA.

C. Select measures that maximize cross-cultural comparability

Another challenge is developing well-being questions that allow for comparison of resulting data across cultural and national borders -- a challenge that is well documented for PISA in general and other international surveys (e.g., Kyllonen & Bertling, 2013). Designing robust reporting elements that allow for cross-country and sub-group comparisons therefore should be a pivotal goal for new survey questionnaire development. Classical measurement approaches based on self-reports often suffer substantial methodological problems, especially limited inter-individual comparability of ratings due to individual or group-specific response styles. Individuals from different countries and cultures may demonstrate different response styles on traditional rating scales. There is ample evidence that responses to seemingly objective questions can often not be interpreted as such objective indicators and suffer limited comparability across countries without accounting for these response style differences (e.g., Kim, Schimmack, & Oishi, 2012). Anchoring Vignettes have been successfully applied here (e.g., Angelini, Cavapozzi, Corazzini, & Paccagnella, 2013; Kristensen & Johansson, 2008; Kyllonen & Bertling, 2014; Salomon, Tandon, & Murray, 2004; Van Soest, Delaney, Harmon, Kapteyn, & Smith, 2011). At the same time, while alternative survey methods (e.g., anchoring vignettes, situational judgment tests, forced choice) could be explored for the well-being questionnaire as well, some of these methods might be less valuable for the well-being construct in specific as many components of the construct explicitly involve a subjective component, which – by definition – will be influenced by cultural norms and the personality of the respondent, in addition to general response sets and generalizable response behaviors. White (2008) identified ‘culture’ as a key influence on the way one’s perception of well-being is constructed and suggested, therefore, that wellbeing should be understood as a ‘process’ grounded in a specific ‘time’ and ‘place’. While the construct at hand, especially its subjective indicators, might to some extent capture culture-specific aspects of student perceptions, ensuring cross-cultural comparability might be maximized by choosing clear, translatable, and where possible quantifiable response formats and include short definitions as part of the item stems whenever a questions requires knowledge of certain assumptions that might limit cross-cultural comparability. If the field trial design allows, additional comparison of alternative item formats might be considered.

D. Consider item formats beyond traditional self-reports

Measuring experienced well-being requires a survey approach that goes beyond the typical self-report questionnaires traditionally used in large-scale assessments. Implementing a shortened form of the day reconstruction method and event reconstruction method could offer a real alternative here to assess how students spent their time, especially the time outside of school, and sample their experienced well-being during various activities. A day or event reconstruction questionnaire would aim at collecting data on two crucial pieces of information: (a) what activities did a student engage in, and (b) how did the student feel during these activities. Given the space constraints, including questions on time spent on each activity might not be feasible whereas a focus on experienced affect and activities only would be a viable alternative here. The proposed questions build upon a time use module proposed for PISA 2015 (Bertling & Kyllonen, 2012), which was partly implemented in the 2015 main survey as well as on methods and question formats recommended by the authors of the original day reconstruction and event reconstruction methods (Grube, Schroer, Hentzschel & Hertel, 2008; Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004; Schwarz, Kahneman, Xu, Belli, Stafford & Alwin, 2009).

E. Consider alternative questionnaire designs to reduce respondent burden

Constraints of overall testing time and the large sample sizes in large-scale assessments make matrix sampling approaches where different respondents receive different sets of items a viable option to reduce burden while maintaining content coverage across relevant areas. Matrix sampling approaches are the standard practice for the subject-area tests in educational large-scale assessments (Comber & Keeves, 1973; OECD, 2014b) and have more recently been proposed as a potentially viable alternative to fixed questionnaires. A 3-form matrix sampling design has been applied to the student questionnaire in PISA 2012 leading to an increase in questionnaire content by 33 percent (OECD, 2013b). A design with ten partly overlapping questionnaire booklets was implemented in the 2013 pilot for the 2015 NAEP Technology and Engineering Learning (TEL) assessment (Almonte, McCullough, Lei, & Bertling, 2014).

These designs create data with missing data structures that require new analytical approaches. Research findings to date are inconclusive regarding the risks and benefits of questionnaire matrix sampling in practical scenarios. While many researchers reported substantial increases in content coverage with very small to negligible impact on the overall measurement model including conditioning and estimation of plausible values (Adams, Lietz, & Berezner, 2013; Almonte et al., 2014; Kaplan & Wu, 2014; Monseur & Bertling, 2014), methodological concerns about possible biases in estimation results have also been raised (von Davier, 2014). Application of mass imputation for all questions that were not administered to a given student might address these issues by creating full datasets with large proportions of imputed data. This approach has been explored in research contexts (e.g., Kaplan & Wu, 2014) but so far was not implemented for operational questionnaires in LSAs. As argued in Bertling, Borgovoni, & Almonte (in press) more generally for survey questionnaires, explorations of potential use of a matrix sampling design for the well-being questionnaire in specific might be also beneficial. Such a design could allow for inclusion of questions for a larger number of facets of the larger well-being construct or for inclusions of more questions for a given construct (e.g., a larger set of affective states in the experienced well-being questionnaire) without increasing individual student burden.

3. SUGGESTED QUALITY OF LIFE INDICATORS

A. Quality of Life as a Whole

Well-being in regards to life as a whole (or overall life satisfaction) is often used as a single indicator for individual subjective well-being, even though it does not capture subjective indicators of more specific quality of life dimensions (e.g., quality of relationships). In addition, no direct objective indicators for well-being in regards to life as a whole are available. Despite the importance of including indicators of this framework component as an important yardstick in any well-being instrument, questions on well-being in regards to life as a whole are deemed insufficient to capture the entire well-being construct.

a. Life Evaluation and Life Satisfaction

Life satisfaction, an evaluation of an individual's quality of life, is an important aspect of well-being (Diener, 1999) that is closely linked to subjective health (Huebner, Valois, Suldo, Smith, McKnight, Seligson, & Zullig, 2004). Classical approaches of assessing subjective well-being rely mostly on unanchored self-report ratings from a large number of individuals. A typical question for the assessment of general life satisfaction is "Overall, how satisfied are you with life as a whole these days?" Respondents have to give a response on a scale from 0-10 with zero indicating "not at all satisfied" and 10 meaning "completely satisfied". Sub-domains can be incorporated into the assessment by asking a block of questions related to specific domains (e.g., health, personal relationships, security, etc.). Domain scores can then be treated as stand-alone scores or aggregated into an overall index representing the average satisfaction across all domains. Questions targeting at domain satisfaction are an integral part of the proposed well-being assessment framework as illustrated also in Figure 16. A composite index capturing life satisfaction may be created by aggregating subjective satisfaction ratings across all framework dimensions (see section I.1.A.d in this framework).

There are two alternative approaches that are widely used in the well-being literature and have sometimes been considered exchangeable or equivalent. One approach is the so-called *Cantril Ladder* (Cantril, 1965), which is used in major international surveys such as the Gallup World Poll or the Gallup Student Poll. By asking respondents to indicate where, on a ladder with steps from 1-10 - representing the range from "the worst possible life" to "the best possible life" - they would see themselves at the current point in time, the question targets at the evaluative aspect of well-being, i.e., how individuals perceive or evaluate their life. The alternative question targets at the same overarching construct but focused more on the satisfaction (rather than the evaluation) component. The question, typically worded "How satisfied are you with your life overall these days?" is part of the core well-being questions OECD recommends in their guidelines on measuring subjective well-being (OECD, 2013a).

From a conceptual point of view, the questions measure two distinct albeit related aspects of subjective well-being, live *evaluation* and life *satisfaction*. Empirical findings are somewhat inconclusive regarding the comparative validity of the two questions, particularly for adolescents (e.g., Casas et al., 2012). Research is currently lacking large-scale studies that systematically compare the nuanced difference between the two questions. Key findings to date are that the Cantril ladder question tends to create data that varies more within samples and produces average scores closer to the mid-point on the scale. Data from multiple large studies on the direct life satisfaction-rating question shows that score distributions are skewed more with the mean closer to the upper end-point of the scale. Data from the PISA 2015 field trial, where both questions were compared, confirmed these findings. Both questions were field trialed in all

participating countries and no clear advantages of the ladder over the satisfaction question, apart from slightly better differentiation across the scale, were found. The two questions correlate strongly. In their guidelines on measuring well-being, OECD has recommended using the 11-point life satisfaction scale as part of the Core well-being module (OECD, 2013a). The question is shorter and therefore requires less time to answer. It is also easier to understand because of lower cognitive burden and reduced reading load. Lastly, the satisfaction question might be classified as less intrusive than the ladder question because it does not explicitly introduce a concept of social rank or comparison with other individuals, which might elicit negative emotionality in some respondents. Minimizing the perceived intrusiveness of survey questions and keeping the cognitive burden low are important considerations in any international study especially with younger respondent groups. OECD included the 11-point satisfaction question as a new survey question in the 2015 PISA student questionnaire. It is therefore recommended to include the 0-10 satisfaction questions as a core question in the PISA well-being questionnaire as well. If space allows, getting additional data from the Cantril ladder question could be beneficial to further increase the robustness of the measure.

While several multi-item scales for overall subjective well-being have been proposed (e.g., Adelman, Taylor, & Nelson, 1989; Huebner, 2001; Rees & Main, 2015), their incremental value over the single item indicators described above is somewhat unclear. Given the space and burden constraints for PISA, it is therefore recommended not to include any multi-item scale on students' overall life evaluation. Rather, aggregating satisfaction ratings across multiple domains into a potential life satisfaction composite index (see section I.1.A.d in this framework) would allow optimization of questionnaire length and tapping into important facets of the well-being construct. This strategy also aligns with the direction well-being research has taken from investigation of global subjective well-being to domain-specific subjective well-being (e.g., Elmore & Huebner, 2010; Gilman & Huebner, 2000; Long, Huebner, Wedell, & Hills., 2012; Tian, Wang, & Huebner, 2015), and empirical findings that adolescents' domain-based reports on aspects of well-being and satisfaction (e.g., family, school) show incremental validity over global life satisfaction reports (Haranin, Huebner, & Suldo., 2007), and valuable indicators of adaptive functioning in children and adolescents (Whitley, Huebner, Hills, & Volois, 2012).

b. Affect/ Emotional Well-being

Several subjective indicators for emotional well-being are proposed, drawing on (a) both positive and negative affect as indicators of emotional well-being and (b) experienced well-being questions in regards to very specific activities and emotional states experienced over extended periods.

Affect can be measured by asking individuals whether or to what degree they have felt specific emotions (e.g., "Overall, how happy did you feel yesterday?" or "Overall, how angry did you feel yesterday?") during a certain period. This corresponds to Watson's positive and negative affect schedule (PANAS; e.g., Watson, Clark, & Tellegen, 1988), which has been used extensively in psychological research. A measure of hedonic balance has been proposed as the difference between positive and negative affect (Watson et al., 1988). There is less consensus about the specific emotions that need to be sampled. The circumplex model of affect (Russell, 1980) represents an important reference point for choosing relevant emotional states to sample in the experienced well-being questionnaire. Laurent et al. (1999) presented a children version of the Positive and Negative Affect Scale (PANAS-C). In addition to PANAS-type measures, several other instruments are available to measure the affective or emotional aspects of individual well-being, both for adult and children or adolescent populations. While a PANAS-type measure would provide general data on experienced positive and negative affect, there is not sufficient

research showing that this type of instrument works well in an international student population. Instead, a specific measure that is recommended for adaptation to PISA is the KIDSCREEN-10 (Ravens-Sieberer et al., 2014). The KIDSCREEN embraces the conceptualization of quality of life as a multidimensional construct that includes physical, emotional, mental, social, and behavioral components. The short 10-item measure has been shown to be Rasch-scalable (Ravens-Sieberer, Erhart, Rajmil, et al., 2010) and is since 2005 used in a large number of countries as part of the HBSC survey (Currie, Gabhainn, & Godeau, 2009). These properties make the measure an excellent choice for PISA as a short unidimensional measure of global well-being. If space in the questionnaire allowed, the WHO-5 (see Topp, Oestergaard, Sondergaard, & Bech, 2015, for a recent review), a short 5-item affective well-being measure that is widely used and well established in clinical research, might serve as an additional benchmark to link PISA with other surveys. In direct comparison, however, the KIDSCREEN-10 should be given priority given its prior international use with students and the less clinical focus.

The cardinal method for measuring experienced well-being with regard to specific events or behaviors is by applying the so-called day reconstruction method (DRM, Kahneman). In the DRM, respondents are asked to revisit a previous day and report in detail on activities as well as emotional states they experienced. The original DRM is not viable for inclusion in PISA given the time and scoring requirements the method imposes. However, a PISA Well-being questionnaire bears the chance of asking students to report on their emotional states experienced in regard to specific events of interest to PISA, such as well-being during specific classes, during homework, during leisure activities with friends or time spent with their parents or guardians. Thereby, affective states may be related to specific before-school, at-school, and after-school activities. In addition to sampling specific emotional states successfully used in adolescent or child emotional well-being instruments, the emerging literature on the psychological concept of flourishing (Seligman, 2012) represents an important anchor point for new question development. The rationale here is that, while many of the established scales have a strong clinical and mental health focus, it seems important for PISA to also focus on engagement and flow-related emotional states that relate more directly to academic achievement, such as feeling challenged or inspired. The specific proposed event reconstruction questions targeted at measuring experienced well-being are referenced in the following sections based on their classification under one of the well-being dimensions (i.e., self, school environment, or out-of-school environment). All questions are also listed in the appendix.

B. Self-related well-being

The first broad quality of life domain under the larger umbrella of quality of life as a whole is quality of life in regards to the student as an individual, with the three sub-dimensions health, education and skills, and psychological functioning.

a. Health

In order to capture the overall health construct, it is proposed to collect data on a set of key objective and subjective indicators⁶, some of which have already been introduced to the PISA 2015 student questionnaires or been field trialed (Bertling & Kyllonen, 2012).

⁶ Note, PISA will have to rely on student self-reported data for both sets of indicators. The key difference is that objective indicators are clearly quantifiable and behavioral indicators that require minimal judgment or interpretation on the respondent side (for example, a student does not need to provide an interpretation of his weight when providing info on his/her weight in kilograms).

- *Objective Indicators*

In terms of a student's physical condition, *Quetelet's index* (Weight/Height²), better known as the *Body Mass Index* (BMI; see e.g., WHO, 2000, 2010; Garrow & Webster, 1985), which is a key health indicator in the literature and widely used in international studies with adults and younger populations, serves as an important anchor for any other health measure. BMI is based on a student's height and weight and can be utilized as an indicator of overweight and obesity, which are a growing health problem among adolescents in many countries (Lobstein, Baur, & Uauy, 2004; Haug et al., 2009; Rokhol, Baker, & Sorensen, 2010; WHO, 2010). Previous research has shown correlations between overweight and behaviors associated with health risks (e.g., skipping breakfast, being less physically active, or watching more television), overall impaired quality of life (Haug et al., 2009; Must & Tybor, 2005; Williams et al., 2005), and increased likelihood to become a victim of bullying (Janssen, Craig, Boyce, & Pickett, 2004). Research with adolescent populations further suggests that dieting and unhealthy weight control behaviors are related to significant weight gain over time (Neumark-Sztainer, Wall, Story, & Standish, 2012). BMI may be used as a measure to track weight status in student populations across PISA countries and economies and as a screening tool to identify potential weight problems in individuals. BMI should not be used as a single diagnostic tool for body fat or overall student health (see e.g., Nihiser et al., 2007), and interpretations need to account for potential differences in meaning across racial or ethnic groups (e.g., James, Chen, & Inoue, 2002) or gender groups (Dupuy, Godeau, Vignes, & Ahluwalia, 2011). To better account for potential inaccuracies in student provided weight and height information, it is proposed to add two qualifying questions "When did you last weigh yourself?" and "When did you last measure your height?", both of which are currently used in the HBSC survey.

Engagement with physical exercise (both moderate and vigorous) constitutes not only an important positive indicator of student health but also a protective factor against excessive body-image concerns (Gaspar, Amaral, Oliveira, & Borges, 2011) and long-term negative physical and mental health outcomes particularly as habits established in adolescence are likely to be carried through into adulthood (Malina, 1991; Hallal, Victoria, Azevedo, & Wells, 2006; Iannotti et al., 2009; McMurray, Harrell, Creighton, Wang, & Bangdiwala, 2008; Sibley & Etnier, 2003). Children who more frequently play sports or exercise report higher levels of subjective well-being (Abdallah, Main, Pople & Rees, 2014). Moreover, research findings indicate that physical activity may also improve cognitive performance (Martinez-Gomez et al., 2011; Sibley & Etnier, 2003). The World Health Organization recommends that children participate in at least 60 minutes of moderate-to-vigorous physical activity daily (Strong et al, 2005), and this standard has been adopted by several government and professional organizations. A small set of questions about students' engagement with physical exercise were introduced to PISA 2015, covering both moderate and vigorous activities, participation in PE classes, and physical exercise on a previous day. These questions are also valuable questions for considering as part of a PISA well-being questionnaire. In addition, information regarding a student's typical sleep duration and health risk behaviors might be collected via a brief day reconstruction checklist. All proposed questions are listed in the appendix.

- *Subjective Indicators*

In addition to these objective indicators, subjective indicators are important to capture the overall health construct. It is suggested that PISA uses established measurement strategies for these variables by relying on instruments validated in other contexts where available. Among the indicators proposed are: Perception of and satisfaction with body image (Rudd & Lennon, 2000); satisfaction with sleep, perceived overall health, psychosomatic complaints, and satisfaction with overall health.

Across all industrialized countries, body-weight concerns and adolescent's dissatisfaction with their body image have increased (Cash, Morrow, Hrabosky, & Perry, 2004). Findings from the most recent HBSC survey and other studies indicate that female students report greater dissatisfaction with their body image (Marcotte, Fortin, Potvin, & Papillon, 2002) and that dissatisfaction with their body-weight increases the likelihood to engage in unhealthy weight-reduction activities (Ojala et al., 2007) and report substance abuse, risky sexual behavior, and poor mental health (e.g., Kaufman & Augustson, 2008; Kvaem, von Soest, Traen, & Singaas, 2011; Verplanken & Velsvik, 2008; HBSC 2010).

The HBSC includes a short symptom checklist as a non-clinical measure of mental health that can be used to measure psychosomatic complaints (Ravens-Sieberer et al., 2008). The checklist reflects both psychological health (e.g., nervousness or irritability) and somatic health (e.g., headaches or backaches), both of which are strongly related (Petersen et al., 1997, Brosschot, 2002) and important facets of the overall well-being construct (Petersen et al., 1997; Vingilis, Wade, & Seeley, 2002, Hetland, Torsheim, & Aaro, 2002).

All questions are listed in the appendix.

b. Education and Skills

A student's education and skills, as well as his or her self-perceptions of the ability to perform specific academic tasks and general confidence in one's capabilities are important aspects of the overall well-being construct.

- *Objective Indicators*

Objective indicators for students' knowledge and skills are available from the cognitive PISA scores and not further elaborated on here.

- *Subjective Indicators*

Questions about students' competency beliefs or their academic self-efficacy (Bandura, 1994) directly address one of the three main basic psychological needs distinguished in self-determination theory (Ryan & Deci, 2000), namely competence. Research findings have shown that adolescents' perceptions of their school performance and own competency predicts perceived health and well-being (Suldo, Riley, & Shaffer, 2006; Ravens-Sieberer et al., 2004), life satisfaction (Sulder & Huebner, 2006), and lower rates of bullying (Nansel et al., 2001). Qualitative studies further point to the importance of positive attitudes (Edwards & Lopez 2006), personal strengths (Shikako-Thomas et al. 2009), and a positive self-image (Helseth & Misvaer, 2010) as important determinants of student well-being. The PISA trend academic self-efficacy questions included in the student questionnaire might provide data on subjective competency perceptions. Questions about students' satisfaction with their own knowledge and skills and their self-confidence are proposed as part of a question set focusing on domain satisfaction. Specific items are listed in the appendix.

c. Psychological Functioning

Psychological Functioning (also referred to as "eudaimonic well-being" or "flourishing", Seligman, 2012) has been proposed in the literature as an additional component of the subjective well-being construct. Eudaimonic well-being is concerned with people's sense of meaning, purpose, and engagement. People might be considered 'flourishing' if they score highly on measures of subjective and psychological wellbeing (The Children's Society, 2015). Well-being in terms of flourishing describes optimal functioning – or "flow" as a gratifying experiential state that can "make life worth living" (Csikszentmihalyi, 1975; Csikszentmihalyi and Csikszentmihalyi 2006) -, personal growth, self-expressiveness, and the pursuit of meaningful goals (Ryan & Deci, 2001). Some researchers conceptualize psychological functioning as being part of the overall subjective well-being construct (e.g., Seligman, 2012; Kern, Waters, Adler, & White, 2015) others do not (e.g., The Children's Society, 2015). While there is large consensus on two key building blocks of the subjective well-being construct (life satisfaction, and affect), there is less consensus on the nature and role of psychological or eudaimonic well-being. This might be (partly) due to the overlap of the construct with other aspects of well-being. For instance, questions targeting eudaimonic well-being (e.g., "I like being the way I am") are very similar to questions measuring overall subjective well-being (e.g. "My life is just right", Huebner, 1991) and some authors have conceptualized flourishing as a higher-level construct comprising positive and negative affect as facets. Three main facets of eudaimonic well-being that are described in the literature (e.g., OECD, 2013a) are competence, autonomy, and meaning/purpose, as well as optimism. The proposed framework conceptualizes psychological functioning as part of the self-related dimension of well-being and not as a measure of overall well-being as psychological functioning focuses explicitly on the self and does not encompass environmental factors whereas the variables described as potential indicators of overall well-being did not have that level of specificity. Large-scale data gathered through PISA will provide an important database to gain a better understanding whether psychological functioning variables relate more strongly to overall well-being or the more specific well-being dimensions.

Several questionnaires for the assessment of eudaimonic well-being have been proposed (e.g., Huppert et al., 2009; Kern et al., 2015). These questionnaires resemble instruments used to assess psychological constructs such as personality, self-concept, locus of control, or attribution. Sample items are "I am always optimistic about my future" or "I am free to decide for myself how to live my life". Ryff (1995) proposed six dimensions of psychological functioning: self-acceptance, positive relations with others, personal growth, purpose in life, environmental

mastery, and autonomy. Others have distinguished the following dimensions: Self-acceptance, environmental mastery, positive relationships, autonomy, purpose in life, personal growth (The Children's Society, 2015). Note, these areas show considerable overlap with components of the proposed well-being framework as well as with many of the currently used attitudinal and self-related questions in the PISA student questionnaire. For instance, the PISA 2012 and PISA2015 questions tapping into personality facets (i.e., perseverance, openness for problem solving) overlap with the dimensions personal growth and autonomy. Openness to new experiences plays an important role particularly in promoting optimal experience among adolescents (Bassi, Steca, Monzani, Greco, & Delle Fave, 2014).

The concept of psychological functioning adds an important additional facet to the overall well-being construct. However, rather than creating a separate eudaimonic well-being scale based on a unique set of questions, it is proposed to conceptualize psychological functioning as a potential composite index. Such an index might be created based on framework components capturing various subjective perceptions, such as perceptions of competence, knowledge and skills; autonomy, personal freedom, and opportunities; meaning and purpose; as well as other subjective perceptions of well-being dimensions (such as perceived relationships). Specific suggested items are listed in the appendix.

C. School-environment: School-related well-being

Students spend large proportions of their awake time at school. The experiences and relationships at school have an important impact on students' perceived quality of life, and schools might be considered to have roles beyond nurturing academic achievement, including promoting students' health and well-being (Jourdan, Samdal, Diagne, & Carvalho, 2008). A positive school climate, including supportive, caring teachers, is associated not only with higher academic achievement but also with better self-reporting of students' health, well-being and health behaviors (Cohen, McCabe, Michelli, & Pickeral, 2009; Jia et al., 2009), as well as reduced negative effects of school-related stress (Torsheim & Wold, 2001) and more positive students reactions to demands at school (Huebner, Suldo, Smith, & McKnight, 2004). Researchers have called for specialized measures of subjective well-being in school to account for potential differences in students' assessments of well-being at school compared to global well-being perceptions (Huebner, Valois, Paxton, & Drane, 2005). Yet, only few studies have so far explicitly focused on examining students' subjective well-being in the school environment, such as the school subscale of the MSLSS (Huebner, 2001), the QSL (Epstein & McPartland 1976) or the QSL (Karatzias, Power, & Swanson, 2001), or the BASWBSS (Tian et al., 2015).

Two main sub-dimensions are proposed for this framework dimension, namely social connections, and schoolwork. A few additional potential indicators are outlined as well. Most proposed indicators fall into the category of subjective indicators as they concern student perceptions of school life and their school environment rather than objective circumstances. It is proposed to include especially indicators of students' social connections and students' workload and fewer indicators on school infrastructure and security as other indicators might be available for this area (e.g., from school records).

a. Social connections at school

Social connections includes social relationships between teachers and students, with peer students, and general patterns of student interactions and school climate that might foster a sense of belonging to school – the feeling of being accepted, respected, included, and socially supported in the school environment (Goodenow 1993) – or a sense of discrimination and

loneliness at school. Several different instruments focusing on sense of belonging for adolescents have been proposed over the past decades (e.g., Anderman, 2002) and PISA has included a sense of belonging scale in its main student questionnaire for several assessment cycles. Sense of belonging correlates with measures of life satisfaction as well as experienced emotional well-being (e.g., Gilman and Anderman 2006; Millings, Buck, Montgomery, Spears, & Stallard, 2012).

Prior survey research found that teacher–student relationships and classmate support are important predictors of student adjustment and adolescent life satisfaction (Reddy, Rhodes, & Mulhall, 2003; Suldo et al. 2009).

HBSC findings show that students who perceive their school as supportive more often report positive health behaviors and health and well-being outcomes (Ravens-Sieberer et al., 2004; Due, Lynch, Holstein, & Modvig, 2003; Vieno et al., 2007; Freeman et al., 2009). Student who indicate that they like school are less likely to become victims of bullying (Harel-Fisch et al., 2011), take fewer sexual risks (3Dias, Matos, & Goncalves, 2005) and less frequently report smoking, alcohol and drug use (e.g., Fletcher, Bonell, & Hargreaves, 2008). In contrast, disliking school has been related to increased risk of drop out (Archambault, Janosz, Morizot, & Pagani, 2009) and health problems (Shochet et al., 20167).

Bullying, defined as negative physical or verbal actions that have hostile intent, cause distress to victims, are repeated and involve a power differential between perpetrators and victims (Craig, Pepler & Atlas, 2000; Mahady, Wilton, Craig, & Pepler, 2000; Olweus, 1991) has received increasing policy attention in recent years (Farrington and Ttofi, 2009), and several countries have launched bullying prevention programs given the severity of bullying consequences for victimized students. Victims of physical or mental bullying, for example, are more likely to exhibit poor school performance or to drop out of the education system (Moore et al., 2008, Currie et al. 2012; Olweus, 1994; Glew, Fan, Katon, & Rivara, 2008), experience depression, anxiety, and loneliness, as well as a range of psychosomatic symptoms (Olweus, 1991, Craig, 1998, Nansel et al., 2001, Due et al., 2005), and are more likely to abuse drugs and alcohol (Molcho, Harel, & Lash, 2004). Adolescents who have recently been bullied tend to report levels of subjective well-being substantially below population averages and recent reports suggest that the effects of bullying on well-being greatly outweigh the effects of other contextual variables (The Children’s Society, 2015).

Evaluation of school-based bullying prevention programs shows high likelihood of success (HBSC 2010). Results from major well-being and health studies further suggest that reducing and preventing bullying could be strongly linked to improving students’ well-being not only in adolescence but also in adulthood as research findings to date suggest that victimization at school has lasting negative effects into adulthood (Bond et al., 2001, Clapper et al., 1995, Ttofi et al., 2011). It is further recommended to focus not only on physical bullying but also on cyberbullying (i.e., bullying involving modern digital communication technologies), a recommendation that is consistent with the recommendation of the HBSC that “research on the effectiveness of prevention and intervention programs on cyberbullying is therefore strongly encouraged” (Ahlfors, 2010). Research further suggests that not only the victim perspective can yield important data on well-being but that questions focusing on the perpetrator and bystander perspectives could add substantively to a well-being module as well (Rigby & Slee, 1991; Veenstra et al., 2005). Perpetrator behaviors are associated with a range of negative health, social, and academic behaviors (Glew et al, 2008; Nansel et al., 2001, Harel, 1999; Olweus, 2011, Farrington, Loeber, Stallings, & Ttofi, 2011).

- *Objective Indicators*

While there are a larger number of established bullying instruments available (e.g., Olweus, 1996; HBSC), the questions on students' bullying experiences, which were introduced to PISA with the 2015 cycle, may serve as one objective indicator of negative or dysfunctional social relationships and the lack of social integration. These questions are objective in the sense that specific, clearly described and quantifiable behaviors are listed in a checklist and students are asked which of the listed experiences they have made.

- *Subjective Indicators*

In addition to questions on bullying, currently used PISA student questionnaire items on student-teacher relationships, sense of discrimination, and sense of belonging represent key subjective indicators for this module.. Questions about students' sense of belonging and social connectedness at school also address one of the three main basic psychological needs distinguished in self-determination theory (Ryan & Deci, 2000), namely relatedness. Perceived discrimination, on the other hand, can have detrimental effects on student well-being. Meta-analytical evidence points to the pervasiveness of perceived discrimination in lowering student well-being (Schmitt, Branscombe, Postmes, & Garcia, 2014). The questions may further be complemented by a set of event reconstruction type questions focusing on experienced well-being as well as general satisfaction questions. A full list of questions is provided in the appendix.

b. Schoolwork

Students' workload and time spent at school is an important component of well-being and serves as one aspect of the proposed work-life-balance composite index. Research with adult populations shows that well-being and health suffer if individuals live under extreme working conditions. So far, comprehensive findings for adolescents are lacking but it is expected that extreme hours of school might have negative consequences (Karasek & Theorell, 1990). Feeling pressured or stressed by schoolwork may lead to more frequent health-compromising behaviors (such as smoking, drinking alcohol and drunkenness), more frequent health complaints (such as headache, abdominal pain and backache) and experience of psychological problems (such as feeling sad, tense and nervous) (Torsheim & Wold, 2001; Simetin et al., 2011), and overall lower life satisfaction (Ravens-Sieberer et al., 2004). Furthermore, students are likely to show different patterns of well-being in regards to different subjects and classroom activities at school.

- Objective Indicators

As objective indicators of the quality of students' leisure time, inclusion of both the total available time as well as how students use this time is recommended. A proxy of the former can be derived as the difference between awake hours minus hours spent at school, spend on the way to and from school, and spent on homework and studying for school. Most of these variables can be derived from questions currently included in the main student questionnaire. Additional questions in the well-being questionnaire on awake hour and hours spent at school, as well as time spent on the commute to school, will ensure that gaps are filled and sufficient information is available to create an estimate of a student's leisure time. Data on specific leisure time activities a student engages in may be collected via a short day reconstruction protocol focusing on selected key events concerning student's time use, such as watching TV/Video, Reading a book, browsing/reading on the Internet, spending time of chat/social networks/e-mail, playing video-games, meeting friends, talking to parents, eating, or practicing a sport. The particular choice of activities is informed by questions included in an abbreviated time use protocol introduced to PISA 2015 (Bertling & Kyllonen) as well as other studies concerned with student time use (Rees & Main, 2015; Bureau of Labor Statistics, 2015; Carson, Staiano, & Katzmarzyk, 2015; Larson & Verma, 1999).

- Subjective Indicators

Student reports on experienced emotions during selected episodes associated with schoolwork are proposed as indicators of subjective well-being in regards to school work. These indicators are important because students who report frequent positive emotions in school tend to have stronger school engagement, higher academic achievement outcomes, and evaluate their overall well-being more positively (Patrick, Knee, Canevello, & Lonsbary, 2007; Sellstroem & Bremberg 2006). In contrast, students who report frequent negative emotions in school are more likely to withdraw from school, show antisocial behavior, and drug abuse (Roeser, 2001). It is proposed to sample affective states that are especially relevant to the school environment versus states mostly relevant in a clinical context. Specific classes recommended for inclusion in event reconstruction questions are mathematics, language of instruction, and art/creativity classes. These three school subjects were chosen to represent a broad range of contents and classroom practices. In addition, questions about experienced well-being while doing homework or studying for school are recommended for inclusion. Affective states are limited to short set of both positive and negative affective states. A potentially valuable extension of this measurement approach could be achieved by implementing matrix sampling approaches for questionnaires that would allow sampling a larger set of events and affective states (see section 2.E in this framework). The current proposal assumes that no questionnaire matrix sampling will be implemented for the 2018 PISA well-being questionnaire, which is why the number of questions are kept at a minimal number.

c. Other Potential Indicators

Students' safety perceptions at school and on their way to school as well as their satisfaction with their safety and the general infrastructure in the school (e.g., availability of ICT resources for learning and instruction) are relevant facets of student well-being at school as well. Information on these facets component could be drawn from subject-specific survey questions and/or based on school records of reported incidents as well as police/safety statistics of the area of the school. In addition, aggregate measures of prevalence of bullying in the school or other disciplinary problems could serve as evidence fragments for this sub-dimension.

D. Out-of school environment: Well-being outside of school

Students' integration with and experiences in their out-of-school environment constitute the third broad well-being dimension distinguished in the proposed framework. Key sub-dimensions that are distinguished are students' social connections outside of school (including friendships and relationships with parents), a student's material living conditions, and leisure time activities.

a. *Social Connections outside of school*

In addition to students' social connections in the school environment, relationships with parents and other family members as well as friendships are important factors for student overall well-being. Social connections with family and friends are deemed especially important for the adolescent population as research points to the importance of friendships for adolescent well-being (e.g., Ma & Huebner, 2008; Levin & Currie, 2010). Findings from the HBSC survey and other studies support that having high-quality peer relationships has positive effects on adolescent health (Barker & Galambos, 2003; Zambon et al, 2010). Furthermore, having fewer friends in adolescence may be related to a lack in opportunities to learn social skills (Gifford & Brownell, 2003), potentially leading to lowered life satisfaction, more frequent negative affect, and bullying experiences (Larson & Richards, 1991: 8). Findings from qualitative studies also point to the importance of family relationships and friendships as two main factors that determine self-satisfaction (e.g., Edwards & Lopez, 2006; Suldo, Frank, Chappel, 2014). Quantitative research findings indicate that self-reported ease of communication with their parents is associated with a range of positive health outcomes as well as absence of risk behaviors and negative outcomes (Currie et al., 2012). Moreover, children who report talking more frequently to family members about things that mattered to them also tend to report higher levels of subjective well-being (Abdallah, Main, Pople & Rees, 2014). Social connections also include student's sense of and identification with their community (Davidson & Cotter, 1991; Farrell, Aubry, & Coulombe, 2004; Prezza, Amici, Roberti, & Tedeschi, 2001; Prezza & Constantini, 1998).

- *Objective Indicators*

Time spent on activities with friends and parents may serve as objective indicators of student's social connections outside of school. This information can be collected via a short time use/ day reconstruction protocol focusing on selected key events, such as having dinner with the parents and spending time with friends outside the home. As a complement to these questions about a specific day, it is recommended to adapt a short set of questions about leisure time spent with friends that is part of HBSC for use in PISA. Questions inquire about the number of days per week on which students spend time with friends right after school, in the evenings, or communicate via electronic media. In addition, including a question about where a student's come from (i.e., whether friends come from a student's current school, previous school, neighborhood, or family) could add value to this set of objective indicators.

- *Subjective Indicators*

In order to capture students' subjective perceptions about their social connections, and their affect and satisfaction in regards to these relationships, it is proposed to include a series of questions that build up on questions validated as part of the HBSC survey and, as a complement to these, measure experienced well-being in regards to social connections with a short set of event reconstruction questions.

Proposed questions on friendships cover the number of perceived close female and male friends (adapted from HBSC)⁷, their satisfaction with the number of friends they have, the degree to which students felt they had fun with their friends over the past week (part of KidScreen-10), the perceived easiness of talking to one's best friend about things that really bother a student (adapted from HBSC), and students' experienced well-being while spending time outside the home with their friends (newly developed for PISA following the event reconstruction approach). Capturing information beyond the mere number of friends is important as the quality of relationships is a stronger predictor of well-being than the quantity of relationships or friendships (The Children's Society, 2015).

Proposed questions on the subjective quality of relationships with parents, guardians, or other family members include: the degree to which students felt they were treated fairly by their parents over the past week (part of KidScreen-10); the degree to which students think their friends are well-accepted by their parents (adapted from HBSC); the perceived easiness of talking to their father, stepfather, mother, stepmother, elder brother or elder sister about things that really bother a student (adapted from HBSC); students' perceptions of parents' or guardians' general behaviors and attitude towards themselves (adapted from HBSC); and students' experienced well-being while having dinner at home with their parents (newly developed for PISA following the event reconstruction approach).

The full set of proposed questions is provided in the appendix.

b. Material Living Conditions

The material living conditions a student is experiencing as measured by his/her family's socioeconomic status (SES Expert Panel, 2012) constitute an important determinant of overall well-being with small but robust associations between household income and adolescent subjective well-being (Rees, Pople & Goswami, 2011). Children from highly affluent families also rate their health higher (Torsheim et al., 2004, Richter et al., 2009). Although PISA already includes a measure of socioeconomic status, adding additional components especially important for well-being might be beneficial. Specifically, the literature indicates that poverty, and even more so perceived poverty, is a crucial limiting factor for a student's well-being (e.g., Goswami, 2014). A child is deprived if he/she does not have a consumption pattern that follows the generally accepted norm (Bastos & Machado, 2009). Students' basic needs and desires are more likely to be met when they live in rich nations (e.g., Tay & Diener, 2011; Diener, Kahneman, Tov, & Arora, 2010). Data from the 2015 good childhood report indicate that child-reported material deprivation explained larger proportions of variation in children's subjective well-being than overall family socioeconomic status did (The Children's Society, 2015). The authors conclude that "adolescents tend to talk about money and possessions in relative terms – e.g., having 'enough' or 'the same amount' as – rather than 'more' than others so that they fit in and are not excluded from things that others can do" (The Children's Society, 2015). These findings point to the importance of subjective socioeconomic status (Diemer et al., 2013; Quon & McGrath, 2014), a facet of the larger SES construct that has not received as much attention as its objective components.

⁷ Note, while this set of questions might be classified also as objective, it is considered here as part of the subjective indicators as individuals might differ in their perception of friends as "close" and number are, therefore, student perceptions rather than objective indicators (e.g., Keller, 2004). It is proposed to add a short qualifying statement about the definition of "close friends" to the beginning of the question to ensure a minimal degree of interindividual and cross-cultural comparability of the question.

- *Objective Indicators*

PISA measures a student's objective quality of material living conditions through a composite *Index of Economic, Cultural, and Social Status* (ESCS) that is derived from questions about general wealth (based on several proxy variables including home possessions), parental education, and parental occupation. Based on the assumption that this established protocol will be continued, no additional objective indicators for material conditions are proposed under the current framework for the well-being questionnaire. However, I am referencing a few facets of the construct and questions that might be considered valuable additions to a future PISA ESCS index. In an attempt to keep the number of questions on this sensitive topic at a minimum, I suggest exploring inclusion of questions focusing on whether students get pocket money (used in HBSC), whether they have been on a vacation with their family (used in HBSC and Children's Worlds), and whether students had to go to bed hungry (used in HBSC). It is expected that a relatively small set of indicators could add substantial value to the current ESCS indicator. Additional candidate questions that could further inform measurement of poverty and deprivation as risk factors for student well-being are questions around the broader concept of unmet needs (Diemer et al., 2013).

- *Subjective Indicators*

In addition to the objective indicators listed above, which are part of the main PISA student questionnaire, it is recommended to include a short indicator of subjective socioeconomic status that would allow capturing students' subjective perspective of their economic standing in addition to the objective wealth indicators. Subjective SES questions serve as relative poverty measures, in contrast to absolute poverty measures. These measures focus on perceptions of the adequacy of one's standard of living (Conger, Conger, & Martin, 2010; Mistry & Lowe, 2006) as well as psychological experiences of material deprivation and hardship (Gershoff et al., 2007, Iceland, 2003; Mayer & Jencks, 1989). As such, relative poverty measures constitute important factors for subjective well-being, particularly domain satisfaction. Moreover, research on poverty and aspirations failure (e.g., Dalton, Ghosal, & Mani, 2014; Ray, 2006) suggests that poverty and incapability of aspiring to change one's life for the better may lead to underutilization of available resources, and that subjective perceptions might play an equally important or maybe even larger role (e.g., The Children's Society, 2015). These findings underline the importance of paying attention to the subjective "experience" of poverty in addition to objective measures of socioeconomic status. A question that has been used with school-aged children is the question "How well off do you think your family is?". Alternatives to this question would be use of a Cantril ladder type question for subjective SES (Diemer et al., 2013), but this question is longer, more cognitively complex, and has been less widely used with relevant populations. In addition, it is recommended to include a question validated as part of the Children's Worlds Survey about whether students worry about the money their family has. Recent meta-analytical results confirmed that perception of financial constraints is strongly associated with adolescent health outcomes (Quon & McGrath, 2014).

All proposed questions are provided in the appendix.

c. Leisure time

One crucial factor for an individual's capability to flourish is the sheer amount of available leisure time for self-chosen activities. A student's leisure time can be described as the awake time not spent on school or schoolwork, commute to school, or other work students have to engage in. For adults, there is overwhelming evidence that that long and difficult commutes are typically perceived as unpleasant and are associated with reduced subjective well-being (e.g., Kahneman, Krueger, Schkade, Schwarz, & Stone, 2004; Office for National Statistics, 2014a). In order to thrive in their lives, adolescents need to be able to spend time with their friends and family and do activities that they enjoy (e.g., Rees, Goswami, & Bradshaw, 2010; Abdallah et al., 2014). Including indicators of leisure time use and the experienced emotional quality of these experiences therefore constitutes an important element in measuring overall well-being.

- Objective Indicators

As objective indicators of the quality of students' leisure time, inclusion of both the total available time as well as how students use this time is recommended. A proxy of the former can be derived as the difference between awake hours minus hours spent at school, spend on the way to and from school, and spent on homework and studying for school. Most of these variables can be derived from questions currently included in the main student questionnaire. Additional questions in the well-being questionnaire on awake hour and hours spent at school, as well as time spent on the commute to school, will ensure that gaps are filled and sufficient information is available to create an estimate of a student's leisure time. Data on specific leisure time activities a student engages in may be collected via a short day reconstruction protocol focusing on selected key events concerning student's time use, such as watching TV/Video, Reading a book, browsing/reading on the Internet, spending time of chat/social networks/e-mail, playing video-games, meeting friends, talking to parents, eating, or practicing a sport. The particular choice of activities is informed by questions included in an abbreviated time use protocol introduced to PISA 2015 as well as other studies concerned with student time use (e.g., Children's Worlds, 2015; American Time Use Survey). Due to the increased time adolescents spend interacting with electronic media, it is recommended to include social media use as a separate activity in the time use protocol. A consensus from findings to date seems to be that using electronic media up to some extent is associated with mostly positive effects whereas excessive use is associated mostly with negative consequences. Recent research findings suggest that greater use of digital media may lead to also more face-to-face contact with friends (13). Excessive use, on the contrary, may lead to poorer health, sleeping habits, loneliness, and great engagement in risk behaviors (Prezz, Pacilli, & Dinelli, 2004, Punamaeki et al., 2007, Leena, Tomi, & Arja, 2005). In a recent study, Sampasa-Kanyinga & Lewis (2015) found that more than two hours of daily social networking site use was associated with poor self-rating of mental health and experiences of high levels of psychological distress.

The full set of proposed questions is provided in the appendix.

- *Subjective Indicators*

A combination of selected event reconstruction questions and a set of self-report questions focusing on students' perception of and satisfaction with their general time use and leisure time activities in specific is recommended to gauge data on subjective indicators of the quality of students' leisure time.

Event reconstruction questions are recommended with regard to students' experienced well-being during breaks between classes at school, time spent outside of their home with friends, and dinner time spent with their parents. A short set of both positive and negative affective states is recommended to sample key emotions while keeping student burden low. A desired extension of this measurement approach could be achieved by implementing matrix sampling approaches for questionnaires that would allow sampling a larger set of events and affective states (see section 2.E in this framework). The current proposal assumes that no questionnaire matrix sampling will be implemented for the 2018 PISA well-being questionnaire, which is why the number of questions are kept at a minimal number. Additional self-report questions proposed for inclusion include questions on: students' overall satisfaction with how they spend their time and what they do in their free time (part of the larger set of domain satisfaction question); their satisfaction with specific activities engaged in on the previous day (as part of the day reconstruction protocol); and their evaluation of the amount of time students have for themselves (part of the KIDSCREEN-10).

The full set of proposed questions is provided in the appendix.

d. Other Potential Indicators

Students' safety perceptions at home and in their local areas as well as their satisfaction with their safety and the opportunities in their local area are relevant facets of this framework component. Due to overall space constraints and as part of an overall prioritization of themes based on the research literature, a few additional domain satisfaction questions covering these themes are recommended, specifically students' satisfaction with the people who live in their area, how they are listened to by adults in general, and the area they live in in general. Additional information on this framework component might be drawn from other sources, as available, such as records about the local area or geographical region a student is living in.

E. Possible composite indicators

In addition to the proposed indicators representing each cell of the framework, composite indices drawing on multiple cells of the framework might be generated as broader reporting elements of policy interest. Composite indicators of this kind would resemble indicators like the *Index of Economic, Social, and Cultural Status* (ESCS) created in PISA based on main student questionnaire items. As other authors have remarked, aggregating indicators into composite indices bears the risk of increased opaqueness as to which are the most critical areas of well-being (UNICEF 2007). At the same time, a number of previous studies proposed well-being composite indicators that are widely used in applied contexts (e.g., Bradshaw et al., 2007, 2009; Land et al., 2001, 2007; Moore et al., 2008), and creating such indicators in addition to more specific indices may facilitate measuring progress over time and comparisons across sub-groups (Ben-Arieh 2008). The following list briefly describes six potential composite indicators that are seen especially promising for policy and practice.

a. Index of Overall Quality of Life (QoL)

A composite index of overall quality of life could be created based on aggregation across all framework components. Such an index would capture students' overall well-being drawing on both objective and subjective indicators for quality of life as a whole, self-related well-being, well-being at school, and well-being outside of school.

b. Index of Overall Subjective Well-being (SWB)

Creation of a composite index of subjective well-being might be explored based on aggregation of all subjective indicators of quality of life across all content dimensions. Such an index would capture students' overall subjective well-being drawing on subjective indicators for quality of life as a whole, self-related well-being, well-being at school, and well-being outside of school.

c. Index of Overall Emotional Well-being (EWB)

Creation of a composite index of emotional well-being might be explored based on aggregation of subjective indicators of affective well-being across all content dimensions. Such an index would capture students' overall experienced emotional well-being drawing on reported affective states within the past week (part of KIDSCREEN-10) and experienced well-being during selected events relevant to learning in and outside of school and students' leisure time activities.

d. Index of Overall Life Satisfaction (LS)

Creation of a composite index of life satisfaction might be explored based on aggregation of subjective indicators of satisfaction across all content dimensions. Such an index would combine self-reported data on satisfaction with life as a whole and satisfaction across all sampled domains.

e. Index of Work/School-life Balance (WLB)

Creation of a composite index of work/school-life balance might be explored based on data on objective and subjective indicators of school-related workload and leisure time.

f. Index of Overall Social Well-being (SWB)

Creation of a composite index of social well-being might be explored based on aggregation of objective and subjective indicators of the quality of social connections both at school and outside of school.

REFERENCES

- Adams, R. J., Lietz, P., & Berezner, A. (2013). On the use of rotated context questionnaires in conjunction with multilevel item response models. *Large-scale Assessments in Education*, 1(1), 1-27. Adamson (2007)
- Abdallah S, Main G, Pople L & Rees G (2014). Ways to well-being: Exploring the links between children's activities and their subjective well-being. London: The Children's Society
- Adelman, H. S., Taylor, L., & Nelson, P. (1989). Minors' dissatisfaction with their life circumstances. *Child Psychiatry and Human Development*, 20, 135–147
- Ahlfors, R. (2010). Many sources, one theme: Analysis of cyberbullying prevention and intervention websites. *Journal of Social Sciences*, 6(4), 515.
- Almonte, D. E., McCullough, J., Lei, M., & Bertling, J. P. (2014, April). Spiraling of contextual questionnaires in the NAEP TEL pilot assessment. In J. P. Bertling (Chair), *Spiraling contextual questionnaires in educational large-scale assessments*. Symposium conducted at the meeting of the National Council on Measurement in Education, Philadelphia, PA.
- Angelini, V., Cavapozzi, D., Corazzini, L., & Paccagnella, O. (2013). Do Danes and Italians Rate Life Satisfaction in the Same Way? Using Vignettes to Correct for Individual-Specific Scale Biases. *Oxford Bulletin of Economics and Statistics*.
- Anderman, E. M. (2002). School effects on psychological outcomes during adolescence. *Journal of Educational Psychology*, 94(4), 795–809.
- Andrews, F. M., & Robinson, J. P. (1991). Measures of subjective well-being. *Measures of personality and social psychological attitudes*, 1, 61-114.
- Archambault, I., Janosz, M., Morizot, J., & Pagani, L. (2009). Adolescent behavioral, affective, and cognitive engagement in school: Relationship to dropout. *Journal of School Health*, 79(9), 408-415.
- Bandura, A. (1994). *Self-efficacy*. New York, NY: Wiley.
- Barker, E. T., & Galambos, N. L. (2003). Body dissatisfaction of adolescent girls and boys: Risk and resource factors. *The Journal of Early Adolescence*, 23(2), 141-165.
- Bassi, M., Steca, P., Monzani, D., Greco, A., & Delle Fave, A. (2014). Personality and optimal experience in adolescence: Implications for well-being and development. *Journal of Happiness Studies*, 15(4), 829-843.
- Bastos, A., & Machado, C. (2009). Child poverty: a multidimensional measurement. *International Journal of Social Economics*, 36(3), 237-251.
- Beal, D., Rueda-Sabater, E., & Heng, S. L. (2015). *Why well-being should drive growth strategies: The 2015 sustainable economic development assessment*. Boston Consulting Group.
- Ben-Arieh, A. (2008). The Child Indicators Movement: Past, Present, and Future. *Child Indicators Research*, 1(1), 3–16. <http://doi.org/10.1007/s12187-007-9003-1>
- Bertling, J. P. (2014, May). *Improving the contextual questionnaires for the National Assessment of Educational Progress: Plans for NAEP Core contextual modules*. White paper presented at the meeting of the National Assessment Governing Board, Boston, MA.
- Bertling, J. P., Borgonovi, F., & Almonte, D. E. (in press). Psychosocial skills in large-scale assessments: Trends, challenges, and policy implications. In A. Lipnevich, F. Preckel, & R. Roberts (Eds.), *Psychosocial Skills and School Systems*. Berlin: Springer.
- Bertling, J. P. (2015, June). The future of the NAEP survey Questionnaires: Measuring what matters for student achievement in the U.S. Presentation at the 2015 CCSSO-NCSA Conference. San Diego, CA.
- Bertling, J. P. & Kyllonen (2012). Domain-general student attitudes and behaviors: Constructs and Items for PISA 2015 (Module 10). Report submitted to PISA Questionnaire Expert Group.
- Boarini, R., Comola, M., Smith, C., Manchin, R., & de Keulenaer, F. (2012). What makes a better life? The determinants of subjective well-being in OECD countries—Evidence from the Gallup World Poll. OECD Statistics Working Papers, 2012/03. Paris, France: OECD Publishing.

- Boarini, R., Kolev, A., & McGregor, A. (2014). Measuring Well-being and Progress in Countries at Different Stages of Development.
- Bond, L., Carlin, J. B., Thomas, L., Rubin, K., & Patton, G. (2001). Does bullying cause emotional problems? A prospective study of young teenagers. *Bmj*, 323(7311), 480-484.
- Bradshaw, J., Hoelscher, P., & Richardson, D. (2007). An index of child well-being in the European Union 25. *Social Indicators Research*, 80, 133–177.
- Bradshaw, J., Keung, A., Rees, G., & Goswami, H. (2011). Children's subjective well-being: International comparative perspectives. *Children and Youth Services Review*, 33(4), 548–556.
- Brosschot JF. Cognitive-emotional sensitization and somatic health complaints. *Scandinavian Journal of Psychology*, 2002, 43(2):113–121.
- Bureau of Labor Statistics (2015). *American Time Use Survey User's Guide: Understanding ATUS 2003 to 2014*. <http://www.bls.gov/tus/atusersguide.pdf>
- Carson, V., Staiano, A. E., & Katzmarzyk, P. T. (2015). Physical activity, screen time, and sitting among us Adolescents. *Pediatric exercise science*, 27(1), 151-159.
- Campbell, A., Converse, P. E., & Rogers, W. L. (1976). *The quality of American life: Perceptions, evaluations, and satisfactions*. Russell Sage: Nueva York.
- Cantril, H. (1965). *The patterns of human concern*. New Brunswick: Rutgers University Press.
- Casas, F. (2011). Subjective social indicators and child and adolescent well-being. *Child Indicators Research*, 4(4), 555-575.
- Casas, F., Sarriera, J. C., Abs, D., Coenders, G., Alfaro, J., Saforcada, E., & Tonon, G. (2012). Subjective indicators of personal well-being among adolescents. Performance and results for different scales in Latin-language speaking countries: A contribution to the international debate. *Child Indicators Research*, 5(1), 1-28.
- Casas, F., Bălțătescu, S., Bertran, I., González, M., & Hatos, A. (2013). School satisfaction among adolescents: Testing different indicators for its measurement and its relationship with overall life satisfaction and subjective well-being in Romania and Spain. *Social Indicators Research*, 111(3), 665-681.
- Cash, T. F., Morrow, J. A., Hrabosky, J. I., & Perry, A. A. (2004). How has body image changed? A cross-sectional investigation of college women and men from 1983 to 2001. *Journal of consulting and clinical psychology*, 72(6), 1081.
- Clapper, R. L., Buka, S. L., Goldfield, E. C., Lipsitt, L. P., & Tsuang, M. T. (1995). Adolescent problem behaviors as predictors of adult alcohol diagnoses. *Substance Use & Misuse*, 30(5), 507-523.
- Cohen, J., McCabe, L., Michelli, N. M., & Pickeral, T. (2009). School climate: Research, policy, practice, and teacher education. *The Teachers College Record*, 111(1), 180-213.
- Comber, L. C., & Keeves, J. P. (1973). *Science education in nineteen countries: An empirical study*. Stockholm, Sweden: Almqvist & Wiksell.
- Conger, R. D., Conger, K. J., & Martin, M. J. (2010). Socioeconomic status, family processes, and individual development. *Journal of Marriage and Family*, 72(3), 685-704.
- Craig, W. M. (1998). The relationship among bullying, victimization, depression, anxiety, and aggression in elementary school children. *Personality and individual differences*, 24(1), 123-130.
- Craig, W. M., Pepler, D., & Atlas, R. (2000). Observations of bullying in the playground and in the classroom. *School Psychology International*, 21(1), 22-36.
- Csikszentmihalyi, M. (1975). Play and intrinsic rewards. *Journal of humanistic psychology*.
- Csikszentmihalyi, M. E., & Csikszentmihalyi, I. S. E. (2006). *A life worth living: Contributions to positive psychology*. Oxford University Press.
- Cummins R & Lau A (2005) *Personal Wellbeing Index – School Children (Third Edition)*. Melbourne: School of Psychology, Deakin University.
- Currie, C., Nic Gabhainn, S., & Godeau, E. (2009). The health behaviour in school-aged children: WHO collaborative crossnational (HBSC) study: Origins, concept, history and development 1982–2008. *International Journal of Public Health*, 54(Suppl. 2), 131–139. doi:10.1007/s00038-009-5404-x

- Currie, C., et al. (Eds.) (2012). *Social determinants of health and well-being among young people. Health Behaviour in School-aged Children (HBSC) study: international report from the 2009/2010 survey*. World Health Organization Regional Office for Europe Copenhagen
- Dalton, P. S., Ghosal, S., & Mani, A. (2014). Poverty and aspirations failure. *The Economic Journal*.
- Davidson, W. B., & Cotter, P. R. (1991). The relationship between sense of community and subjective well-being: A first look. *Journal of Community Psychology, 19*(3), 246-253.
- Dias SF, Matos MG, Goncalves AC. Preventing HIV transmission in adolescents: an analysis of the Portuguese data from the Health Behaviour Schoolaged Children study and focus groups. *European Journal of Public Health, 2005, 15*:200–204.
- Diemer, M. A., Mistry, R. S., Wadsworth, M. E., López, I., & Reimers, F. (2013). Best practices in conceptualizing and measuring social class in psychological research. *Analyses of Social Issues and Public Policy, 13*(1), 77-113.
- Diener E. Subjective wellbeing: three decades of progress. *Psychological Bulletin, 1999, 125*:276–301
- Diener, E., & Chan, M. Y. (2011). Happy people live longer: Subjective well-being contributes to health and longevity. *Applied Psychology: Health and Well-Being, 3*(1), 1-43.
- Diener, E., Lucas, R. E., & Oishi, S. (2005). Handbook of positive psychology. *Handbook of positive psychology*.
- Diener, E., Oishi, S., & Lucas, R. E. (2015). National accounts of subjective well-being. *American Psychologist, 70*(3), 234.
- Diener, E., Seligman, M.E.P., 2004. Beyond money: Toward an economy of well-being. *Psychological Science in the Public Interest 5*, 1–31.
- Diener, E., Kahneman, D., Tov, W., & Arora, R. (2010). Income's association with judgments of life versus feelings. *International differences in well-being, 3*-15.
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: the tailored design method*. John Wiley & Sons.
- Dolan, P., Peasgood, T., & White, M. (2008). Do we really know what makes us happy? A review of the economic literature on the factors associated with subjective well-being. *Journal of economic psychology, 29*(1), 94-122.
- Due, P., Lynch, J., Holstein, B., & Modvig, J. (2003). Socioeconomic health inequalities among a nationally representative sample of Danish adolescents: the role of different types of social relations. *Journal of Epidemiology and Community Health, 57*(9), 692-698.
- Due, P., Holstein, B. E., Lynch, J., Diderichsen, F., Gabhain, S. N., Scheidt, P., & Currie, C. (2005). Bullying and symptoms among school-aged children: international comparative cross sectional study in 28 countries. *The European Journal of Public Health, 15*(2), 128-132.
- Dupuy, M., Godeau, E., Vignes, C., & Ahluwalia, N. (2011). Socio-demographic and lifestyle factors associated with overweight in a representative sample of 11-15 year olds in France: results from the WHO-Collaborative Health Behaviour in School-aged Children (HBSC) cross-sectional study. *BMC Public Health, 11*(1), 442.
- Edwards, L. M., & Lopez, S. J. (2006). Perceived family support, acculturation, and life satisfaction in mexican american youth: A mixed-methods exploration. *Journal of Counseling Psychology, 53*(3), 279.
- Elmore, G. M., & Huebner, E. S. (2010). Adolescents' satisfaction with school experiences: Relationships with demographics, attachment relationships, and school engagement behavior. *Psychology in the Schools, 47*(6), 525-537.
- Epstein, J. L., & McPartland, J. M. (1976). The concept and measurement of the quality of school life. *American Educational Research Journal, 13*(1), 15-30.
- Evans J., Macrory, I. & Randall, C. (2015) *Measuring National Well-being: Life in the UK, 2015*. London: Office for National Statistics
- Farrell, S. J., Aubry, T., & Coulombe, D. (2004). Neighborhoods and neighbors: Do they contribute to personal well-being?. *Journal of community psychology, 32*(1), 9-25.
- Farrington, D. P., Loeber, R., Stallings, R., & Ttofi, M. M. (2011). Bullying perpetration and victimization as predictors of delinquency and depression in the Pittsburgh Youth Study. *Journal of Aggression, conflict and peace research, 3*(2), 74-81..

- Federal Interagency Forum on Child and Family Statistics (2009) *America's Children: Key National Indicators of Well-being*. Washington, DC: US Government Printing Office. http://www.childstats.gov/pdf/ac2009/ac_09.pdf
- Fernandes, L., Mendes, A., & Teixeira, A. A. (2012). A review essay on the measurement of child well-being. *Social Indicators Research*, 106(2), 239-257.
- Fletcher, A., Bonell, C., & Hargreaves, J. (2008). School effects on young people's drug use: a systematic review of intervention and observational studies. *Journal of Adolescent Health*, 42(3), 209-220.
- Foundation for Child Development (2012). National Child and Youth Well-being Index (CWI).
- Freeman, J. G., Samdal, O., Klinger, D. A., Dur, W., Griebler, R., Currie, D., & Rasmussen, M. (2009). The relationship of schools to emotional health and bullying. *International journal of public health*, 54(2), 251-259.
- Garrow, J.S. & Webster, J., 1985. Quetelet's index (W/H²) as a measure of fatness. *Int. J. Obes.*, 9(2), pp.147–153.
- Gaspar, M. J. M., Amaral, T. F., Oliveira, B. M., & Borges, N. (2011). Protective effect of physical activity on dissatisfaction with body image in children—A cross-sectional study. *Psychology of Sport and Exercise*, 12(5), 563-569.
- Gehlbach, H. (2015). Seven survey sins. *The Journal of Early Adolescence*, 0272431615578276. Gershoff et al., 2007
- Gifford-Smith, M. E., & Brownell, C. A. (2003). Childhood peer relationships: Social acceptance, friendships, and peer networks. *Journal of School Psychology*, 41(4), 235-284.
- Gilman, R., & Huebner, E. S. (2000). Review of life satisfaction measures for adolescents. *Behaviour Change*, 17(03), 178-195.
- Gilman, R., & Huebner, S. (2003). A review of life satisfaction research with children and adolescents. *School Psychology Quarterly*, 18(2), 192.
- Gilman, R., & Anderman, E. M. (2006). The relationship between relative levels of motivation and intrapersonal, interpersonal, and academic functioning among older adolescents. *Journal of School Psychology*, 44(5), 375-391.
- Glew, G. M., Fan, M. Y., Katon, W., & Rivara, F. P. (2008). Bullying and school safety. *The Journal of pediatrics*, 152(1), 123-128.
- Goodenow, C. (1993). The psychological sense of school membership among adolescents: Scale development and educational correlates. *Psychology in the Schools*, 30(1), 79-90.
- Goswami (2014) Children's Subjective Well-being: Socio-demographic Characteristics and Personality. Volume 7, Issue 1, pp. 119-140
- Grube, A., Schroer, J., Hentzschel, C., & Hertel, G. (2008). The event reconstruction method: An efficient measure of experience-based job satisfaction. *Journal of Occupational and Organizational Psychology*, 81(4), 669-689
- Hallal, P. C., Victora, C. G., Azevedo, M. R., & Wells, J. C. (2006). Adolescent physical activity and health. *Sports Medicine*, 36(12), 1019-1030.
- Harel-Fisch, Y., Walsh, S. D., Fogel-Grinvald, H., Amitai, G., Pickett, W., Molcho, M., ... & Craig, W. (2011). Negative school perceptions and involvement in school bullying: A universal relationship across 40 countries. *Journal of adolescence*, 34(4), 639-652.
- Harel, Y. (1999). A cross-national study of youth violence in Europe. *International journal of adolescent medicine and health*, 11(3-4), 121-134.
- Harandin, E. C., Huebner, E. S., & Suldo, S. M. (2007). Predictive and incremental validity of global and domain-based adolescent life satisfaction reports. *Journal of Psychoeducational Assessment*, 25(2), 127-138.
- Hardy, C. L., Bukowski, W. M., & Sippola, L. K. (2002). Stability and change in peer relationships during the transition to middle-level school. *Journal of Early Adolescence*, 22, 117–142. doi:10.1177/02724316020220 02001
- Haug, E., Rasmussen, M., Samdal, O., Iannotti, R., Kelly, C., Borraccino, A., ... & Ahluwalia, N. (2009). Overweight in school-aged children and its relationship with demographic and lifestyle factors: results from the WHO-Collaborative Health Behaviour in School-aged Children (HBSC) study. *International Journal of Public Health*, 54(2), 167-179.
- Helliwell J, Layard R, and Sachs J (eds.) 2015, World Happiness Report. The Earth Institute, Columbia University.
- Helseth, S., & Misvær, N. (2010). Adolescents' perceptions of quality of life: what it is and what matters. *Journal of clinical nursing*, 19(9-10), 1454-1461.

- Hetland, J., Torsheim, T., & Aarø, L. E. (2002). Subjective health complaints in adolescence: dimensional structure and variation across gender and age. *Scandinavian Journal of Public Health, 30*(3), 223-230.
- Huebner, E. S., Valois, R. F., Paxton, R., & Drane, W. (2005). Middle school students' perceptions of quality of life. *Journal of Happiness Studies, 6*(1), 15–24. doi:10.1007/s10902-004-1170-x
- Huebner, S. (2001). Multidimensional students' life satisfaction scale. *University of South Carolina, Department of Psychology, Columbia, SC, 29208*.
- Huebner, E. S., Suldo, S. M., Smith, L. C., & McKnight, C. G. (2004). Life satisfaction in children and youth: empirical foundations and implications for school psychologists [Special issue]. *Psychology in the Schools, 41*, 81–93
- Huebner, E. S., Valois, R. F., Suldo, S. M., Smith, L. C., McKnight, C. G., Seligson, J. L., & Zullig, K. J. (2004). Perceived quality of life: A neglected component of adolescent health assessment and intervention. *Journal of Adolescent Health, 34*(4), 270-278.
- Huebner, E. S., & Dew, T. (1996). The interrelationships of positive affect, negative affect, and life satisfaction in an adolescent sample. *Social Indicators Research, 38*(2), 129-137.
- Huppert, F. A., Marks, N., Clark, A., Siegrist, J., Stutzer, A., Vittersø, J., & Wahrendorf, M. (2009). Measuring well-being across Europe: Description of the ESS well-being module and preliminary findings. *Social Indicators Research, 91*(3), 301-315.
- Iannotti, R. J., Janssen, I., Haug, E., Kololo, H., Annaheim, B., & Borraccino, A. (2009). HBSC Physical Activity Focus Group: Interrelationships of adolescent physical activity, screen-based sedentary behaviour, and social and psychological health. *Int J Public Health, 54*(Suppl 2), 191-198.
- Iceland, J. (2003). Why poverty remains high: The role of income growth, economic inequality, and changes in family structure, 1949–1999. *Demography, 40*(3), 499-519.
- James W.P.T., Chen C., & Inoue S. (2002). *Appropriate Asian body mass indices? Obesity Review, 3*, 139.
- Janssen, I., Craig, W. M., Boyce, W. F., & Pickett, W. (2004). Associations between overweight and obesity with bullying behaviors in school-aged children. *Pediatrics, 113*(5), 1187-1194.
- Jourdan, D., Samdal, O., Diagne, F., & Carvalho, G. S. (2008). The future of health promotion in schools goes through the strengthening of teacher training at a global level. *Promotion & education, 15*(3), 36-38.
- Jia, Y., Way, N., Ling, G., Yoshikawa, H., Chen, X., Hughes, D., ... & Lu, Z. (2009). The influence of student perceptions of school climate on socioemotional and academic adjustment: A comparison of Chinese and American adolescents. *Child development, 80*(5), 1514-1530.
- Kahneman, D., Krueger, A. B., Schkade, D. A., Schwarz, N., & Stone, A. A. (2004). A survey method for characterizing daily life experience: The day reconstruction method. *Science, 306*, 1776–1780.
- Kahneman, D. & Krueger, A.B. (2006), Developments in the Measurement of Subjective Well-Being, *Journal of Economic Perspectives*, Vol. 20(1), pp. 19-20.
- Kaplan, D., & Wu, D. (2014, April). Imputation issues relevant to context questionnaire rotation. In J. P. Bertling (Chair), *Spiraling contextual questionnaires in educational large-scale assessments*. Symposium conducted at the meeting of the National Council on Measurement in Education, Philadelphia, PA.
- Karatzias, A., Power, K. G., & Swanson, V. (2001). Quality of school life: Development and preliminary standardisation of an instrument based on performance indicators in Scottish secondary schools. *School Effectiveness and School Improvement, 12*(3), 265–284.
- Kristensen, N., & Johansson, E. (2008). New evidence on cross-country differences in job satisfaction using anchoring vignettes. *Labour Economics, 15*, 96–117.
- Krosnick, J. A., & Fabrigar, L. R. (1997). Designing rating scales for effective measurement in surveys. *Survey measurement and process quality, 141-164*.
- Karasek, R., & Theorell, T. (1992). *Healthy work: stress, productivity, and the reconstruction of working life*. Basic books.
- Kaufman, A. R., & Augustson, E. M. (2008). Predictors of regular cigarette smoking among adolescent females: does body image matter?. *Nicotine & tobacco research: official journal of the Society for Research on Nicotine and Tobacco, 10*(8), 1301.

- Keller, M. (2004). A cross-cultural perspective on friendship research. *Newsletter of the International Society for the Study of Behavioral Development*, 28, 10–14
- Kern, M. L., Waters, L. E., Adler, A., & White, M. A. (2015). A multidimensional approach to measuring well-being in students: Application of the PERMA framework. *The journal of positive psychology*, 10(3), 262-271.
- Keyes, C. L., & Grzywacz, J. G. (2005). Health as a complete state: The added value in work performance and healthcare costs. *Journal of Occupational and Environmental Medicine*, 47(5), 523-532.
- Kim, H., Schimmack, U., & Oishi, S. (2012). Cultural differences in self-and other-evaluations and well-being: A study of European and Asian Canadians. *Journal of Personality and Social Psychology*, 102(4), 856–873. <http://doi.org/10.1037/a0026803>
- Kvalem, I. L., von Soest, T., Træen, B., & Singaas, K. (2011). Body evaluation and coital onset: A population-based longitudinal study. *Body image*, 8(2), 110-118.
- Kyllonen, P. C., & Bertling, J. P. (2013). Innovative questionnaire assessment methods to increase cross-country comparability. In L. Rutkowski, M. von Davier, & D. Rutkowski, (Eds.) *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis* (pp. 277-286). Boca Raton, FL: CRC Press.
- Kyllonen, P. C., & Bertling, J. P. (2014). *Anchoring Vignettes Reduce Bias in Noncognitive Rating Scale Responses*. Report submitted to OECD.
- Land, K. C., Lamb, V. L., & Mustillo, S. K. (2001). Child and youth well-being in the United States, 1975–1998: Some findings from a new index. *Social indicators research*, 56(3), 241-318.
- Land, K.C., V.L. Lamb, S.O. Meadows and A. Taylor (2007). “Measuring trends in child well-being: an evidence-based approach.” *Social Indicators Research*, 80: 105-132.
- Larson, R., & Richards, M. H. (1991). Daily companionship in late childhood and early adolescence: Changing developmental contexts. *Child development*, 62(2), 284-300.
- Larson, R. W., & Verma, S. (1999). How children and adolescents spend time across the world: work, play, and developmental opportunities. *Psychological bulletin*, 125(6), 701.
- Laurent, J., Catanzaro, S. J., Joiner Jr, T. E., Rudolph, K. D., Potter, K. I., Lambert, S., ... & Gathright, T. (1999). A measure of positive and negative affect for children: scale development and preliminary validation. *Psychological assessment*, 11(3), 326.
- Leena, K., Tomi, L., & Arja, R. (2005). Intensity of mobile phone use and health compromising behaviours—how is information and communication technology connected to health-related lifestyle in adolescence?. *Journal of adolescence*, 28(1), 35-47.
- Levin, K. A., & Currie, C. (2010). Family structure, mother-child communication, father-child communication, and adolescent life satisfaction: A cross-sectional multilevel analysis. *Health Education*, 110(3), 152-168.
- Lippman, L. H., Moore, K. A., & McIntosh, H. (2011). Positive indicators of child well-being: a conceptual framework, measures, and methodological issues. *Applied Research in Quality of Life*, 6(4), 425-449.
- Lobstein, T., Baur, L., & Uauy, R. (2004). Obesity in children and young people: a crisis in public health. *Obesity reviews*, 5(s1), 4-85.
- Long, R. F., Huebner, E. S., Wedell, D. H., & Hills, K. J. (2012). Measuring school-related subjective wellbeing in adolescents. *American Journal of Orthopsychiatry*, 82(1), 50–60.
- Ma, C. Q., & Huebner, E. S. (2008). Attachment relationships and adolescents life satisfaction: Some relationships matter more to girls than boys. *Psychology in the Schools*, 45(2), 177.
- Mahady Wilton, M. M., Craig, W. M., & Pepler, D. J. (2000). Emotional regulation and display in classroom victims of bullying: Characteristic expressions of affect, coping styles and relevant contextual factors. *Social development*, 9(2), 226-245.
- Malina, R. M. (1991). Fitness and performance: adult health and the culture of youth. *New possibilities, new paradigms*, 24, 30-38.
- Marcotte, D., Fortin, L., Potvin, P., & Papillon, M. (2002). Gender differences in depressive symptoms during adolescence role of gender-typed characteristics, self-esteem, body image, stressful life events, and pubertal status. *Journal of Emotional and Behavioral Disorders*, 10(1), 29-42.

- Martínez-Gómez, D., Ruiz, J. R., Gómez-Martínez, S., Chillón, P., Rey-López, J. P., Díaz, L. E., ... & Marcos, A. (2011). Active commuting to school and cognitive performance in adolescents: the AVENA study. *Archives of pediatrics & adolescent medicine*, 165(4), 300-305.
- Mayer, S. E., & Jencks, C. (1989). Growing up in poor neighborhoods: How much does it matter?. *Science*, 243(4897), 1441-1445.
- McGill, R. K., Hughes, D., Alicea, S., & Way, N. (2012). Academic adjustment across middle school: The role of public regard and parenting. *Developmental Psychology*, 48, 1003–1018. doi:10.1037/a0026006
- McMurray, R. G., Harrell, J. S., Creighton, D., Wang, Z., & Bangdiwala, S. I. (2008). Influence of physical activity on change in weight status as children become adolescents. *International Journal of Pediatric Obesity*, 3(2), 69-77.
- Mistry, R. S., & Lowe, E. (2006). What earnings and income buy: The “basics”, plus “a little extra”: Implications for family and child well-being. *Lowe (Eds.), Making it work: Low-wage employment, family life, and child development*, 173-205.
- Millings, A., Buck, R., Montgomery, A., Spears, M., & Stallard, P. (2012). School connectedness, peer attachment, and self-esteem as predictors of adolescent depression. *Journal of adolescence*, 35(4), 1061-1067.
- Molcho, M., Harel, Y., & Lash, D. (2004). The co-morbidity of substance use and youth violence among Israeli school children. *International Journal of Adolescent Medicine and Health*, 16(3), 223-251.
- Monseur, C., & Bertling, J. P. (2014, April). Questionnaire rotation in international surveys: Findings from PISA. In J. P. Bertling (Chair). *Spiraling contextual questionnaires in educational large-scale assessments*. Symposium conducted at the meeting of the National Council on Measurement in Education, Philadelphia, PA.
- Morgan, A., Currie, C., Due, P., Gabhain, S. N., Rasmussen, M., Samdal, O., & Smith, R. (2008). Mental well-being in school-aged children in Europe: associations with social cohesion and socioeconomic circumstances. *Social cohesion for mental well-being among adolescents*.
- Moore, K. A., Theokas, C., Lippman, L., Bloch, M., Vandivere, S., & O'Hare, W. (2008). A microdata child well-being index: Conceptualization, creation, and findings. *Child Indicators Research*, 1(1), 17-50.
- Must, A., & Tybor, D. J. (2005). Physical activity and sedentary behavior: a review of longitudinal studies of weight and adiposity in youth. *International journal of obesity*, 29, S84-S96.
- Nansel, T. R., Overpeck, M., Pilla, R. S., Ruan, W. J., Simons-Morton, B., & Scheidt, P. (2001). Bullying behaviors among US youth: Prevalence and association with psychosocial adjustment. *Jama*, 285(16), 2094-2100.
- Natsuaki, M. N., Biehl, M. C., & Ge, X. (2009). Trajectories of depressed mood from early adolescence to young adulthood: The effects of pubertal timing and adolescent dating. *Journal of Research on Adolescence*, 19, 47–74.
- Neumark-Sztainer, D., Wall, M., Story, M., & Standish, A. R. (2012). Dieting and unhealthy weight control behaviors during adolescence: associations with 10-year changes in body mass index. *Journal of Adolescent Health*, 50(1), 80-86.
- Nihiser AJ, Lee SM, Wechsler H, McKenna M, Odom E, Reinold C, Thompson D, Grummer-Strawn L. Body Mass Index Measurement in Schools. *Journal of School Health*. 2007; 77:651-671
- OECD (2013a), *OECD Guidelines on Measuring Subjective Well-being*, OECD Publishing. <http://dx.doi.org/10.1787/9789264191655-en>
- OECD. (2013b). *PISA 2015: Draft questionnaire framework*. Paris, France: OECD Publishing. Retrieved from <http://www.oecd.org/pisa/pisaproducts/PISA-2015-draft-questionnaire-framework.pdf>
- OECD (2015). *How's Life? 2015: Measuring Well-being*. OECD Publishing, Paris.
- Ojala, K., Vereecken, C., Välimaa, R., Currie, C., Villberg, J., Tynjälä, J., & Kannas, L. (2007). Attempts to lose weight among overweight and non-overweight adolescents: a cross-national survey. *International Journal of Behavioral Nutrition and Physical Activity*, 4(1), 50.
- Olweus D. Bully/victim problems among school children: some basic facts and effects of a school-based intervention program. In: Pepler D, Rubin K eds. *The development and treatment of childhood aggression*. Hillsdale, NJ, Erlbaum, 1991:411–448. Olweus, 1994
- Olweus, D. (1996). Bully/Victim Problems at School: Facts and Effective Intervention. *Reclaiming Children and Youth: Journal of Emotional and Behavioral Problems*, 5(1), 15-22.

- Olweus, D. (2011). Bullying at school and later criminality: Findings from three Swedish community samples of males. *Criminal behaviour and mental health*, 21(2), 151-156.
- Patrick, H., Knee, C. R., Canevello, A., & Lonsbary, C. (2007). The role of need fulfillment in relationship functioning and well-being: a self-determination theory perspective. *Journal of personality and social psychology*, 92(3), 434.
- Petersen, A. C., Leffert, N., Graham, B., & Alwin, J. (1997). Promoting mental health during the transition into adolescence.
- Pollard, E. L., & Lee, P. D. (2003). 'Child Well-being: A Systematic Review of the Literature'. *Social Indicators Research*, 61, 1, 59-78.
- Prezza, M., Pacilli, & Dinelli, 2004—replace Prezza, M., Pacilli, M. G., & Dinelli, S. (2004). Loneliness and new technologies in a group of Roman adolescents. *Computers in Human Behavior*, 20(5), 691-709.
- Prezza, M., & Costantini, S. (1998). Sense of community and life satisfaction: Investigation in three different territorial contexts. *Journal of Community & Applied Social Psychology*, 8(3), 181-194.
- Prezza, M., Amici, M., Roberti, T., & Tedeschi, G. (2001). Sense of community referred to the whole town: Its relations with neighboring, loneliness, life satisfaction, and area of residence. *Journal of community psychology*, 29(1), 29-52.
- Punamäki, R. L., Wallenius, M., Nygård, C. H., Saarni, L., & Rimpelä, A. (2007). Use of information and communication technology (ICT) and perceived health in adolescence: The role of sleeping habits and waking-time tiredness. *Journal of adolescence*, 30(4), 569-585.
- Quon, E. C., & McGrath, J. J. (2014). Subjective socioeconomic status and adolescent health: A meta-analysis. *Health Psychology*, 33(5), 433.
- Ravens-Sieberer, U., Erhart, M., Torsheim, T., Hetland, J., Freeman, J., Danielson, M., & Thomas, C. (2008). An international scoring system for self-reported health complaints in adolescents. *The European Journal of Public Health*, 18(3), 294-299.
- Ravens-Sieberer, U., Erhart, M., Rajmil, L., Herdman, M., Auquier, P., Bruil, J., et al. (2010). Reliability, construct and criterion validity of the KIDSCREEN-10 score: A short measure for children and adolescents' well-being and health-related quality of life. *Quality of Life Research*, 19(10), 1487–1500.
- Ravens-Sieberer, U., Herdman, M., Devine, J., Otto, C., Bullinger, M., Rose, M., & Klasen, F. (2014). The European KIDSCREEN approach to measure quality of life and well-being in children: development, current application, and future advances. *Quality of life research*, 23(3), 791-803.
- Ravens-Sieberer U, Kokonyei G, Thomas C. School and health. In: In Currie C et al., eds. Young people's health in context. Health Behaviour in Schoolaged Children study: international report from the 2001/2002 survey. Copenhagen, WHO Regional Office for Europe, 2004 (Health Policy for Children and Adolescents, No.4).
- Ray, D. (2006). Aspirations, poverty, and economic change. *Understanding poverty*, 409-421.
- Rees G, Pople L & Goswami H (2011) Links between family economic factors and children's subjective wellbeing: Initial findings from Wave 2 and Wave 3 quarterly surveys. London: The Children's Society
- Reddy, R., Rhodes, J. E., & Mulhall, P. (2003). The influence of teacher support on student adjustment in the middle school years: A latent growth curve study. *Development and psychopathology*, 15(01), 119-138.
- Rees G, Goswami H & Bradshaw J (2010) Developing an Index of Children's Subjective Well-Being in England. London: The Children's Society
- Rees, G. & Main, G. (eds) (2015) Children's views on their lives and well-being in 15 countries: An initial report on the Children's Worlds survey, 2013-14. York, UK: Children's Worlds Project (ISCWeB).
- Richter, M., Erhart, M., Vereecken, C. A., Zambon, A., Boyce, W., & Gabhainn, S. N. (2009). The role of behavioural factors in explaining socio-economic differences in adolescent health: a multilevel study in 33 countries. *Social science & medicine*, 69(3), 396-403.
- Rigby, K & Slee, P.T. (1991). Bullying among Australian school children: Reported behaviour and attitudes toward victims. *Journal of Social Psychology*, 131, 615-627.
- Rokholm, B., Baker, J. L., & Sørensen, T. I. A. (2010). The levelling off of the obesity epidemic since the year 1999—a review of evidence and perspectives. *Obesity reviews*, 11(12), 835-846.

- Rudd, N. A., & Lennon, S. J. (2000). Body image and appearance-management behaviors in college women. *Clothing and Textiles Research Journal*, 18(3), 152-162.
- Russell, J. (1980). A Circumplex Model of Affect, *Journal of Personality and Social Psychology*, Vol. 39(6), 1161-1178.
- Ryan RM, Deci EL. Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being. *The American Psychologist*, 2000, 55(1):68–78.
- Ryan, R. M., & Deci, E. L. (2001). On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual review of psychology*, 52(1), 141-166.
- Ryff, C. D. (1995). Psychological well-being in adult life. *Current directions in psychological science*, 99-104.
- Sacks, D. W., Stevenson, B., & Wolfers, J. (2010). *Subjective well-being, income, economic development and growth* (No. w16441). National Bureau of Economic Research.
- Saha, R., Huebner, E. S., Suldo, S. M., & Valois, R. F. (2010). A longitudinal study of adolescent life satisfaction and parenting. *Child Indicators Research*, 3(2), 149–165.
- Salomon, J. A., Tandon, A., & Murray, C. J. (2004). Comparability of self rated health: cross sectional multi-country survey using anchoring vignettes. *Bmj*, 328(7434), 258.
- Sampasa-Kanyinga, H., & Lewis, R. F. (2015). Frequent use of social networking sites is associated with poor psychological functioning among children and adolescents. *Cyberpsychology, Behavior, and Social Networking*, 18(7), 380-385.
- Schleicher, A. (2015, March). *Proposals for PISA 2018 international questionnaire options*. Report presented at 39th meeting of the PISA Governing Board. OECD.
- Schmitt, M. T., Branscombe, N. R., Postmes, T., & Garcia, A. (2014). The consequences of perceived discrimination for psychological well-being: A meta-analytic review. *Psychological Bulletin*, 140(4), 921.
- Schwarz, N., Kahneman, D., Xu, J., Belli, R., Stafford, F., & Alwin, D. (2009). Global and episodic reports of hedonic experience. *Using calendar and diary methods in life events research*, 157-174.
- Seligman, M. E. (2012). *Flourish: A visionary new understanding of happiness and well-being*. Simon and Schuster.
- Sellström, E., & Bremberg, S. (2006). Review Article: The significance of neighbourhood context to child and adolescent health and well-being: A systematic review of multilevel studies. *Scandinavian journal of public health*, 34(5), 544-554.
- SES Expert Panel. (2012). *Improving the measurement of socioeconomic status for the National Assessment of Educational Progress: A theoretical foundation*. White Paper prepared for the National Center for Education Statistics. Washington, DC: U.S. Department of Education.
- Shochet IM et al. School connectedness is an underemphasized parameter in adolescent mental health: results of a community prediction study. *Journal of Clinical Child and Adolescent Psychology*, 2006, 35(2):170–179.
- Sibley, B. A., & Etnier, J. L. (2003). The relationship between physical activity and cognition in children: a meta-analysis. *Pediatric Exercise Science*, 15(3), 243-256.
- Statham, J. & Chase, E. (2010). Childhood Wellbeing: A brief overview. Childhood Wellbeing Research Centre.
- Stiglitz, J., Sen, A. & Fitoussi, J.-P. (2009). Report of the commission on the measurement of economic performance and social progress. http://www.stiglitz-sen-fitoussi.fr/documents/rapport_anglais.pdf.
- Strong WB et al. Evidence based physical activity for school-age youth. *The Journal of Pediatrics*, 2005, 146(6):732–737.
- Suldo, S. M., & Huebner, E. S. (2006). Is extremely high life satisfaction during adolescence advantageous?. *Social Indicators Research*, 78(2), 179-203.
- Suldo, S. M., Riley, K. N., & Shaffer, E. J. (2006). Academic correlates of children and adolescents' life satisfaction. *School Psychology International*, 27(5), 567-582.
- Suldo, S. M., Frank, M. J., Chappel, A. M., Albers, M. M., & Bateman, L. P. (2014). American high school students' perceptions of determinants of life satisfaction. *Social Indicators Research*, 118(2), 485-514.
- Suldo, S. M., Friedrich, A. A., White, T., Farmer, J., Minch, D., Michalowski, J., et al. (2009). Teacher support and adolescents' subjective well-being: A mixed-methods investigation. *School Psychology Review*, 38(1), 67–85.
- Tay, L., & Diener, E. (2011). Needs and subjective well-being around the world. *Journal of personality and social psychology*, 101(2), 354.

- The Children's Society (2015). The Good Childhood Report 2015. <https://www.childrenssociety.org.uk/sites/default/files/TheGoodChildhoodReport2015.pdf>
- Torsheim, T., & Wold, B. (2001). School-related stress, support, and subjective health complaints among early adolescents: a multilevel approach. *Journal of adolescence*, 24(6), 701-713.
- Torsheim, T., Currie, C., Boyce, W., Kalnins, I., Overpeck, M., & Haugland, S. (2004). Material deprivation and self-rated health: a multilevel study of adolescents from 22 European and North American countries. *Social science & medicine*, 59(1), 1-12.
- Tian, L., Wang, D., & Huebner, E. S. (2015). Development and Validation of the Brief Adolescents' Subjective Well-Being in School Scale (BASWBSS). *Social Indicators Research*, 120(2), 615-634.
- The Children's Society (2006) Good Childhood? A Question for Our Times. London: The Children's Society
- Topp, C. W., Østergaard, S. D., Søndergaard, S., & Bech, P. (2015). The WHO-5 Well-Being Index: A Systematic Review of the Literature. *Psychotherapy and psychosomatics*, 84(3), 167-176.
- Ttofi, M. M., Farrington, D. P., Lösel, F., & Loeber, R. (2011). Do the victims of school bullies tend to become depressed later in life? A systematic review and meta-analysis of longitudinal studies. *Journal of Aggression, Conflict and Peace Research*, 3(2), 63-73.
- UNICEF. (2007). Child poverty in perspective: An overview of child well-being in rich countries, Innocenti Report Card 7. Florence: UNICEF.
- UNICEF Spain (2012). *Children's well-being from their own point of view. What affects the children's wellbeing in the first year of Compulsory Secondary Education in Spain?* Madrid: UNICEF Spain
- Van Soest, A., Delaney, L., Harmon, C., Kapteyn, A., & Smith, J. P. (2011). Validating the use of anchoring vignettes for the correction of response scale differences in subjective questions. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, 174(3), 575-595.
- Veenstra, R., Lindenberg, S., Oldehinkel, A. J., De Winter, A. F., Verhulst, F. C., & Ormel, J. (2005). Bullying and victimization in elementary schools: A comparison of bullies, victims, bully/victims, and uninvolved preadolescents. *Developmental Psychology*, 41(4), 672-682.
- Verplanken, B., & Velsvik, R. (2008). Habitual negative body image thinking as psychological risk factor in adolescents. *Body Image*, 5(2), 133-140. Vieno et al., 2007
- Vingilis, E. R., Wade, T. J., & Seeley, J. S. (2002). Predictors of adolescent self-rated health: analysis of the National Population Health Survey. *Canadian Journal of Public Health/Revue Canadienne de Sante'e Publique*, 193-197.
- von Davier, M. (2014). Imputing proficiency data under planned missingness in population models. In L. Rutkowski, M. von Davier, & D. Rutkowski (Eds.), *Handbook of international large-scale assessment: Background, technical issues, and methods of data analysis*. Boca Raton, FL: CRC Press.
- Wang, M. T., & Eccles, J. S. (2012). Adolescent behavioral, emotional, and cognitive engagement trajectories in school and their differential relations to educational success. *Journal of Research on Adolescence*, 22(1), 31-39.
- Watson, D., Clark, L. A., & Tellegen, A. (1988). Development and validation of brief measures of positive and negative affect: The PANAS scales. *Journal of Personality and Social Psychology*, 54(6), 1063-1070.
- Way, N., & Greene, M. L. (2006). Trajectories of perceived friendship quality during adolescence: The patterns and contextual predictors. *Journal of Research on Adolescence*, 16, 293-320.
- Way, N., Reddy, R., & Rhodes, J. (2007). Students' perceptions of school climate during the middle school years: Associations with trajectories of psychological and behavioral adjustment. *American Journal of Community Psychology*, 40, 194-213
- Whitley, A. M., Huebner, E. S., Hills, K. J., & Valois, R. F. (2012). Can students be too happy in school? The optimal level of school satisfaction. *Applied Research in Quality of Life*, 7(4), 337-350
- WHO (2000). *Obesity: preventing and managing the global epidemic. Report of a WHO Consultation. WHO Technical Report Series 894*. Geneva: World Health Organization.
- WHO (2010). *Global recommendations on physical activity for health*. http://apps.who.int/iris/bitstream/10665/44399/1/9789241599979_eng.pdf
- Williams, J., Wake, M., Hesketh, K., Maher, E., & Waters, E. (2005). Health-related quality of life of overweight and obese children. *Jama*, 293(1), 70-76.

- Winkelmann, L., & Winkelmann, R. (1998). Why are the unemployed so unhappy? Evidence from panel data. *Economica*, 65(257), 1-15.
- Wright, T. A., Bonett, D. G., & Sweeney, D. A. (1993). Mental health and work performance: Results of a longitudinal field study. *Journal of Occupational and Organizational Psychology*, 66(4), 277-284.
- Zambon, A., Morgan, A., Vereecken, C., Colombini, S., Boyce, W., Mazur, J., ... & Cavallo, F. (2010). The contribution of club participation to adolescent health: evidence from six countries. *Journal of epidemiology and community health*, 64(01), 89-95.

APPENDIX: DETAILED ITEM-LEVEL ASSESSMENT PLAN

The appendix is provided in the accompanying Excel file (APPENDIX_PISA2018_WB_Framework_11-09-2015(Prepared by J Bertling for OECD).xlsx)

