Organisation for Economic Co-operation and Development

DIRECTORATE FOR EDUCATION AND SKILLS

Mixed-Worded Scales and Acquiescence in Educational Large-Scale Assessments

OECD Education Working Paper No. 269

Janine Buchholz, OECD

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

Janine Buchholz (Janine.Buchholz@oecd.org)

JT03490862

<div align="center">**OECD EDUCATION WORKING PAPERS SERIES**</div>

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

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

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

The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

The use of this work, whether digital or print, is governed by the Terms and Conditions to be found at http://www.oecd.org/termsandconditions.

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

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

<div align="center">

----

www.oecd.org/edu/workingpapers

</div>

Abstract

Self-report data such as those regularly administered with questionnaires in the OECD's educational large-scale assessments are subject to response biases such as acquiescence, i.e., the tendency to agree with questionnaire items regardless their content. Research has shown that acquiescence affects the psychometric quality of such data, posing a threat to validity. Using a simple index that can be computed in the presence of both positively and negatively keyed items, the author examined the prevalence, the individual-level correlates, the impact on associations between indicators, as well as the county-level consistency of acquiescence for 16 questionnaires administered in four study programmes (*Programme for International Student Assessment*, PISA; *Teaching and Learning International Survey*, TALIS; *Survey on Social and Emotional Skills*, SSES; and *International Early Learning and Child Well-being Study*, IELS). Findings suggest that variation in acquiescence exists both between and within countries, the latter of which is determined by factors largely in line with prior research. Impact on associations as well as high levels of country-level consistency are evident. Based on these findings, recommendations for the construction of questionnaires to be administered in future assessments are derived.

Acknowledgments

Table of contents

FIGURES

I

TABLES

Chapter 1. Background

Educational large-scale assessments typically administer questionnaires comprised of Likert-type self-report items. Although being very economic, a "long-lasting issue" associated with the use of self-report items is the fact that they are affected by response biases (Primi et al., 2019, p. 447[1]), i.e., systematic factors influencing responses that are unrelated to the construct of interest. Examples for such biasing response styles are acquiescence, disacquiescence, mid-point responding, extreme responding, and non-contingent responding (for a comprehensive overview, see van Vaerenbergh and Thomas (2013[2])). McCrae (2018[3]) found that response biases explain up to 40% of the variance in self-report personality assessments.

This working paper focuses on acquiescent response style (ARS) – or "yea-saying" – i.e., the tendency of an individual to consistently agree to questionnaire items regardless of the content of the items (Rammstedt, Kemper and Borg, 2013[4]). Smith and Fischer (2007[5]) suggest ARS to be a function of both situational factors (such as item format and the context of administration) and respondent personality. Wetzel et al. (2016[6]) demonstrated a remarkably stable tendency to acquiescence over a period of eight years, thus discussing whether it could be considered a trait.

It is widely recognised that ARS poses a threat to the validity of self-report questionnaire data (Rammstedt, Danner and Bosnjak, 2017[7]; Smith and Fischer, 2007[5]). Research has demonstrated a number of methodological problems resulting from acquiescence affecting the psychometric quality of such data (Rammstedt, Danner and Bosnjak, 2017[7]). For example, as acquiescence can affect item means, mean differences between groups who differ systematically in their tendency to acquiesce are misleading. In addition, ARS holds the potential to distort the intended factorial structure by biasing item variances and covariances (Rammstedt, Goldberg and Borg, 2010[8]). It has been shown repeatedly that the use of both positively and negatively keyed items (see below) led to the emergence of separate factors because acquiescing respondents agree with the negatively keyed items more than non-acquiescent respondents do. In fact, a portion of only 10% of respondents failing to attend to the wording of the items is sufficient to cause a second factor to occur (Schmitt and Stuits, 1985[9]). Finally, it has been shown that acquiescence can substantially bias the associations of the target construct with other variables (Danner, Aichholzer and Rammstedt, 2015[10]; Primi et al., 2019[1]).

1.1. Measuring and controlling for acquiescence

Summing up, research has shown that acquiescence is to be regarded an "important, stable and influential response bias" (Primi et al., 2019[1]) which needs to be addressed in the context of studies relying on self-report data.

One approach to examine, and potentially control for, acquiescence is the inclusion of both positively and negatively keyed items in the assessment of the target construct, i.e., "markers of opposite poles of a trait" (Primi et al., 2019[1]). The following antonym item pair serves as an example for the two types of items: "If I could decide again, I would still choose to work as a teacher" and "I wonder whether it would have been better to choose another profession" (Items TT3G53B and TT3G53F in the TALIS 2018 Teacher Questionnaire, respectively). The Likert-scale response options for both items are "Strongly disagree" (1), "Disagree" (2), "Agree" (3), and "Strongly agree" (4). In the absence of acquiescence, a teacher strongly satisfied with his or her profession should

strongly agree with the first (raw response is 4) and strongly disagree with the second item (raw response is 1), resulting in a mean score of 2.5. In the presence of acquiescence, the teacher would indicate agreement with both items, e.g. responding in category 4 on both items, thus receiving a mean score of 4 – with other words, a result shifted by +1.5 units from the score expected in the absence of acquiescence (or the mid-point of the scale). Averaging raw responses over a set of such mixed-worded antonym pairs is a simple method to derive an indicator of acquiescent response style (ARS) and will present the basis for all analyses in the present paper. The indicator can be used for simple corrections such as ipsatization, within-person centering, or partialling the within-person mean response (Soto et al., 2008[11]; Ten Berge, 1999[12]). Other, more elaborated methods for controlling ARS have also been proposed (e.g., Primi et al., (2019[1]);Steinmann, Strietholt and Braeken (2021[13])).

While the use of mixed-worded scales allows for an examination of the degree of acquiescence – which is the scope of this paper – recommendations on their use for substantial research is controversial (Suárez-Álvarez et al., 2018[14]). Researchers in favour of using these scales argue that response biases such as acquiescence would be completely masked if only positively keyed items were present (Smith and Fischer, 2007[5]; Weijters and Baumgartner, 2012[15]). Arguments against the use of mixed-worded items include that negations violate common guidelines on item construction (Haladyna, Downing and Rodriguez, 2002[16]) and that a substantive portion of respondents fails to notice the changing item wording (Steinmann, Strietholt and Braeken, 2021[13]), particularly when their reading skills are poor (Marsh, 1996[17]), thus distorting the factor structure (Rammstedt, Goldberg and Borg, 2010[8]). Other negative consequences are the flawed measurement precision of the instrument and decreased variability in the response data (Suárez-Álvarez et al., 2018[14]).

1.2. Individual- and country-level correlates of acquiescence

Much research on characteristics of acquiescing respondents has been conducted, suggesting correlates to be on both the individual and the country level.

On the individual level, the strongest evidence exists with respect to the respondent's educational background. It has been shown repeatedly that acquiescence occurs more frequently among persons with a lower level of educational attainment (Meisenberg and Williams, 2008[18]; Narayan and Krosnick, 1996[19]; Rammstedt, Goldberg and Borg, 2010[8]; Rammstedt and Kemper, 2011[20]) and lower cognitive abilities (Gnambs and Schroeders, 2020[21]). Such respondents might have less clear self-concepts, smaller vocabulary and less developed comprehension skills which could lead to uncertainty when responding to questionnaires, administered in written form, thus leaving more room for systematic response biases such as acquiescence (Rammstedt, Danner and Bosnjak, 2017[7]). Another explanation suggests that respondents with relatively little education are especially susceptible to satisficing-induced response effects such as acquiescence (Narayan and Krosnick, 1996[19]). In contrast to the cognitively more demanding optimizing response behaviour for which multiple steps are executed (i.e. interpreting the meaning of a question, retrieving relevant information, integrating this information into a summary judgment, and reporting that judgment), satisficing respondents provide answers that appear to be satisfactory or acceptable to the interviewer without executing all the steps necessary for optimizing. While the negative relationship between educational level and acquiescence has been demonstrated repeatedly, evidence suggests this effect to not being universal but to differ systematically between countries. For example, Rammstedt, Kemper and Borg (2013[4]) found the influence of education to be stronger for individualistic countries (e.g. Germany). Prior research also demonstrated a relationship of acquiescence

with reading proficiency (Dunbar et al., 2000[22]; Quilty, Oakman and Risko, 2006[23]) and verbal ability (Marsh, 1996[17]). Mixed findings, however, exist for the relationship of acquiescence with age and with gender (Rammstedt, Danner and Bosnjak, 2017[7]). With respect to age, some studies found evidence for a positive relationship (Meisenberg and Williams, 2008[18]; Weijters, Geuens and Schillewaert, 2010[24]) while others failed to do so (Eid and Rauber, 2000[25]) or even demonstrated a negative relationship (Smith and Fischer, 2007[5]). The higher tendency to acquiesce with increasing age could be explained by increasing cognitive limitations (Weijters, Geuens and Schillewaert, 2010[24]). With respect to gender, some studies found higher levels of acquiescence for female respondents (Eid and Rauber, 2000[25]; Weijters, Geuens and Schillewaert, 2010[24]) while others did not (Marin, Gamba and Marin, 1992[26]). Finally, Quilty, Oakman and Risko (2006[23]) found a negative relationship of acquiescence with personality factors (self-reported conscientiousness and emotional stability).

Although not being the scope of this working paper, research has also pointed at systematic country differences in acquiescence. For example, van Herk, Poortinga and Verhallen (2004[27]) demonstrated that, among six European countries, higher acquiescence occurred in Mediterranean countries compared to North-western countries. Both a country's social and economic situation and its cultural orientations, especially the degree of collectivism, have been suggested as correlates of country-level acquiescence. For example, Meisenberg and Williams (2008[18]) found that acquiescence was most prevalent in less developed countries and could best be explained by the country's corruption level. Regarding culture, Johnson et al. (2005[28]) found acquiescence to be negatively related to individualism, uncertainty avoidance, power distance, and masculinity. Similarly, Smith and Fischer (2007[5]) found that individualism explained about 35% of the variance between countries, and suspect the need for harmony in collectivistic countries to cause a press for acquiescent responding.

1.3. Research questions

To examine the extent of acquiescence across OECD studies, its individual-level correlates, its impact on associations between variables, and its consistency across studies and countries, the present working paper addresses the following four research questions (RQ):

1. How prevalent is acquiescent response style (ARS) in the questionnaires administered in OECD studies?

2. Is there systematic variation in ARS across respondents and if so, can interindividual differences be explained by person characteristics (e.g., educational attainment, gender, age)?

3. Does the strength of the relationship between variables change in the theoretically expected direction when controlling for ARS?

4. On the country level, are findings about the average level of ARS consistent for populations (students, teachers, school principals, parents) across OECD studies?

Chapter 2. Method

2.1. Data

All studies conducted by the OECD in the field of educational large-scale assessments within the last decade were of potential interest for the purpose of this study, i.e. PISA, , TALIS, SSES, IELS, and PIAAC administered after 2011. In a first step, all questionnaires administered in these studies were examined in search of matrix questions (or "item batteries") containing both positively and negatively keyed items. As a result, a total of 16 questionnaires across seven studies and four study programmes (PISA, TALIS, IELS, and SSES) were identified and will form the basis for all subsequent analyses.

- The Programme for International Student Assessment (PISA) is a triannual assessment of 15-year-old students which has been conducted since 2000. The assessment consists of a two-hour test of mathematics, science and reading proficiency as well as a subsequently administered questionnaire of (at least) 35 minutes. In addition to the student questionnaire (STQ), questionnaires are administered to school principals of the sampled schools (SCQ) and, as national options, teachers (TCQ; since 2015) and the sampled students' parents (PAQ). Data from PISA 2018 (OECD, 2021[29]), 2015 (OECD, 2017[30]) and 2012 (OECD, 2014[31]) will be used for the analyses. With up to 80 countries participating in each of the three cycles, PISA forms the largest study programme in the present analyses.

- The Teaching and Learning International Survey (TALIS) is an assessment of teachers (TCQ) and school principals (SCQ) that has been conducted since 2008. The questionnaires take about 45 to 60 minutes to answer. Data from TALIS 2018 (OECD, 2019[32]) and 2013 (OECD, 2014[33]) will be used for the present analyses.

- The Survey on Social and Emotional Skills (SSES) is an assessment of children (STQ) at the ages of 10 and 15, their teachers (TCQ) and their parents (PAQ) in ten cities located in nine countries. Scales on the assessment of students' social and emotional skills that are administered in all three populations were explicitly constructed with the aim to include both positively and negatively keyed items. SSES is the newest study programme in the current analyses and has been administered only once in 2020 (OECD, 2021[34]).

- The International Early Learning and Child Well-being Study (IELS) is an assessment of children's early learning through questionnaires administered to educators (TCQ) and parents (PAQ) of five-year-old children (OECD, 2020[35]). It has been conducted once in 2018, and with three countries participating in the assessment, IELS forms the smallest study under investigation here.

- No mixed-worded scales administered in the Programme for the International Assessment of Adult Competencies (PIAAC) (OECD, 2013[36]) could be identified.

For each of the 16 questionnaires, Table 1 shows the total number of respondents, the number of cases with a valid score on the ARS index (ARSI), the number of countries with at least one valid score on the index, the number of pairs underlying the computation of the index, the mean and standard deviation across senate-weighted countries, the number of Likert-scale response options of the items underlying the computation of the index, as well as the deviance of ARSI from the value expected under the absence of acquiescence (i.e. the scale's mid-point). Regarding the latter, positive values are indicative of acquiescence.

Annex A contains a detailed overview regarding the countries participating in each of the studies as well as their mean ARSI score.

Table 1. Summary of studies and questionnaires included in the analyses

Study	Questionnaire	N	N (valid ARSI)	# CNT	# ARSI pairs	ARSI: M (SD)	Likert scale, mid-point	Δ_{ARSI}
(1) PISA 2018	(1) STQ	612 004	590 869	80	13	2.506 (0.248)	4-point, 2.5	0.006
	(2) TCQ	107 367	89 731	19	3	2.537 (0.283)	4-point, 2.5	0.037
	(3) PAQ	112 011	91 017	17	3	2.293 (0.404)	4-point, 2.5	-0.207
(2) PISA 2015	(4) STQ	519 334	487 311	71	3	2.390 (0.336)	4-point, 2.5	-0.110
	(5) TCQ	108 292	87 935	18	2	2.476 (0.323)	4-point, 2.5	-0.024
(3) PISA 2012	(6) STQ	480 174	470 902	65	15	2.641 (0.339)	4-point, 2.5	0.141
	(7) SCQ	18 139	17 540	65	3	3.046 (0.396)	4-point, 2.5	0.546
(4) TALIS 2018	(8) TCQ	261 426	254 017	48	4	2.461 (0.244)	4-point, 2.5	-0.039
	(9) SCQ	15 980	15 046	48	3	2.462 (0.278)	4-point, 2.5	-0.038
(5) TALIS 2013	(10) TCQ	186 346	180 683	36	8	2.556 (0.271)	4-point, 2.5	0.056
	(11) SCQ	11 121	10 362	36	2	2.492 (0.293)	4-point, 2.5	-0.008
(6) SSES	(12) STQ	60 985	60 985	10	28	3.127 (0.270)	5-point, 3.0	0.127
	(13) TCQ	51 085	51 084	10	5	3.106 (0.317)	5-point, 3.0	0.106
	(14) PAQ	34 650	34 649	10	28	3.123 (0.240)	5-point, 3.0	0.123
(7) IELS	(15) TCQ	6 921	6 431	3	8	2.939 (0.299)	5-point, 3.0	-0.061
	(16) PAQ	6 921	4 993	3	8	3.187 (0.277)	5-point, 3.0	0.187

Note: STQ: Student Questionnaire; TCQ: Teacher Questionnaire; PAQ: Parent Questionnaire; SCQ: Questionnaire administered to School Principals; Δ_{ARSI}: Difference between ARSI and mid-point of scale, with positive values indicating acquiescence.

2.2. Construction of the ARS index

The construction of the ARS index (ARSI) is based on mixed-worded item pairs, i.e., pairs of items with one being positively and one being negatively related to the measured construct. The most important criterion for the selection of these items and the construction of pairs consisted in a high conceptual overlap, i.e. items aiming to measure the same construct. To facilitate this search, only items within a given matrix question were considered for pairing (OECD, 2021[37])[1]. In addition, Technical Reports were consulted to check whether these items were initially considered to form a unidimensional construct. As an example, the PISA 2018 TCQ contained one question (TC198) with mixed-worded items aiming to measure two different constructs: satisfaction with the job and satisfaction with the teaching profession (SATJOB and SATTEACH, respectively). Only items aiming to measure the same construct were considered to form a pair. As a second criterion, items were preferred to form a pair based on semantic similarity, i.e. when the same verb or noun was present. Finally, only pairs comprising items with the same number of response categories were kept within each of the questionnaires as this is a necessary prerequisite for the calculation of the index (see below).Chapter 7. Annex A contains a comprehensive summary of all item pairs including their wording as well as descriptive statistics.

Before calculating the index, items were reverse-coded so that higher values indicated stronger agreement. As a result, ARSI indicates the tendency to choose a Likert-type response option representing agreement, not a response option "on the right-hand side" of

---

[1] Because of the random item order design regarding the measurement of the 15 constructs in the SSES study, the search was conducted across questions, guided by the design underlying the item development (OECD, 2021[37])

the response scale. Table 2 lists all questions and their initial response format for which such recoding was conducted.

Table 2. Questions for which reverse-coding of the items was necessary to indicate stronger agreement

Study	QQQ	Question ID	Sample item	Original response options
PISA 2018	STQ	ST034	I make friends easily at school.	1: Strongly agree; 2: Agree; 3: Disagree; 4: Strongly disagree
PISA 2015	STQ	ST034	I feel like I belong at school.	
PISA 2012	STQ	ST42	I learn mathematics quickly.	
		ST43	If I put in enough effort I can succeed in mathematics.	
		ST85	My teacher starts lessons on time.	
		ST87	Other students seem to like me.	
		ST88	School has helped give me confidence to make decisions.	
		ST91	It is completely my choice whether or not I do well at school.	
	ICQ	IC22	The computer is a very useful tool for my schoolwork.	
	SCQ	SC27	Mathematics teachers are interested in trying new methods and teaching practices.	
		SC28	There is consensus among mathematics teachers that academic achievement must be kept as high as possible.	
		SC29	There is consensus among mathematics teachers that the social and emotional development of the students is as important as their acquisition of mathematical skills and knowledge in mathematics classes.	

Note: QQQ: Questionnaire; STQ: Student Questionnaire; ICQ: ICT Familiarity Questionnaire administered to students (national option); SCQ: Questionnaire administered to School Principals

Based on the mean of two items i constituting an item pair j, ARSI is computed as the mean across all J item pairs:

$$ARSI = \frac{\sum_j^J \frac{\sum_{i=1}^2 item_{ij}}{2}}{J}$$

The index, therefore, can be computed even when responses to some of the items are missing – as long as at least one item pair has two valid responses. Within a given questionnaire, the index can thus potentially be based on different sets of item pairs. This procedure, however, became necessary in the context of this study as individual countries have the option to exclude individual questions from an assessments so that listwise deletion would have excluded these countries completely.

Chapter 3. Prevalence of ARS across studies and countries (RQ 1)

3.1. Method

To answer Research Question 1, univariate analyses of the ARS indicator were conducted, both within and between countries. Descriptive statistics as well as visual displays of the distribution of ARSI within countries show whether variation in the index exists. For the computation of the mean index across all countries participating in a given questionnaire, the data were senate-weighted in line with general recommendations on the use of large-scale assessment data (Rutkowski et al., 2010[38]). Chapter 7. Annex C documents the variables used for weighting. All analyses are conducted with R (version 4.1.0) (R Core Team, 2021[39]); weighting was implemented using the R-package Hmisc (Harrell and Dupont, 2021[40]).

3.2. Results

For each of the studies and questionnaires, Table 1 contains descriptive statistics on the mean and standard deviation of the ARS index. In addition, Figure 1 shows the distribution of ARSI across countries for each of the 16 questionnaires. Through the varying number of lines in these diagrams, the plots immediately show the variation in the number of countries participating in each of the studies. In addition, the plots show more or less continuous distributions, depending on the number of pairs underlying the computation of ARSI (see Table 1). Most important to all of the subsequent analyses, the plots consistently demonstrate great variation on ARSI – both within countries (relevant for RQ 2) and between countries (relevant for RQ 4). Finally, the plots include a line reflecting the expected ARSI score under the absence of acquiescence, thus allowing to gain a quick insight into the presence of acquiescence: a country's tendency to show acquiescence is reflected by a distribution's shift to the right-hand side of the plotted area.

Several patterns are noticeable and will be discussed in the following. The PISA 2018 parent questionnaire (PAQ) shows a negative deviation from the ARSI expected under the absence of acquiescence, implying a rather surprising tendency to disagree. The finding can be explained by the fact that the index is based on two item pairs only, stemming from a question aiming to measure the enjoyment of reading (JOYREAD; e.g., ST160Q02IA: "Reading is one of my favourite hobbies."). The index, therefore, is confounded with the content of the measured construct and might rather reflect parents' general tendency to dislike reading. The PISA 2018 PAQ data will therefore need to be interpreted with caution in the following. Although the PISA 2012 School Questionnaire (SCQ) shows a strong positive deviation from the expected ARSI and the distribution looks quite continuous, the index is based on three items only. Another interesting pattern is present in the IELS data: although rating the behaviour of the same children, parents (PAQ) show higher ARSI than teachers (TCQ), implying a higher tendency to agree with the items regardless their content, thus expressing a generally more positive view of their children.

Figure 1. Distribution of the ARS Index by country for each of the studies and questionnaires
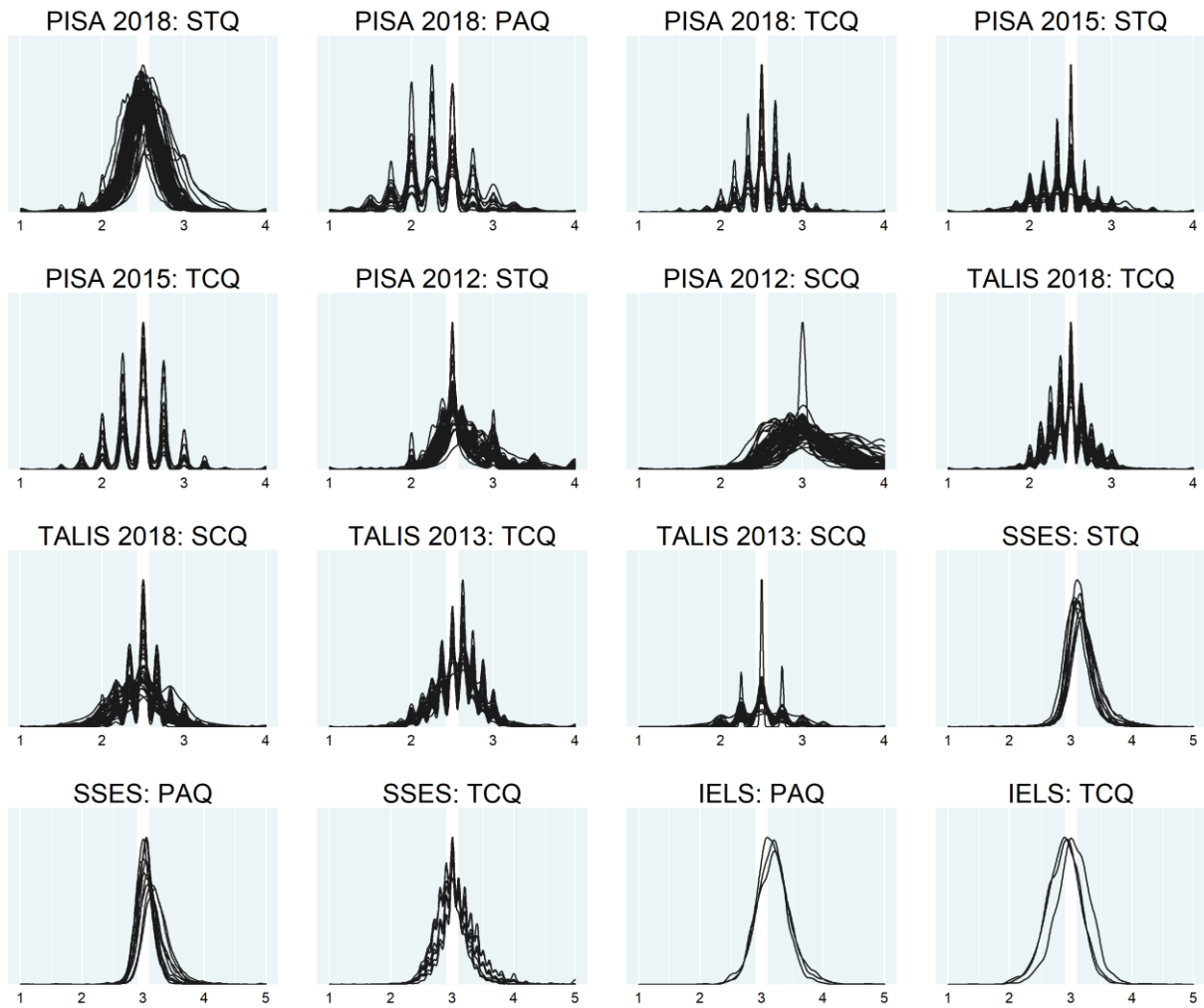
Note: The thick white vertical line represents the score expected in the absence of acquiescence

Chapter 7. Annex A contains the ARSI mean for each of the countries participating in each of the questionnaires.

Chapter 4. Individual-level correlates of ARS (RQ 2)

After the previous section demonstrated variation of ARS within and between countries, research question 2 investigates whether this variation can be explained by person characteristics.

4.1. Method

Due to their differences in population and scope, each study and questionnaire collected different information about the respondent. Across the various questionnaires, individual-level characteristics relevant for the purpose of this study were available for the following categories: (a) educational attainment, (b) migration background and language use, (c) age, (d) gender, and (e) other. Table 3 provides an overview, detailing which category was available for each of the questionnaires. Proxies for educational attainment, the strongest correlate of acquiescence based on prior research, were administered in most of the questionnaires. Gender, a characteristic prominent in the literature on ARS, has been administered in almost all of the studies. No information relevant for the purpose of this study was available for school principals administered in PISA 2012 and teachers administered in IELS.

Table 3. Availability of relevant individual-level correlates of ARS

Study	QQQ	Educational attainment	Migration / language use	Age	Gender	Other
PISA 2018	STQ	X	X		X	X
	TCQ		X	X	X	
	PAQ	X	X		X	X
PISA 2015	STQ	X	X		X	X
	TCQ			X	X	
PISA 2012	STQ	X			X	X
	SCQ					
TALIS 2018	TCQ	X		X	X	
	SCQ	X		X	X	
TALIS 2013	TCQ	X		X	X	
	SCQ	X		X	X	
SSES	STQ	X		X	X	
	TCQ	X		X	X	
	PAQ	X	X	X	X	
IELS	TCQ					
	PAQ		X		X	X

Note: QQQ: Questionnaire, administered to students (STQ), teachers (TCQ), school principals (SCQ) and parents (PAQ).

To answer Research Question 2, bivariate correlations between the ARS index and the respective indicator for each of the five categories (gender, educational attainment and/or cognitive skills, age, migration background/language proficiency, and other) were computed within countries. For the majority of cases, linear correlations were computed; Kendall's Tau was computed for ordinal data which will be indicated in the subsequent presentation of findings. Data were weighted based on the respondent weight

(see Annex C). IBM SPSS (version 26.0) (IBM Corp., 2019[41]) was used to compute the weighted correlations while their summary statistics were computed using R (version 4.1.0) (R Core Team, 2021[39]).

4.2. Results

The presentation of results is organised according to the five categories of individual-level characteristics expected to relate to acquiescence.

4.2.1. Educational Attainment

Table 4 contains a list of all education-related indicators for each of the studies and questionnaires, the number of countries for which the correlation could be computed, and summary statistics regarding the respective correlation within countries. This table, as well as those following afterwards, is organised according to the respondent's population (students, teachers, school principals, parents). As all variables are coded so that higher values indicate higher levels of education, negative correlations would be expected based on previous research. With only a few exceptions, the present findings support such a negative relationship.

Table 4. Correlation between ARSI and variables on educational attainment

| QQQ | Study | Variable | Variable meaning [coding] | # CNT | Correlation | | | |
					Md	[Min , Max]	% pos.	% sig. ($p<.05$)
STQ	PISA 2018	PV[1:10]READ	PVs for Reading (mean correlations across PVs)	80	-0.013	[-0.19 , 0.20]	41.2	62.5
		PV[1:10]MATH	PVs for Math (mean correlations across PVs)	80	-0.028	[-0.19 , 0.16]	26.2	53.8
		PV[1:10]SCIE	PVs for Science (mean correlations across PVs)	80	-0.036	[-0.20 , 0.18]	28.7	62.5
		BSMJ	Students' expected occupational status (SEI)	80	0.028	[-0.08 , 0.12]	76.2	50.0
		ISCEDL*	ISCED level [1-5]	80	-0.002	[-0.07 , 0.12]	41.2	26.2
		REPEAT (R)	Grade Repetition [0: no, 1: yes]	80	-0.003	[-0.14 , 0.07]	58.8	37.5
	PISA 2015	PV[1:10]READ	PVs for Reading (mean correlations across PVs)	73	-0.083	[-0.30 , 0.08]	9.6	80.8
		PV[1:10]MATH	PVs for Math (mean correlations across PVs)	73	-0.079	[-0.26 , 0.04]	4.1	78.1
		PV[1:10]SCIE	PVs for Science (mean correlations across PVs)	73	-0.083	[-0.26 , 0.05]	4.1	79.5
		BSMJ	Students' expected occupational status (SEI)	73	-0.025	[-0.15 , 0.04]	20.5	39.7
		REPEAT (R)	Grade Repetition [0: yes, 1: no]	73	0.048	[-0.05 , 0.14]	12.3	68.5
	PISA 2012	PV[1:5]READ	PVs for Reading (mean correlations across PVs)	65	-0.034	[-0.19 , 0.05]	3.1	92.3
		PV[1:5]MATH	PVs for Math (mean correlations across PVs)	65	-0.200	[-0.30 , 0.05]	3.1	93.8
		PV[1:5]SCIE	PVs for Science (mean correlations across PVs)	65	-0.186	[-0.30 , 0.06]	3.1	93.8
		iscedl*	ISCED level [1-8]	65	-0.191	[-0.30 , 0.07]	26.2	52.3
	SSES	Sgrade_Math	Standardised grade - Mathematics	10	-0.063	[-0.12 , -0.03]	0	90.0
		Sgrade_Read_Lang	Standardised grade - Reading / Language	10	-0.041	[-0.15 , 0.01]	10.0	80.0
TCQ	TALIS 2018	TT3G03*	Highest level of formal education [ISCED 3-8]	48	0.016	[-0.04 , 0.11]	64.6	37.5
	TALIS 2013	TT2G10*	Highest level of formal education [ISCED 3-8]	36	0.013	[-0.03 , 0.06]	66.7	19.4
	SSES	TCQM00601	Highest level of formal education [ISCED 1-8]	10	-0.016	[-0.13 , 0.03]	30.0	90.0
SCQ	TALIS 2018	TC3G03*	Highest level of formal education (ISCED) [3-8]	48	0.032	[-0.17 , 0.15]	60.4	58.3
	TALIS 2013	TC2G03*	Highest level of formal education (ISCED) [3-8]	36	-0.014	[-0.12 , 0.11]	38.9	58.3

QQQ	Study	Variable	Variable meaning [coding]	# CNT	Correlation			
					Md	[Min , Max]	% pos.	% sig. (*p*<.05)
PAQ	PISA 2018	MISCED*	Mother's Education (ISCED) (if mother responded)	17	-0.004	[-0.16 , 0.09]	47.1	58.8
		FISCED*	Father's Education (ISCED) (if father responded)	17	-0.009	[-0.16 , 0.07]	47.1	35.3
		BMMJ1	ISEI of Mother (if mother responded)	17	0.003	[-0.16 , 0.10]	52.9	41.2
		BFMJ2	ISEI of Father (if father responded)	17	-0.036	[-0.19 , 0.12]	35.3	17.6
	SSES	PAQM00601*/ PAQM00602*	Highest level of formal education (ISCED) [1-8] of respondent (either mother/female guardian or father/male guardian)	10	-0.106	[-0.22 , 0.04]	10.0	100.0

Note: QQQ: Questionnaire, administered to students (STQ), teachers (TCQ), school principals (SCQ) and parents (PAQ); (R): variable has been recoded; '# CNT': number of countries for which the information was available; Md: Median; '% pos.': percentage of positive coefficients across countries; '% sig. (*p*<.05)': percentage of significant coefficients across countries; * indicates that a rank correlation was computed.

For students, measures of achievement in the cognitive test (PISA 2018, 2015, 2012) as well as self-reported grades (SSES) show strong evidence for the assumed negative relationship, with the picture being least clear in PISA 2018. Figure 2 provides more detailed information, showing the average correlation coefficient between ARSI and the 10 (2018, 2015) or 5 (2012) domain-related Plausible Values (Reading, Science, Mathematics), respectively, for each of the countries and the three PISA cycles under investigation. While higher values on ARSI relate to lower test scores in the vast majority of countries in PISA 2012 and 2015, the pattern is mixed for PISA 2018. For a large number of countries (particularly, OECD countries), no or even a positive relationship was found. Similarly, the other three indicators (expected occupational status, ISCED level, and the absence of grade repetition) show an unexpected, although small, positive relationship with ARSI in PISA 2018. As the very same indicators show relationships in the expected direction for PISA 2015 and 2012, it is hard to argue that the positive relationship could be due to overclaiming on these self-report measures. Instead, the unexpected findings might point at problems with the student questionnaire ARS index in PISA 2018.

For teachers, school principals and parents, the patterns are rather mixed, with positive and negative relationships across countries and studies. A potential explanation for the heterogeneous findings and negligible effect sizes among teachers and school principals is the little within-country variation that can be expected on the ISCED, an indicator of formal education which, in turn, presents a requirement to become a teacher or school principal, respectively.

Figure 2. Mean correlation between the ARS Index and Plausible Values for Reading, Science and Mathematics within countries for three cycles of PISA
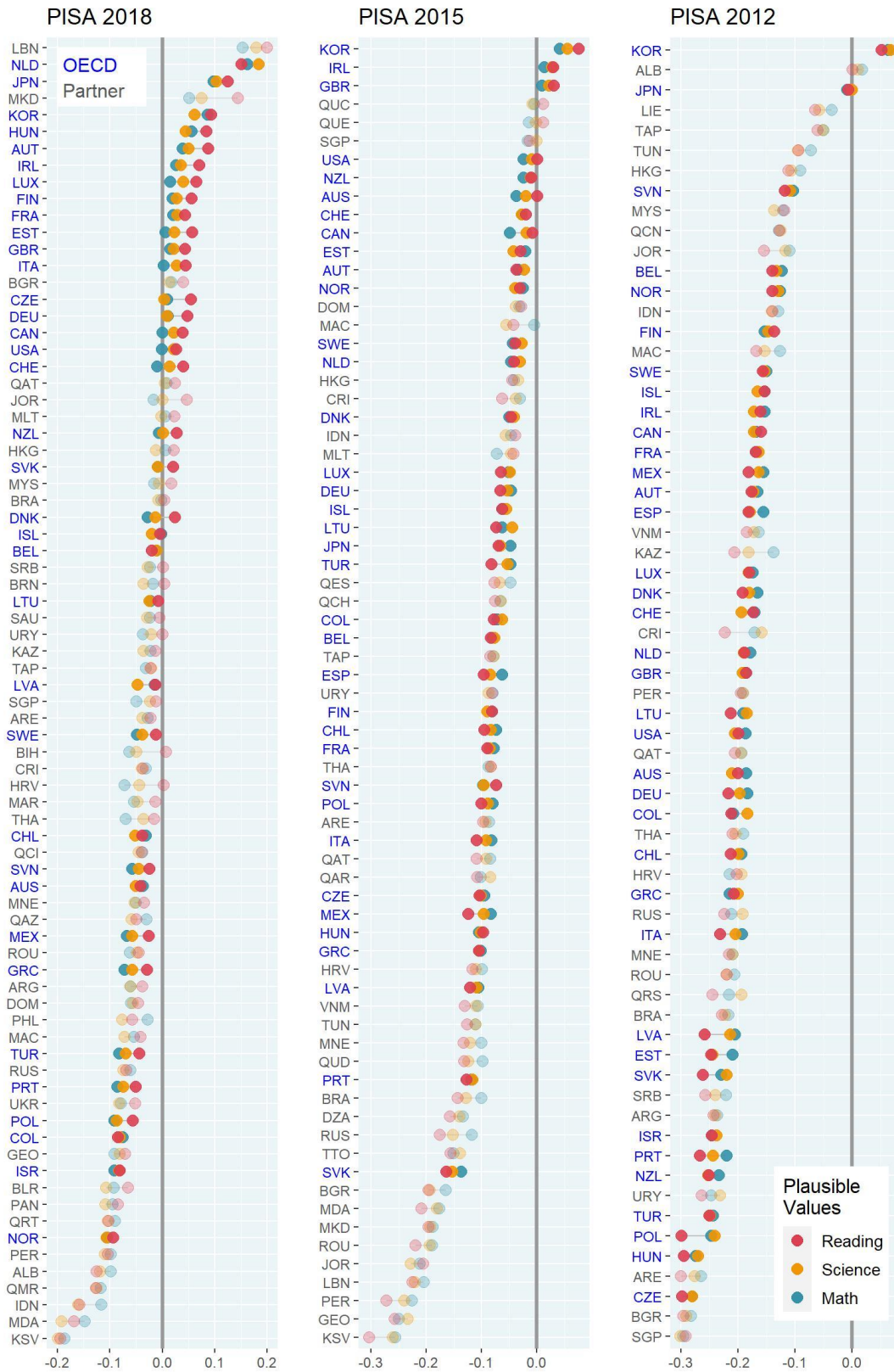
4.2.2. Migration background and use of the test language at home

Prior research demonstrated a negative relationship between ARSI and reading/verbal ability. While indicators for reading proficiency are available for students in PISA (Plausible Values for Reading) and SSES (standardised grade in Reading; see previous section), no such information is available for the other populations. However, many studies assessed a potential migration background, either directly or indirectly by asking whether the language spoken at home corresponds with the test language. Although these indicators present rather distal determinants of reading/verbal ability, the following section reviews findings for the studies at hand.

Table 5 contains a list of all indicators relating to a migration background and the use of the test language at home. As all variables are coded so that higher values indicate a migration background or a language different from the language of the assessment, positive correlations would be expected based on previous research. With only a few exceptions, the findings support such a positive relationship. Just as in the previous section, findings for students in PISA 2018 are rather unexpected for about half of the countries. A similarly unexpected finding occurred for teachers in PISA 2018 where positive correlations occurred in only about 37% of countries. With the average correlation being close to zero and hardly significant in any of the countries, this finding could instead point at the negligible impact of language proficiency on the response behaviour of teachers. It is plausible to assume that little variation exists in the language proficiency among in-service teachers who must have received a teaching qualification in the country of the assessment.

Table 5. Correlation between ARSI and variables on migration background and language use

QQQ	Study	Variable	Variable label / coding	# CNT	Correlation			
					Md	[Min , Max]	% pos.	% sig. (*p*<.05)
STQ	PISA 2018	ST022Q01TA	Language at home (1: test language, 2: other)	80	0.004	[-0.07 , 0.10]	56.2	30.0
		IMMIG*	Index immigration status (1: Native, 2: Second gen., 3: first gen.)	80	0.000	[-0.14 , 0.03]	48.8	26.2
	PISA 2015	ST022Q01TA	Language at home (1: test language, 2: other)	73	0.022	[-0.05 , 0.16]	69.9	43.8
		IMMIG*	Index immigration status (1: Native, 2: Second gen., 3: first gen.)	73	0.014	[-0.04 , 0.16]	74.0	31.5
TCQ	PISA 2018	TC186Q01HA	Country of birth (1: Country of test, 2: other country)	19	-0.002	[-0.07 , 0.09]	36.8	10.5
PAQ	PISA 2018	PA155Q01IA	In what language did most of the activities in the previous question take place? (1: test language, 2: other language)	17	0.006	[-0.05 , 0.23]	58.8	35.3
	SSES	PAQM01302/ PAQM01303	Country of birth of respondent (either mother/female or father/male guardian) (1: country of test, 2: other)	10	0.042	[-0.03 , 0.12]	80.0	100.0
	IELS	STUD_LANG	Student language most often spoken at home (Recoded: 0: language of assessment, 1: other)	3	0.032	[0.02 , 0.07]	100.0	66.7
		PARENT_LANG	At least one parent primarily speaks a language other than the assessment language at home (0: no, 1: yes)	3	0.091	[0.02 , 0.11]	100.0	100.0
		IMMIG	Immigration background - both parents born abroad (0: no, 1: yes)	3	0.100	[0.08 , 0.13]	100.0	100.0

Note: QQQ: Questionnaire, administered to students (STQ), teachers (TCQ), and parents (PAQ); '# CNT': number of countries for which the information was available; Md: Median; '% pos.': percentage of positive coefficients across countries; '% sig. (*p*<.05)': percentage of significant coefficients across countries; * indicates that a rank correlation was computed.

4.2.3. Age

Table 6 contains all variables assessing age across the various questionnaires as well as summary statistics for their relationship with ARSI. Prior evidence on the relationship between acquiescence and age was not clear; however, those studies that found significant findings suggested a positive relationship. Contrary to these findings, the median correlation in all but one questionnaire was negative, the exception being the only student questionnaire. For all others, i.e. teachers, principals and parents, the respective correlation is negative in the vast majority of countries, pointing at a higher tendency of younger respondents to show acquiescence.

Table 6. Correlation between ARSI and variables assessing age

| QQQ | Study | Variable | Variable meaning/coding | # CNT | Correlation | | | |
					Md	[Min , Max]	% pos.	% sig. (*p*<.05)
STQ	SSES	CohortID	Age (1: Younger; 2: Older)	10	0.030	[-0.13 , 0.15]	60.0	100.0
TCQ	PISA 2018	TC002Q01NA	Age (in years)	19	-0.097	[-0.23 , 0.04]	15.8	84.2
	PISA 2015	TC002Q01NA	Age (in years)	18	-0.051	[-0.16 , 0.05]	22.2	72.2
	TALIS 2018	TCHAGEGR*	Age (in 6 categories: <25, 25-29, …)	48	-0.089	[-0.19 , 0.00]	0	95.8
	TALIS 2013	TT2G02	Age (in years)	36	-0.046	[-0.13 , 0.06]	19.4	88.9
	SSES	TCQM00201	Age (in years)	10	-0.073	[-0.12 , -0.04]	0	100.0
SCQ	TALIS 2018	PRAGEGR*	Age (in 4 categories: <40, 40-49, …)	48	-0.019	[-0.18 , 0.13]	35.4	58.3
	TALIS 2013	TC2G02	Age (in years)	36	-0.019	[-0.21 , 0.21]	44.4	66.7
PAQ	SSES	PAQM00301/ PAQM00302*	Age of female/male guardian (in 7 categories: <25, 25-29, …)	10	-0.063	[-0.13 , -0.03]	0	100.0

Note: QQQ: Questionnaire, administered to students (STQ), teachers (TCQ), school principals (SCQ) and parents (PAQ); '# CNT': number of countries for which the information was available; Md: Median; '% pos.': percentage of positive coefficients across countries; '% sig. (*p*<.05)': percentage of significant coefficients across countries; * indicates that a rank correlation was computed.

The analyses above were conducted within populations. The relationship between acquiescence and age can, however, also be analysed across the four populations under investigation in this study (students, teachers, principals, and parents). These four populations belong to two rather distinct age groups – adolescents and adults. Table 1 contains descriptive statistics for ARSI for each of the questionnaires, allowing to compare the average ARSI between these two age groups. According to Table 1, students tend to show the lowest levels of ARSI, followed by teachers, school principals and parents. Taken together, these findings point at a positive relationship between ARSI and age.

4.2.4. Gender

Table 7 contains all variables assessing gender across the various questionnaires as well as summary statistics for their relationship with ARSI. Prior findings on the relationship between acquiescence and gender were unclear, with studies showing either no relationship or a higher tendency for females to acquiesce. As all variables were recoded so that higher values indicate female gender, the tentative expectation based on previous findings would be reflected by positive correlations. The present results, however, are mixed. Parents are the only group in which findings are rather consistent, pointing at higher levels of acquiescence for females.

Table 7. Correlation between ARSI and female gender

QQQ	Study	Variable	Original codes	# CNT	Correlation			
					Md	[Min , Max]	% pos.	% sig. (*p*<.05)
STQ	PISA 2018	ST004D01T	1: Female, 2: Male	80	0.098	[-0.01 , 0.23]	98.8	92.5
	PISA 2015	ST004D01T	1: Female, 2: Male	73	-0.031	[-0.08 , 0.10]	28.8	65.8
	PISA 2012	ST04Q01	1: Female, 2: Male	65	-0.042	[-0.17 , 0.06]	15.4	32.3
	SSES	Gender_Std	1: Female, 2: Male	10	0.024	[-0.02 , 0.08]	80.0	80.0
TCQ	PISA 2018	TC001Q01NA	1: Female, 2: Male	19	-0.006	[-0.10 , 0.03]	31.6	10.5
	PISA 2015	TC001Q01NA	1: Female, 2: Male	18	-0.008	[-0.06 , 0.03]	44.4	38.9
	TALIS 2018	TT3G01	1: Female, 2: Male	48	-0.003	[-0.06 , 0.05]	50.0	77.1
	TALIS 2013	TT2G01	1: Female, 2: Male	36	0.019	[-0.05 , 0.06]	66.7	86.1
	SSES	TCQM00101	1: Female, 2: Male	10	-0.033	[-0.08 , 0.03]	20.0	100.0
SCQ	TALIS 2018	TC3G01	1: Female, 2: Male	48	-0.036	[-0.38 , 0.19]	16.7	56.2
	TALIS 2013	TC2G01	1: Female, 2: Male	36	-0.002	[-0.23 , 0.26]	47.2	52.8
PAQ	PISA 2018	PA001Q01TA/ PA001Q02TA	0: Not checked, 1: checked	17	0.042	[-0.04 , 0.09]	64.7	76.5
	SSES	PAQM00201	1: Mother, 2: Other female guardian, 3: Father, 4: Other male guardian	10	0.039	[-0.01 , 0.10]	80.0	70.0
	IELS	ELPAQ1301	1: Mother or female guardian, 2: Father or male guardian	3	0.014	[0.00 , 0.03]	66.7	66.7

Note: QQQ: Questionnaire, administered to students (STQ), teachers (TCQ), school principals (SCQ) and parents (PAQ); '# CNT': number of countries for which the information was available; Md: Median; '% pos.': percentage of positive coefficients across countries; '% sig. (*p*<.05)': percentage of significant coefficients across countries. All variables have been recoded so that higher levels indicate female gender.

4.2.5. Other

This section covers the analysis of person characteristics, which, to the best of the author's knowledge, have not been subject to prior research on acquiescence. These characteristics consist of the respondent's socio-economic status (SES) as well as the effort invested into the completion of the assessment – two variables typically associated with higher performance in achievement tests. The operationalisation of *SES* differed across populations: for students in PISA and parents in IELS, the index is a composite score reflecting the respondent's self-reported home resources as well as their parent's or their own highest educational and occupational level, respectively; for parents in PISA 2018, the annual household income serves as a proxy for SES. *Effort* was assessed using the effort thermometer, asking students in PISA on a 10-point Likert-type rating scale to indicate the level of effort they have invested into the completion of the test as well as the level of effort they would have invested if the marks from the test were going to be counted into the student's school marks.

Table 8 lists the corresponding variables as well as summary statistics for their relationship with ARSI. As all variables are coded so that higher values indicate higher levels of SES and effort, respectively, a finding corresponding to the above-mentioned expectations would be reflected by negative correlations. With the exception of the student questionnaire in PISA 2018, findings support such a negative relationship with SES: respondents reporting a higher socio-economic background tend to show less acquiescence. Contrary to the expectations above, students' self-reported effort tends to relate positively to ARSI. This finding could be explained by a general tendency of acquiescing respondents to respond in higher response categories not only on Likert-type items regardless their content, but to indicate higher levels of effort on the effort thermometer as well.

Table 8. Correlation between ARSI and other variables of interest

QQQ	Study	Variable	Variable label / coding	# CNT	Correlation			
					Md	[Min , Max]	% pos.	% sig. (p<.05)
STQ	PISA 2018	ESCS	Socio-economic background	80	0.041	[-0.05 , 0.16]	86.2	70.0
		EFFORT1	How much effort did you put into this test?	80	0.045	[-0.05 , 0.10]	97.5	75.0
		EFFORT2	How much effort would you have invested?	80	0.039	[-0.03 , 0.10]	91.2	57.5
	PISA 2015	ESCS	Socio-economic background	73	-0.015	[-0.16 , 0.12]	35.6	35.6
	PISA 2012	ESCS	Socio-economic background	65	-0.056	[-0.15 , 0.07]	18.5	49.2
		CLCUSE301	How much effort did you put into this test?	65	0.019	[-0.06 , 0.17]	72.3	3.1
		CLCUSE302	How much effort would you have invested?	65	0.001	[-0.08 , 0.14]	50.8	10.8
PAQ	PISA 2018	PA042Q01TA*	Annual household income (6 categories)	17	-0.055	[-0.15 , 0.03]	17.6	76.5
	IELS	SES	Socio-economic background	3	-0.109	[-0.13 , -0.08]	0	100.0

Note: QQQ: Questionnaire, administered to students (STQ) and parents (PAQ); '# CNT': number of countries for which the information was available; Md: Median; '% pos.': percentage of positive coefficients across countries; '% sig. (p<.05)': percentage of significant coefficients across countries; * indicates that a rank correlation was computed.

Chapter 5. Impact of ARS on relations between variables (RQ 3)

This research question examines the impact of acquiescence on substantive research questions involving self-report questionnaire data. Very elaborated methods for controlling ARS in the computation of scale scores have been suggested (e.g. Primi et al. (2019[1])); their application, however, would exceed the scope of this paper. Instead, the present section examines whether the relationship between two variables changes in the theoretically expected direction when ARS is controlled for. Instead of a thorough presentation of findings across the heterogeneous study programmes, the large number of studies and potential pairs of variables to look at – an almost impossible endeavour – the subsequent presentation of results will be restricted to only a few examples to demonstrate possible effects.

5.1. Method

The presentation of subsequent findings is based on data from the PISA 2012 student questionnaire for several reasons. With 15 pairs underlying the computation of ARSI, it represents an average number of pairs in the studies under investigation (see Table 1). Also, PISA is not only a prominent but also large study programme, contributing seven of the 16 questionnaires in this study. The large number of participating countries allows one to see some variation in findings, particularly as Figure 1 demonstrated considerable variance of ARSI within and between countries. Finally, PISA 2012 was selected after the previous section demonstrated unexpected findings regarding ARSI computed for PISA 2018.

To investigate the impact of ARS on the relationship between variables, two sets of analyses were conducted for each pair of variables: (1) a simple linear regression with one variable serving as the only predictor, and (2) a multiple regression in which ARSI serves as additional predictor, thus controlling for the effect of acquiescence in the analysis of the bivariate relationship. Results are based on the comparison of standardised regression weights. All analyses were conducted within countries and individual-level weights were applied (see Annex C). The regression analyses were conducted in IBM SPSS (version 26.0; IBM Corp., 2019) while summary statistics were computed using R (version 4.1.0; R Core Team, 2021).

Table 9 contains the selected pairs of variables, including sample items, and the expectation regarding the direction of their association. Each of the pairs includes Plausible Values (PVs) for the mathematics test score due to its prominence in reporting and because mathematics presented the major domain in PISA 2012. Accordingly, all regression analyses were replicated across the five PV variables each, and the subsequent findings are based on the mean of the standardised regression weights across the five regression analyses each. All of the independent variables in Table 9 are scale scores based on IRT scaling (for a documentation of the procedure, see OECD (2014[31]).

It is important to note that none of these scales included mixed-worded items to avoid a potential confounding of the respective variable and ARSI.

Table 9. Pairs of variables underlying the regression analyses

Dependent variable	Independent variable: Name (label)	Sample item of scale	Expected association
PV[1:5]MATH	Mathematics anxiety (ANXMAT)	I often worry that it will be difficult for me in mathematics classes	Negative
	Cognitive activation in mathematics lessons (COGACT)	The teacher asks questions that make us reflect on the problem	Positive
	Index of economic, social and cultural status (ESCS)	How many books are there in your home?	Positive
	Mathematics interest (INTMAT)	I enjoy reading about mathematics	Positive
	Mathematics behaviour (MATBEH)	I talk about mathematics problems with my friends	Positive
	Mathematics self-efficacy (MATHEFF)	How confident do you feel about having to do the following mathematics tasks?: Using a <train timetable> to work out how long it would take to get from one place to another	Positive

5.2. Results

Figure 3 contains a visual presentation of results for the six selected pairs of variables. For each of the 65 countries participating in PISA 2012, dots represent the average regression weight across Plausible Values with and without controlling for ARS, respectively.

Across the six variables of interest, different patterns occurred. While controlling for ARS seems to have no or only a small effect for the relationship of mathematics performance with ESCS and mathematics self-efficacy, controlling for ARS causes increasing (i.e. more positive) associations between variables in the remaining four cases. These findings correspond with the expectations regarding COGACT, INTMAT, and MATBEH: When controlling for acquiescence, higher levels of cognitive activation, mathematics interest and mathematics behaviour are associated with better performance in the mathematics test. With other words, the relationship between variables changes in the theoretically expected positive direction, creating a clearer picture about their association. An explanation for the absence of such an improved prediction for ESCS and mathematics self-efficacy could be the distinct nature of their indicator items. ESCS is an index based on three different indicators: a scale score derived from 23 self-report items on the presence of home possessions and another two indices based on self-report items asking about their parents' educational and occupational levels, i.e. items that are very different from typical rating scale items, asking about the degree of agreement with statements about attitudes, values and beliefs that are common across the student questionnaire. The same is true for the items measuring mathematics self-efficacy: asking about the student's confidence about doing things is very different from the items asking to indicate agreement.

Finally, a negative association was expected for the relationship between the performance in the mathematics test and mathematics anxiety (ANXMAT). Controlling for ARS should have made this association stronger, i.e. more negative. The opposite pattern, however, occurred across the vast majority of countries: controlling for ARS caused weaker associations. The pattern is likely due to the fact that the items were all negatively keyed (see example item in Table 9).

Figure 3. Regression weights for the prediction of mathematics performance and six variables of interest
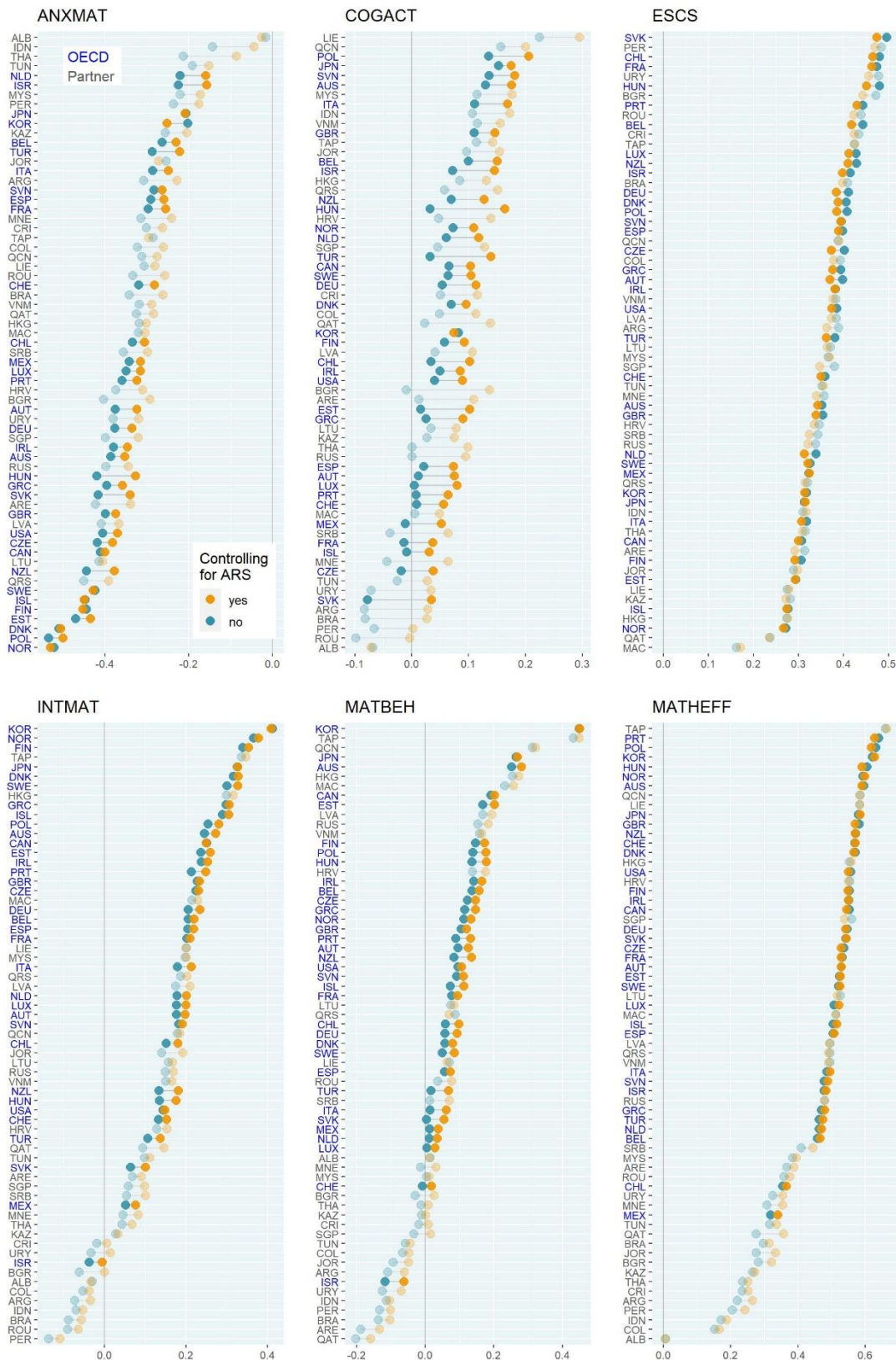
Chapter 6. Consistency across countries (RQ 4)

While the previous sections covered the analysis of ARS within each of the questionnaires individually, Research question 4 investigates whether these results are consistent across studies and countries. The stability and generalizability of previous findings would only hold if respondents of a particular population in one country show similar levels of ARS across multiple questionnaires. As the same or comparable populations have been administered in multiple studies, it is possible to derive specific expectations about the consistency of ARS across countries. In particular, similar levels of country-ARS should hold for the following populations across studies:

•   Students in PISA 2018, PISA 2015, PISA 2012, and SSES

•   Teachers in PISA 2018, PISA 2015, TALIS 2018, TALIS 2013, and SSES

•   School principals in PISA 2012, TALIS 2018, and TALIS 2013

•   Adults in general (PAQ, TCQ, SCQ)

6.1. Method

To relate the level of ARS approximated in multiple studies and countries to one another, the country-level mean ARS was computed for each questionnaire and country (seeChapter 7. Annex A). Whenever available, the data were weighted based on individual-level weights (see Chapter 7. Annex C). To align SSES data to all other studies, nine of the ten sites were recoded to represent the country in which the particular site is located in (see Table 10 for details). Next, a linear correlation was computed for all pair-wise combinations of questionnaires with on overlap of at least five countries. This implied, however, that data from IELS could not be included in the analysis as only three countries participated. Positive correlation coefficients are indicative of consistent levels of ARSI across countries and studies.

Both the computation of country means and correlation coefficients were conducted in R (version 4.1.0) (R Core Team, 2021[39]). Weighted country means were computed using the R-package Hmisc (version 4.5-0) (Harrell and Dupont, 2021[40]).

Table 10. Recoding of SSES grouping variable "SiteID" to represent the country of administration

SiteID	Site	Country name	ISO Alpha-3 code
1	Ottawa	Canada	CAN
2	Houston	United States	USA
3	Bogota	Colombia	COL
4	Manizales	Colombia	COL
5	Helsinki	Finland	FIN
6	Moscow	Russian Federation	RUS
7	Istanbul	Turkey	TUR
8	Daegu	Korea	KOR
9	Sintra	Portugal	PRT
10	Suzhou	-	-

Note: As only regions within the People's Republic of China have participated in the assessments under investigation, Suzhou could not be recoded to match a country.

6.2. Results

Table 11 shows the pair-wise correlations of ARS across studies (lower triangular) as well as the number of countries the calculation is based on (upper triangular).

In line with expectations, the vast majority of correlation coefficients are positive, thus indicating a strong consistency of country-level ARS across studies. The few exceptions occurred for combinations of studies in which the computation of ARSI in at least one study was based on only two or three item pairs (i.e. PISA 2018 TCQ, PISA 2015 TCQ, and PISA 2012 SCQ). When ignoring these studies, the correlation coefficients range from -.337 (SSES-STQ with SSES-TCQ) to .844 (PISA 2018-STQ with SSES-STQ), with a median correlation of $r = .411$.

As the table is organised according to the respondent's population, highest correlations should occur within the blocks representing students, teachers, school principals and parents, respectively. Indeed, findings support this expectation for students ($.208 < r \leq .844$, $Md = .485$), teachers ($.476 < r \leq .931$, $Md = .857$), and school principals ($.260 < r \leq .593$, $Md = .323$). No correlation could be computed among the two parent questionnaires due to an insufficient number of countries participating in both studies ($N = 2$). In addition and in line with expectations, the correlations among all adults (teachers, school principals, parents; $-.390 < r \leq .931$, $Md = .537$) are much higher on average than the correlations between adults and students ($-.481 < r \leq .760$, $Md = .243$).

Table 11. Correlation of ARSI across countries: correlation coefficient (lower triangular) and number of data points (upper triangular)

		STQ				TCQ					SCQ			PAQ	
	PISA 2018	PISA 2015	PISA 2012	SSES	PISA 2018	PISA 2015	TALIS 2018	TALIS 2013	SSES	PISA 2012	TALIS 2018	TALIS 2013	PISA 2018	SSES
STQ														
P 18		62	61	8	19	17	40	31	8	61	40	31	17	8
P 15	.523		56	8	14	18	37	28	8	56	37	28	16	8
P 12	.446	.775		8	15	16	37	30	8	65	37	30	13	8
SSES	.844	.265	.208		(3)	(4)	7	5	9	8	7	5	(2)	9
TCQ														
P 18	.605	.350	-.138	-		13	7	7	(3)	15	7	7	9	(3)
P 15	.386	.243	-.481	-	.904		11	9	(4)	16	11	9	9	(4)
T 18	.203	.176	.001	.316	.931	.918		32	7	37	48	32	10	7
T 13	.078	.160	.144	.494	.904	.794	.721		5	30	32	36	8	5
SSES	.461	.407	.225	-.337	-	-	.476	.810		8	7	5	(2)	9
SCQ														
P 12	.477	.442	.476	.202	.537	.329	.161	.413	-.390		37	30	13	8
T 18	.027	.172	.116	.328	.776	.567	.707	.677	.572	.323		32	10	7
T 13	.106	.067	.337	.218	.619	.313	.467	.498	.824	.260	.593		8	5
PAQ														
P 18	.460	.312	.172	-	.789	.541	.204	.614	-	.516	.500	.415		(2)
SSES	.760	.686	.672	.556	-	-	.305	.297	.583	-.046	.532	.291	-	

Note: Questionnaire administered to students (STQ), teachers (TCQ), school principals (SCQ) and parents (PAQ); P 18: PISA 2018; P 15: PISA 2015; P 12: PISA 2012; T 18: TALIS 2018; T 13: TALIS 2013. Correlation coefficients based on less than 5 countries are omitted.

Chapter 7. Discussion

Evidence from prior research suggests that response styles such as acquiescence affect self-report questionnaire data. The aim of this working paper was to examine the prevalence, correlates, impact and consistency of acquiescence across the questionnaires administered in recent OECD studies in the field of educational large-scale assessments as these heavily rely on self-reports.

Results regarding the research questions guiding the outline of this paper can be summarized as follows:

1. Variation in the acquiescence index ARSI existed both within and between countries.

2. In line with findings of prior research, evidence pointed at a negative relationship of ARSI with measures of educational attainment and, although not being quite as clear, with reading/verbal ability. Also in line with previous research, findings for age and gender were mixed. In addition, findings demonstrated a consistent negative association with socio-economic background. Finally, parents, especially mothers, were particularly prone to acquiesce, although gender and population are confounded as the vast majority of parents responding to the questionnaires were female.

3. Controlling for ARS resulted in stronger associations between test performance and scaled scores under two conditions: when scales are based on typical, agreement-based items measuring attitudes, values, and beliefs; and when these items are not negatively keyed.

4. The country-level ARS was consistent across studies for students, parents, teachers, and among adults in general. As correlations were higher among adults than between adults and students, the finding provides evidence for age-related differences in ARS. Within populations, this pattern has not emerged, but can be supported by examining the data across populations.

All analyses were based on a simple indicator capturing the degree of acquiescence which can easily be computed in the presence of mixed-worded scales based on antonym pairs tapping on the same construct. As such, the indicator was particularly useful for the purpose of this working paper as the study programmes (and studies therein) differ quite a bit in the principles guiding the development and design of questionnaires, potentially due to changing contractors in charge of these studies. A thorough examination of all questionnaires revealed that the number of mixed-worded scales ranged from none in PIAAC to almost all in SSES where the construction of such scales for measuring social and emotional skills was explicitly desired. Making such mechanisms transparent holds great promise to initiate a scientific debate. Two recent developments that address this call for transparency are particularly encouraging: first, the published framework guiding the development of the questionnaires for PISA 2022 contained very concrete principles on item wording (ETS, 2019[42]); second, the OECD's PISA Research, Development and Innovation (RDI) Programme was established to support innovation in the design of future assessments. Among three other projects, one project is designated to develop comprehensive Quality Standards for PISA tests and questionnaires (project 1; (OECD, 2021[43])), thus promising greater consistency across cycles in the future. For the study at hand, 16 questionnaires were identified containing mixed-worded scales, however some of them contained too few antonym pairs to show meaningful, consistent findings

(particularly PISA 2018 TCQ, PISA 2015 TCQ, and PISA 2012 SCQ), providing evidence to conclude that two or three item pairs are not sufficient for the construction of the index.

Although being a necessity for constructing the index underlying all analyses in this paper, the use of mixed-worded items for substantive research is subject to a controversial debate. One concern is that students fail to notice the changing wording – a concern strongly supported by the regularly found association between ARS and reading skills. Another potential explanation for this relationship could be disengagement or careless responding (e.g. Schmitt and Stuits (1985[9])), as such behaviour will likely affect both variables simultaneously in an assessment situation. This is particularly relevant in the context of low-stakes assessments such as the studies under investigation in this paper where no personal consequences are associated with the results of the assessment. For students responding to the PISA questionnaire, engagement is especially relevant due to the fact that the (at least) 35-minute long questionnaire is administered after a two-hour long assessment. At the same time, developing measures to capture engagement using Likert-type items is challenging as, unlike in achievement tests, no objective criterion for validation purposes is available (e.g. Soland, Wise and Gao (2019[44])). Another encouraging development addressing the issue of engagement is the RDI project dedicated to this particular challenge making use of, among others, log-file data from computer-based assessments (project 3; OECD (2021[43])).

The aim of this paper was to gather descriptive information about acquiescence across OECD studies. Whether the effects of acquiescence on scale scores and, subsequently, substantive findings relevant for policy-making while yielding valid cross-country comparisons, can be statistically controlled for, is a question that needs to be addressed in a separate study. Methods for controlling the effects of acquiescence are subject to ongoing research. For the time being, those in charge of developing questionnaires for operational use in international large-scale assessments need to know whether to include mixed-worded scales or not. The approach described by Herbert Marsh more than three decades ago provides a pragmatic solution:

> *[He] specifically chose to include a few negatively worded items on his widely used preadolescent self-concept instrument to disrupt potential response biases even though there was clear evidence that young children have difficulty responding appropriately to the negatively worded items. However, these negatively worded items were not used in the scores based on responses to this instrument. This approach had the added benefit of allowing Marsh (1986) to test for potential response biases (Marsh, 1996, p. 817[17])*

Accordingly, it would be recommended to include a few mixed-worded item pairs in questionnaires to disrupt potential response biases and to allow for a post-hoc analysis of the degree of acquiescence in the data. While the four OECD study programmes which are subject to this study fulfil this to a large degree (with a few exceptions, i.e. School, Teacher and Parent Questionnaires across the different waves of PISA, see Table 1), not a single mixed-worded item pair could be identified in the questionnaires administered in PIAAC.

Another aspect of Marsh's recommendation is to exclude the inversely-worded items from scaling. For the process of scale construction, this implies to develop a sufficient number of equally-keyed items to measure the construct of interest. For operational testing, including inversely-keyed items in the questionnaire means additional testing time – a valuable resource. Based on the findings of this study, acquiescence, as measured by ARSI, is recognizable with as few as five mixed-worded item pairs (SSES Teacher Questionnaire). Thus, an additional number of five items would need to be included in a given questionnaire. Although this corresponds to the typical length of a matrix question, these inversely-keyed items can be spread over the total length of the questionnaire, mixed

in with other items of matrix questions. Additional reading time is therefore only needed for the item itself, not for reading an additional question stem. In light of the advantages of including inversely-keyed items, the additional testing time for administering these items can, therefore, be justified.

References

Danner, D., J. Aichholzer and B. Rammstedt (2015), "Acquiescence in personality questionnaires: Relevance, domain specificity, and stability", *Journal of Research in Personality*, Vol. 57, pp. 119–130, http://dx.doi.org/10.1016/j.jrp.2015.05.004. [10]

Dunbar, M. et al. (2000), "Question wording effects in the assessment of global self-esteem", *European Journal of Psychological Assessment*, Vol. 16/1, pp. 13-19, http://dx.doi.org/10.1027//1015-5759.16.1.13. [22]

Eid, M. and M. Rauber (2000), "Detecting measurement invariance in organizational surveys", *European Journal of Psychological Assessment*, The original data upon which this paper is based are available at www.hhpub.com/journals/ejpa, pp. 20-30, http://dx.doi.org/10.1027//1015-5759.16.1.20. [25]

ETS (2019), *PISA 2021 Context Questionnaires Framework*, https://www.oecd.org/pisa/sitedocument/PISA-2021-questionnaire-framework.pdf (accessed on 29 September 2021). [42]

Gnambs, T. and U. Schroeders (2020), "Cognitive abilities explain wording effects in the Rosenberg self-esteem scale", *Assessment*, Vol. 27/2, pp. 404–418, http://dx.doi.org/10.1177/1073191117746503. [21]

Haladyna, T., S. Downing and M. Rodriguez (2002), "A review of multiple-choice item-writing guidelines for classroom assessment", *Applied Measurement in Education*, Vol. 15/3, pp. 309–333. [16]

Harrell, F. and C. Dupont (2021), *Hmisc: Harrell Miscellaneous (Version 4.5-0)*, https://cran.r-project.org/package=Hmisc. [40]

IBM Corp. (2019), *IBM SPSS Statistics for Windows (Version 26.0)*. [41]

Johnson, T. et al. (2005), "The relation between culture and response styles", *Journal of Cross-Cultural Psychology*, Vol. 36/2, pp. 264–277, http://dx.doi.org/10.1177/0022022104272905. [28]

Marin, G., R. Gamba and B. Marin (1992), "Extreme response style and acquiescence among Hispanics. The role of acculturation and education", Vol. 23/4, pp. 498–509. [26]

Marsh, H. (1996), "Positive and negative global self-esteem: A substantively meaningful distinction or artifactors", *Journal of Personality and Social Psychology*, Vol. 70/4, pp. 810–819, http://dx.doi.org/10.1037//0022-3514.70.4.810. [17]

McCrae, R. (2018), "Method biases in single-source personality assessments", *Psychological Assessment*, Vol. 30/9, pp. 1160–1173, http://dx.doi.org/10.1037/pas0000566. [3]

Meisenberg, G. and A. Williams (2008), "Are acquiescent and extreme response styles related to low intelligence and education?", *Personality and Individual Differences*, Vol. 44/7, pp. 1539–1550, http://dx.doi.org/10.1016/j.paid.2008.01.010. [18]

Narayan, S. and J. Krosnick (1996), "Education moderates some response effects in attitude measurement", *Public Opinion Quarterly*, Vol. 60, pp. 58-88, http://dx.doi.org/10.1086/297739. [19]

OECD (2021), *Innovation in PISA: Frequently Asked Questions (FAQs)*, http:////www.oecd.org/pisa/innovation/Frequently-Asked-Questions.pdf (accessed on 29 September 2021). [43]

OECD (2021), *OECD Survey on Social and Emotional Skills*, https://www.oecd.org/education/ceri/social-emotional-skills-study/sses-technical-report.pdf (accessed on 29 September 2021). [37]

OECD (2021), *OECD Survey on Social and Emotional Skills Technical Report*, https://www.oecd.org/education/ceri/social-emotional-skills-study/sses-technical-report.pdf (accessed on 31 January 2022). [34]

OECD (2021), *PISA 2018 Technical Report*, https://www.oecd.org/pisa/data/pisa2018technicalreport/ (accessed on 29 September 2021). [29]

OECD (2020), *Early Learning and Child Well-being*, https://read.oecd.org/10.1787/3990407f-en?format=pdf (accessed on 29 September 2021). [35]

OECD (2019), *TALIS 2018 Technical Report*, https://www.oecd.org/education/talis/TALIS_2018_Technical_Report.pdf (accessed on 29 September 2021). [32]

OECD (2017), *PISA 2015 Technical Report*, https://www.oecd.org/pisa/data/2015-technical-report/PISA2015_TechRep_Final.pdf (accessed on 29 September 2021). [30]

OECD (2014), *PISA 2012 Technical Report*, https://www.oecd.org/pisa/pisaproducts/PISA-2012-technical-report-final.pdf (accessed on 29 September 2021). [31]

OECD (2014), *TALIS 2013 Technical Report*, https://www.oecd.org/education/school/TALIS-technical-report-2013.pdf (accessed on 29 September 2021). [33]

OECD (2013), *Technical Report of the Survey of Adult Skills (PIAAC)*, https://www.oecd.org/skills/piaac/_Technical%20Report_17OCT13.pdf (accessed on 29 September 2021). [36]

Primi, R. et al. (2019), "Comparison of classical and modern methods for measuring and correcting for acquiescence", *The British Journal of Mathematical and Statistical Psychology*, Vol. 72/3, pp. 447–465, http://dx.doi.org/10.1111/bmsp.12168. [1]

Quilty, L., J. Oakman and E. Risko (2006), "Correlates of the Rosenberg self-esteem scale method effects", *Structural Equation Modeling: A Multidisciplinary Journal*, Vol. 13/1, pp. 99–117, http://dx.doi.org/10.1207/s15328007sem1301_5. [23]

R Core Team (2021), *R: A Language and Environment for Statistical Computing (Version 4.1.0)*, https://www.r-project.org/. [39]

Rammstedt, B., D. Danner and M. Bosnjak (2017), "Acquiescence response styles: A multilevel model explaining individual-level and country-level differences", *Personality and Individual Differences*, Vol. 107, pp. 190–194, http://dx.doi.org/10.1016/j.paid.2016.11.038. [7]

Rammstedt, B., L. Goldberg and I. Borg (2010), "The measurement equivalence of Big Five factor markers for persons with different levels of education", *Journal of Research in Personality*, Vol. 44/4, pp. 53–61, http://dx.doi.org/10.1016/j.jrp.2009.10.005. [8]

Rammstedt, B. and C. Kemper (2011), "Measurement equivalence of the Big Five: Shedding further light on potential causes of the educational bias", *Journal of Research in Personality*, Vol. 45/1, pp. 121–125, http://dx.doi.org/10.1016/j.jrp.2010.11.006. [20]

Rammstedt, B., C. Kemper and I. Borg (2013), "Correcting Big Five personality measurements for acquiescence: An 18–country cross–cultural study", *European Journal of Personality*, Vol. 27/1, pp. 71–81, http://dx.doi.org/10.1002/per.1894. [4]

Rutkowski, L. et al. (2010), "International large-scale assessment data", *Educational Researcher*, Vol. 39/2, pp. 142–151, http://dx.doi.org/10.3102/0013189X10363170. [38]

Schmitt, N. and D. Stuits (1985), "Factors defined by negatively keyed items: The result of careless respondents?", *Applied Psychological Measurement*, Vol. 9/4, pp. 367–373, http://dx.doi.org/10.1177/014662168500900405. [9]

Soland, J., S. Wise and L. Gao (2019), "Identifying disengaged survey responses: New evidence using response time metadata", *Applied Measurement in Education*, Vol. 32/2, pp. 151–165, http://dx.doi.org/10.1080/08957347.2019.1577244. [44]

Soto, C. et al. (2008), "The developmental psychometrics of big five self-reports: acquiescence, factor structure, coherence, and differentiation from ages 10 to 20", *Journal of Personality and Social Psychology*, Vol. 94/4, pp. 718–737, http://dx.doi.org/10.1037/0022-3514.94.4.718. [11]

Steinmann, I., R. Strietholt and J. Braeken (2021), "A constrained factor mixture analysis model for consistent and inconsistent respondents to mixed-worded scales", *Psychological Methods*, p. Epub ahead of print., http://dx.doi.org/10.1037/met0000392. [13]

Suárez-Álvarez, J. et al. (2018), "Using reversed items in Likert scales: A questionable practice", *Psicothema*, Vol. 30/2, pp. 149–158, http://dx.doi.org/10.7334/psicothema2018.33. [14]

Ten Berge, J. (1999), "A legitimate case of component analysis of ipsative measures, and partialling the mean as an alternative to ipsatization", *Multivariate Behavioral Research*, Vol. 34/1, pp. 89–102, http://dx.doi.org/10.1207/s15327906mbr3401_4. [12]

van de Vijver, F., D. van Hemert and Y. Poortinga (eds.) (2007), *Acquiescence, extreme response bias and culture: A multilevel analysis*, Taylor and Francis. [5]

van Herk, H., Y. Poortinga and T. Verhallen (2004), "Response styles in rating scales", *Journal of Cross-Cultural Psychology*, Vol. 35/3, pp. 346–360, http://dx.doi.org/10.1177/0022022104264126. [27]

van Vaerenbergh, Y. and T. Thomas (2013), "Response styles in survey research: A literature review of antecedents, consequences, and remedies", *International Journal of Public Opinion Research*, Vol. 25/2, pp. 195–217, http://dx.doi.org/10.1093/ijpor/eds021. [2]

Weijters, B. and H. Baumgartner (2012), "Misresponse to reversed and negated items in surveys: A review", *Journal of Marketing Research*, Vol. 49/5, pp. 737–747, http://dx.doi.org/10.1509/jmr.11.0368. [15]

Weijters, B., M. Geuens and N. Schillewaert (2010), "The stability of individual response styles", *Psychological Methods*, Vol. 15/1, pp. 96–110, http://dx.doi.org/10.1037/a0018721. [24]

Wetzel, E. (2016), "The stability of extreme response style and acquiescence over 8 years", *Assessment*, Vol. 23/3, pp. 279–291, http://dx.doi.org/10.1177/1073191115583714. [6]

Annex A. Details on countries: Study participation and mean ARS

Table A.1. Details on countries: Study participation and mean ARS

CNT	Name	OECD	PISA 2018			PISA 2015		PISA 2012		TALIS 2018		TALIS 2013		SSES			IELS	
			STQ	PAQ	TCQ	STQ	TCQ	STQ	SCQ	TCQ	SCQ	TCQ	SCQ	STQ	PAQ	TCQ	PAQ	TCQ
ALB	Albania	0	2.572	NA	2.594	NA	NA	2.665	3.398	NA	NA	NA	NA	NA	NA	NA	NA	NA
ARE	United Arab Emirates	0	2.585	NA	2.549	2.424	2.488	2.702	3.412	2.535	2.516	NA	NA	NA	NA	NA	NA	NA
ARG	Argentina	0	2.490	NA	NA	NA	NA	2.737	3.025	NA	NA	NA	NA	NA	NA	NA	NA	NA
AUS	Australia	1	2.427	NA	NA	2.427	2.507	2.615	2.996	2.516	2.572	2.613	2.525	NA	NA	NA	NA	NA
AUT	Austria	1	2.469	NA	NA	2.333	NA	2.534	2.870	2.391	2.407	NA	NA	NA	NA	NA	NA	NA
BEL	Belgium	1	2.310	2.138	NA	2.302	NA	2.570	2.787	2.412	2.340	NA	NA	NA	NA	NA	NA	NA
BGR	Bulgaria	0	2.517	NA	NA	2.440	NA	2.694	3.195	2.429	2.387	2.561	2.474	NA	NA	NA	NA	NA
BIH	Bosnia and Herzegovina	0	2.531	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BLR	Belarus	0	2.443	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
BRA	Brazil	0	2.480	2.143	2.446	2.421	2.375	2.728	2.972	2.430	2.413	2.442	2.492	NA	NA	NA	NA	NA
BRN	Brunei Darussalam	0	2.616	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
CAN	Canada	1	2.525	NA	NA	2.438	NA	2.592	3.031	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHE	Switzerland	1	2.466	NA	NA	2.281	NA	2.511	2.953	NA	NA	NA	NA	NA	NA	NA	NA	NA
CHL	Chile	1	2.488	2.279	2.500	2.378	2.406	2.756	3.084	2.461	2.457	2.592	2.582	NA	NA	NA	NA	NA
COL	Colombia	1	2.559	NA	NA	2.460	2.434	2.733	3.109	2.497	2.559	NA	NA	NA	NA	NA	NA	NA
CRI	Costa Rica	0	2.497	NA	NA	2.434	NA	2.776	3.132	NA	NA	NA	NA	NA	NA	NA	NA	NA
CZE	Czech Republic	1	2.525	NA	NA	2.398	2.327	2.616	2.932	2.348	2.337	2.452	2.401	NA	NA	NA	NA	NA
DEU	Germany	1	2.443	2.068	2.355	2.292	2.325	2.541	2.801	NA	NA	NA	NA	NA	NA	NA	NA	NA
DNK	Denmark	1	2.439	NA	NA	2.312	NA	2.542	3.144	2.483	2.545	2.618	2.502	NA	NA	NA	NA	NA
DOM	Dominican Republic	0	2.549	2.407	2.534	2.504	2.493	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ESP	Spain	1	2.496	NA	2.476	2.337	2.420	2.644	2.999	2.490	2.442	2.498	2.518	NA	NA	NA	NA	NA

EST	Estonia	1	2.452	NA	NA	2.340	NA	2.568	3.229	2.381	2.333	2.465	2.316	NA	NA	NA	3.159	3.052
FIN	Finland	1	2.450	NA	NA	2.331	NA	2.505	2.859	2.478	2.504	2.552	2.450	NA	NA	NA	NA	NA
FRA	France	1	2.464	NA	NA	2.301	NA	2.574	2.828	2.407	2.362	2.480	2.688	NA	NA	NA	NA	NA
GBR	United Kingdom	1	2.535	NA	2.493	2.374	NA	2.559	3.107	NA	NA	NA	NA	NA	NA	NA	NA	NA
GEO	Georgia	0	2.447	2.522	NA	2.287	NA	NA	NA	2.483	2.554	2.606	2.516	NA	NA	NA	NA	NA
GRC	Greece	1	2.511	NA	NA	2.400	NA	2.670	3.053	NA	NA	NA	NA	NA	NA	NA	NA	NA
HKG	Hong Kong	0	2.592	2.344	2.595	2.462	2.584	2.578	3.000	NA	NA	NA	NA	NA	NA	NA	NA	NA
HRV	Croatia	0	2.510	2.197	NA	2.355	NA	2.622	2.929	2.387	2.381	2.508	2.397	NA	NA	NA	NA	NA
HUN	Hungary	1	2.449	NA	NA	2.364	NA	2.614	3.242	2.418	2.454	NA	NA	NA	NA	NA	NA	NA
IDN	Indonesia	0	2.666	NA	NA	2.423	NA	2.753	3.281	NA	NA	NA	NA	NA	NA	NA	NA	NA
IRL	Ireland	1	2.511	2.208	NA	2.378	NA	2.566	3.096	NA	NA	NA	NA	NA	NA	NA	NA	NA
ISL	Iceland	1	2.470	NA	NA	2.371	NA	2.580	3.123	NA	NA	NA	NA	NA	NA	NA	NA	NA
ISR	Israel[2]	1	2.444	NA	NA	NA	NA	2.570	3.124	2.443	2.488	2.547	2.487	NA	NA	NA	NA	NA
ITA	Italy	1	2.366	2.227	NA	2.283	2.303	2.592	2.916	2.408	2.416	2.476	2.506	NA	NA	NA	NA	NA
JOR	Jordan	0	2.599	NA	NA	2.558	NA	2.906	3.213	NA	NA	NA	NA	NA	NA	NA	NA	NA
JPN	Japan	1	2.500	NA	NA	2.336	NA	2.496	2.664	2.362	2.261	2.514	2.307	NA	NA	NA	NA	NA
KAZ	Kazakhstan	0	2.502	NA	NA	NA	NA	2.623	3.392	2.430	2.446	NA	NA	NA	NA	NA	NA	NA
KOR	Korea	1	2.499	2.237	2.579	2.284	2.631	2.411	3.008	2.570	2.486	2.602	2.516	NA	NA	NA	NA	NA
KSV	Kosovo	0	2.581	NA	NA	2.424	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LBN	Lebanon	0	2.703	NA	NA	2.467	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
LTU	Lithuania	1	2.486	NA	NA	2.359	NA	2.575	3.281	2.502	2.718	NA	NA	NA	NA	NA	NA	NA
LUX	Luxembourg	1	2.477	2.140	NA	2.304	NA	2.519	2.956	NA	NA	NA	NA	NA	NA	NA	NA	NA
LVA	Latvia	1	2.457	NA	NA	2.378	NA	2.590	3.152	2.397	2.463	2.492	2.451	NA	NA	NA	NA	NA
MAC	Macao	0	2.514	2.406	2.558	2.379	2.509	2.562	2.974	NA	NA	NA	NA	NA	NA	NA	NA	NA
MAR	Morocco	0	2.489	NA	2.516	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MDA	Moldova	0	2.499	NA	NA	2.356	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
MEX	Mexico	1	2.523	2.326	NA	2.420	NA	2.803	3.131	2.441	2.586	2.616	2.663	NA	NA	NA	NA	NA
MKD	North Macedonia	0	2.725	NA	NA	2.423	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

---

[2] The statistical data for Israel are supplied by and under the responsibility of the relevant Israeli authorities. The use of such data by the OECD is without prejudice to the status of the Golan Heights, East Jerusalem and Israeli settlements in the West Bank under the terms of international law.

Code	Country																	
MLT	Malta	0	2.546	2.560	NA	2.382	NA	NA	NA	2.465	2.426	NA	NA	NA	NA	NA	NA	NA
MNE	Montenegro	0	2.466	NA	NA	2.343	NA	2.596	3.150	NA	NA	NA	NA	NA	NA	NA	NA	NA
MYS	Malaysia	0	2.594	NA	2.599	NA	NA	2.727	3.383	NA	NA	2.660	2.704	NA	NA	NA	NA	NA
NLD	Netherlands	1	2.413	NA	NA	2.307	NA	2.556	2.860	2.454	2.480	2.541	2.534	NA	NA	NA	NA	NA
NOR	Norway	1	2.387	NA	NA	2.331	NA	2.566	3.066	2.470	2.541	2.587	2.490	NA	NA	NA	NA	NA
NZL	New Zealand	1	2.419	NA	NA	2.432	NA	2.639	3.004	2.503	2.503	2.612	2.511	NA	NA	NA	NA	NA
PAN	Panama	0	2.566	2.369	2.561	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
PER	Peru	0	2.544	NA	2.526	2.416	2.400	2.729	3.086	NA	NA	NA	NA	NA	NA	NA	NA	NA
PHL	Philippines	0	2.724	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
POL	Poland	1	2.456	NA	NA	2.368	NA	2.559	3.093	NA	NA	2.581	2.585	NA	NA	NA	NA	NA
PRT	Portugal	1	2.490	2.154	2.500	2.376	2.465	2.664	3.068	2.507	2.451	2.564	2.450	NA	NA	NA	NA	NA
QAT	Qatar	0	2.572	NA	NA	2.441	NA	2.753	3.401	NA	NA	NA	NA	NA	NA	NA	NA	NA
QAZ	Baku (Azerbaijan)	0	2.565	NA	2.584	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QCI	B-S-J-Z (China)	0	2.623	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QMR	Moscow Region (RUS)	0	2.416	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QRT	Tatarstan (RUS)	0	2.440	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
ROU	Romania	0	2.497	NA	NA	2.360	NA	2.690	3.241	2.409	2.297	2.541	2.432	NA	NA	NA	NA	NA
RUS	Russian Federation	0	2.428	NA	NA	2.424	NA	2.659	3.295	2.367	2.366	2.545	2.499	NA	NA	NA	NA	NA
SAU	Saudi Arabia	0	2.513	NA	NA	NA	NA	NA	NA	2.523	2.480	NA	NA	NA	NA	NA	NA	NA
SGP	Singapore	0	2.429	NA	NA	2.455	NA	2.684	3.172	2.552	2.589	2.609	2.562	NA	NA	NA	NA	NA
SRB	Serbia	0	2.479	NA	NA	NA	NA	2.652	3.013	NA	NA	2.567	2.370	NA	NA	NA	NA	NA
SVK	Slovak Republic	1	2.464	NA	NA	2.409	NA	2.657	2.933	2.454	2.391	2.506	2.447	NA	NA	NA	NA	NA
SVN	Slovenia	1	2.469	NA	NA	2.348	NA	2.632	3.029	2.408	2.377	NA	NA	NA	NA	NA	NA	NA
SWE	Sweden	1	2.510	NA	NA	2.360	NA	2.569	3.016	2.446	2.465	2.504	2.476	NA	NA	NA	NA	NA
TAP	Chinese Taipei	0	2.482	NA	2.646	2.394	2.632	2.528	3.061	NA	NA	NA	NA	NA	NA	NA	NA	NA
THA	Thailand	0	2.676	NA	NA	2.445	NA	2.811	3.333	NA	NA	NA	NA	NA	NA	NA	NA	NA
TUR	Turkey	1	2.485	NA	NA	2.469	NA	2.717	2.885	2.511	2.521	NA	NA	NA	NA	NA	NA	NA
UKR	Ukraine	0	2.463	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
URY	Uruguay	0	2.470	NA	NA	2.465	NA	2.779	2.975	NA	NA	NA	NA	NA	NA	NA	NA	NA
USA	United States	1	2.548	NA	2.533	2.472	2.517	2.615	3.047	2.526	2.435	2.581	2.526	NA	NA	NA	3.206	2.882
VNM	Vietnam	0	2.535	NA	NA	2.251	NA	2.453	2.963	2.513	2.469	NA	NA	NA	NA	NA	NA	NA
DZA	Algeria	NA	NA	NA	NA	2.616	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QAR	Argentina (Ciudad Autónoma de Buenos)	NA	NA	NA	NA	2.374	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA

QCH	B-S-J-G (China)	NA	NA	NA	NA	2.392	2.613	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QES	Spain (Regions)	NA	NA	NA	NA	2.332	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QUC	Massachusettes (USA)	NA	NA	NA	NA	2.441	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QUD	Puerto Rico (USA)	NA	NA	NA	NA	2.417	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
QUE	North Carolina (USA)	NA	NA	NA	NA	2.473	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
TTO	Trinidad and Tobago	NA	NA	NA	NA	2.442	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
TUN	Tunisia	NA	NA	NA	NA	2.459	NA	2.797	3.183	NA	NA	NA	NA	NA	NA	NA	NA
LIE	Liechtenstein	NA	NA	NA	NA	NA	NA	2.511	2.875	NA	NA	NA	NA	NA	NA	NA	NA
QCN	Shanghai-China	NA	NA	NA	NA	NA	NA	2.530	3.029	NA	NA	NA	NA	NA	NA	NA	NA
QRS	Perm (Russian Federation)	NA	NA	NA	NA	NA	NA	2.660	3.217	NA	NA	NA	NA	NA	NA	NA	NA
ABA	Ciudad Autónoma de Buenos Aires (Argentina)	NA	NA	NA	NA	NA	NA	NA	NA	2.465	2.445	NA	NA	NA	NA	NA	NA
BFL	Belgium (Flemish)	NA	NA	NA	NA	NA	NA	NA	NA	2.429	2.402	2.505	2.449	NA	NA	NA	NA
CAB	Alberta (Canada)	NA	NA	NA	NA	NA	NA	NA	NA	2.535	2.592	2.619	2.619	NA	NA	NA	NA
CSH	Shanghai (China)	NA	NA	NA	NA	NA	NA	NA	NA	2.492	2.393	2.629	2.498	NA	NA	NA	NA
CYP	Cypr[3]us	NA	NA	NA	NA	NA	NA	NA	NA	2.475	2.541	NA	NA	NA	NA	NA	NA
ENG	England (United Kingdom)	NA	NA	NA	NA	NA	NA	NA	NA	2.491	2.576	2.608	2.535	NA	NA	3.203	2.882
TWN	Chinese Taipei	NA	NA	NA	NA	NA	NA	NA	NA	2.567	2.569	NA	NA	NA	NA	NA	NA
ZAF	South Africa	NA	NA	NA	NA	NA	NA	NA	NA	2.511	2.490	NA	NA	NA	NA	NA	NA
AAD	Abu Dhabi (United Arab Emirates)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	2.624	2.531	NA	NA	NA	NA
Bogota		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.249	3.209	3.079	NA	NA
Daegu		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.152	3.033	3.053	NA	NA
Helsinki		NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.139	3.031	2.968	NA	NA

[3] Note by Turkey:

The information in this document with reference to "Cyprus" relates to the southern part of the Island. There is no single authority representing both Turkish and Greek Cypriot people on the Island. Turkey recognises the Turkish Republic of Northern Cyprus (TRNC). Until a lasting and equitable solution is found within the context of the United Nations, Turkey shall preserve its position concerning the "Cyprus issue".

Note by all the European Union Member States of the OECD and the European Union:
The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of Cyprus.

Houston	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.207	3.154	3.063	NA	NA
Istanbul	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.104	3.139	3.159	NA	NA
Manizales	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.268	3.243	3.054	NA	NA
Moscow	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.118	3.023	2.953	NA	NA
Ottawa	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.180	3.060	2.999	NA	NA
Sintra	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.180	3.090	2.930	NA	NA
Suzhou	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	3.067	3.123	3.181	NA	NA
Sum of participating countries		80	17	19	71	18	65	65	48	48	36	36	10	10	10	3	3

Annex B. Details on ARS pairs: Item wording and descriptive statistics

Table B.1. Details on ARS pairs: Item wording and descriptive statistics

Study	QQQ	#	Item ID and wording	*M*	*SD*	*N*
PISA 2018	STQ	1	FL169Q01HA: I enjoy talking about money matters. FL169Q03HA: Money matters are not relevant for me right now.	2.414	0.580	158 989
PISA 2018	STQ	2	ST034Q01TA: I feel like an outsider (or left out of things) at school. ST034Q05TA: Other students seem to like me.	2.416	0.482	526 063
PISA 2018	STQ	3	ST034Q02TA: I make friends easily at school. ST034Q06TA: I feel lonely at school.	2.377	0.471	526 469
PISA 2018	STQ	4	ST034Q03TA: I feel like I belong at school. ST034Q04TA: I feel awkward and out of place in my school.	2.403	0.491	524 673
PISA 2018	STQ	5	ST160Q01IA: I read only if I have to. ST160Q03IA: I like talking about books with other people.	2.363	0.525	567 362
PISA 2018	STQ	6	ST160Q02IA: Reading is one of my favourite hobbies. ST160Q04IA: For me, reading is a waste of time.	2.162	0.488	566 730
PISA 2018	STQ	7	ST161Q01HA: I am a good reader. ST161Q06HA: I have always had difficulty with reading.	2.348	0.507	559 081
PISA 2018	STQ	8	ST161Q02HA: I am able to understand difficult texts. ST161Q08HA: I find it difficult to answer questions about a text.	2.432	0.481	557 822
PISA 2018	STQ	9	ST161Q03HA: I read fluently. ST161Q07HA: I have to read a text several times before completely understanding it.	2.706	0.529	555 549
PISA 2018	STQ	10	ST186Q02HA: Afraid ST186Q03HA: Cheerful	2.874	0.545	508 085
PISA 2018	STQ	11	ST186Q05HA: Happy ST186Q08HA: Sad	2.938	0.459	510 285
PISA 2018	STQ	12	ST186Q06HA: Scared ST186Q07HA: Lively	2.724	0.548	506 820
PISA 2018	STQ	13	ST186Q09HA: Proud ST186Q10HA: Miserable	2.544	0.553	507 473
PISA 2018	PAQ	1	PA158Q01HA: I read only if I have to. PA158Q03HA: I like talking about books with other people.	2.350	0.524	90 020
PISA 2018	PAQ	2	PA158Q02IA: Reading is one of my favourite hobbies. PA158Q04IA: For me, reading is a waste of time.	2.185	0.441	90 212
PISA 2018	TCQ	1	TC198Q01HA: The advantages of being a teacher clearly outweigh the disadvantages. TC198Q04HA: I regret that I decided to become a teacher.	2.312	0.416	89 007
PISA 2018	TCQ	2	TC198Q02HA: If I could decide again, I would still choose to work as a teacher. TC198Q06HA: I wonder whether it would have been better to choose another profession.	2.616	0.409	89 063
PISA 2018	TCQ	3	TC198Q03HA: I would like to change to another school if that were possible. TC198Q05HA: I enjoy working at this school.	2.632	0.441	88 883
PISA 2015	STQ	1	ST034Q01TA: I feel like an outsider (or left out of things) at school. ST034Q05TA: Other students seem to like me.	2.384	0.460	480 010
PISA 2015	STQ	2	ST034Q02TA: I make friends easily at school. ST034Q06TA: I feel lonely at school.	2.376	0.439	481 864
PISA 2015	STQ	3	ST034Q03TA: I feel like I belong at school. ST034Q04TA: I feel awkward and out of place in my school.	2.403	0.469	478 743
PISA 2015	TCQ	1	TC026Q01NA: The advantages of being a teacher clearly outweigh the	2.300	0.422	87 316

			disadvantages. TC026Q04NA: I regret that I decided to become a teacher.			
PISA 2015	TCQ	2	TC026Q02NA: If I could decide again, I would still choose to work as a teacher. TC026Q06NA: I wonder whether it would have been better to choose another profession.	2.615	0.410	87 156
PISA 2012	STQ	1	ST42Q01: I often worry that it will be difficult for me in mathematics classes. ST42Q07: I have always believed that mathematics is one of my best subjects.	2.553	0.540	308 325
PISA 2012	STQ	2	ST42Q02: I am just not good at mathematics. ST42Q06: I learn mathematics quickly.	2.508	0.431	307 898
PISA 2012	STQ	3	ST42Q04: I get good <grades> in mathematics. ST42Q10: I worry that I will get poor <grades> in mathematics.	2.779	0.557	307 111
PISA 2012	STQ	4	ST42Q08: I feel helpless when doing a mathematics problem. ST42Q09: In my mathematics class, I understand even the most difficult work.	2.269	0.518	307 423
PISA 2012	STQ	5	ST43Q01: If I put in enough effort I can succeed in mathematics. ST43Q06: I do badly in mathematics whether or not I study for my exams.	2.748	0.500	309 541
PISA 2012	STQ	6	ST85Q03: My teacher starts lessons on time. ST85Q04: The teacher has to wait a long time for students to <quiet down>.	2.708	0.553	307 376
PISA 2012	STQ	7	ST87Q01: I feel like an outsider (or left out of things) at school. ST87Q05: Other students seem to like me.	2.410	0.449	305 994
PISA 2012	STQ	8	ST87Q02: I make friends easily at school. ST87Q06: I feel lonely at school.	2.422	0.423	306 723
PISA 2012	STQ	9	ST87Q03: I feel like I belong at school. ST87Q04: I feel awkward and out of place in my school.	2.430	0.457	305 774
PISA 2012	STQ	10	ST88Q01: School has done little to prepare me for adult life when I leave school. ST88Q04: School has taught me things which could be useful in a job.	2.765	0.497	306 819
PISA 2012	STQ	11	ST88Q02: School has been a waste of time. ST88Q03: School has helped give me confidence to make decisions.	2.401	0.447	306 402
PISA 2012	STQ	12	ST91Q02: It is completely my choice whether or not I do well at school. ST91Q06: I perform poorly at school whether or not I study for my exams.	2.605	0.554	305 285
PISA 2012	STQ	13	IC22Q01: The computer is a very useful tool for my schoolwork. IC22Q07: Since anyone can upload information to the internet, it is in general not suitable to use it for schoolwork.	2.855	0.572	288 539
PISA 2012	STQ	14	IC22Q02: Doing my homework using a computer makes it more fun. IC22Q06: Using the computer for learning is troublesome.	2.572	0.583	288 657
PISA 2012	STQ	15	IC22Q04: The Internet is a great resource for obtaining information I can use for my school work. IC22Q08: Information obtained from the internet is generally too unreliable to be used for school assignments.	2.866	0.544	288 180
PISA 2012	SCQ	1	SC27Q01: Mathematics teachers are interested in trying new methods and teaching practices. SC27Q02: There is a preference among mathematics teachers to stay with well-known methods and practices.	2.938	0.428	17 481
PISA 2012	SCQ	2	SC28Q01: There is consensus among mathematics teachers that academic achievement must be kept as high as possible. SC28Q02: There is consensus among mathematics teachers that it is best to adapt academic standards to the students' levels and needs.	3.135	0.539	17 466
PISA 2012	SCQ	3	SC29Q01: There is consensus among mathematics teachers that the social and emotional development of the students is as important as their acquisition of mathematical skills and knowledge in mathematics classes. SC29Q02: There is consensus among mathematics teachers that the development of mathematical skills and knowledge in students is the most important objective in mathematics classes.	3.067	0.523	17 248
TALIS 2018	TCQ	1	TT3G53A: The advantages of being a teacher clearly outweigh the disadvantages TT3G53D: I regret that I decided to become a teacher.	2.276	0.423	251 197
TALIS 2018	TCQ	2	TT3G53B: If I could decide again, I would still choose to work as a teacher TT3G53F: I wonder whether it would have been better to choose another	2.562	0.410	251 021

			profession			
TALIS 2018	TCQ	3	TT3G53C: I would like to change to another school if that were possible TT3G53E: I enjoy working at this school	2.575	0.397	250 745
TALIS 2018	TCQ	4	TT3G41B: Students in this class take care to create a pleasant learning atmosphere TT3G41C: I lose quite a lot of time because of students interrupting the lesson	2.479	0.403	212 294
TALIS 2018	SCQ	1	TC3G44A: The advantages of this profession clearly outweigh the disadvantages TC3G44D: I regret that I decided to become a principal	2.285	0.434	14 999
TALIS 2018	SCQ	2	TC3G44B: If I could decide again, I would still choose this job/position TC3G44F: I wonder whether it would have been better to choose another profession.	2.512	0.429	15 000
TALIS 2018	SCQ	3	TC3G44C: I would like to change to another school if that were possible TC3G44E: I enjoy working at this school.	2.608	0.411	14 995
TALIS 2013	TCQ	1	TT2G41B: Students in this class take care to create a pleasant learning atmosphere. TT2G41B: I lose quite a lot of time because of students interrupting the lesson.	2.791	0.723	155 595
TALIS 2013	TCQ	2	TT2G46A: The advantages of being a teacher clearly outweigh the disadvantages TT2G46F: I wonder whether it would have been better to choose another profession.	2.508	0.444	177 754
TALIS 2013	TCQ	3	TT2G46B: If I could decide again, I would still choose to work as a teacher TT2G46D: I regret that I decided to become a teacher.	2.348	0.357	178 343
TALIS 2013	TCQ	4	TT2G46C: I would like to change to another school if that were possible TT2G46E: I enjoy working at this school.	2.584	0.379	177 726
TALIS 2013	TCQ	5	TT2M15A: I am able to ask questions that get students to think deeply about mathematics. TT2M15F: I have a hard time getting my students to understand underlying concepts in mathematics.	2.675	0.381	5 184
TALIS 2013	TCQ	6	TT2M15B: I have a hard time getting students interested in mathematics. TT2M15E: I am able to get my students to feel confident in mathematics.	2.562	0.339	5 185
TALIS 2013	TCQ	7	TT2M15C: I always know which of my students understand and which do not. TT2M15D: I find it hard to meet the needs of the individual students in my mathematics class.	2.738	0.445	5 190
TALIS 2013	TCQ	8	TT2M16C: Grades are a primary motivator for getting students to learn mathematics. TT2M16D: I would like my students to study mathematics because it is an interesting and worthwhile subject rather than only because they want good marks.	2.979	0.450	5 194
TALIS 2013	SCQ	1	TC2G39B: If I could decide again, I would still choose this job/position TC2G39D: I regret that I decided to become a principal.	2.391	0.385	10 317
TALIS 2013	SCQ	2	TC2G39C: I would like to change to another school if that were possible TC2G39E: I enjoy working at this school.	2.609	0.396	10 312
SSES	STQ	1	STA_ASS05: Dislike leading a team STA_ASS07: *Item removed from scaling, see* OECD (2021[34])	2.899	0.622	60 367
SSES	STQ	2	STA_COO02: Get along well with others STA_COO04: Start arguments with others.	3.091	0.608	60 675
SSES	STQ	3	STA_CRE01: Find new ways to do things STA_CRE07: Find it difficult to create new things.	3.189	0.627	60 590
SSES	STQ	4	STA_CRE02: Original, come up with new ideas STA_CRE08: *Item removed from scaling, see* OECD (2021[34])	3.030	0.630	60 329
SSES	STQ	5	STA_CRE03: *Item removed from scaling, see* OECD (2021[34]) STA_CRE06: Have a good imagination	3.117	0.602	60 539
SSES	STQ	6	STA_CUR05: Like learning new things STA_CUR06: Don't like learning	3.147	0.564	60 564
SSES	STQ	7	STA_EMO01: Not easily upset STA_EMO06: Have unpredictable emotions and moods	3.235	0.747	60 483
SSES	STQ	8	STA_EMO02: Keep my emotions under control	3.477	0.677	60 594

SSES	STQ		STA_EMO05: *Item removed from scaling, see* OECD (2021[34])			
SSES	STQ	9	STA_EMO03: Get mad easily STA_EMO07: Stay calm even in tense situations	2.995	0.721	60 573
SSES	STQ	10	STA_EMO04: Know how to control my anger STA_EMO08: Often feel angry	3.055	0.627	60 320
SSES	STQ	11	STA_EMP02: Important to me that my friends are okay STA_EMP08: *Item removed from scaling, see* OECD (2021[34])	3.523	0.675	60 481
SSES	STQ	12	STA_ENE01: Full of energy STA_ENE04: Have less energy than my classmates	3.115	0.575	60 655
SSES	STQ	13	STA_ENE06: Tire out quickly STA_ENE08: Maintain high energy throughout the day	3.141	0.671	60 347
SSES	STQ	14	STA_OPT01: Often feel sad STA_OPT07: A happy person	3.333	0.635	60 450
SSES	STQ	15	STA_OPT02: Believe good things will happen to me STA_OPT08: Expect bad things to happen	3.234	0.672	60 382
SSES	STQ	16	STA_PER02: Make sure that I finish tasks STA_PER03: Give up easily	3.007	0.625	60 648
SSES	STQ	17	STA_PER04: Finish what I start STA_PER05: Leave things unfinished	3.088	0.531	60 641
SSES	STQ	18	STA_PER06: *Item removed from scaling, see* OECD (2021[34]) STA_PER08: Finish things despite difficulties in the way	3.149	0.594	60 379
SSES	STQ	19	STA_RES02: *Item removed from scaling, see* OECD (2021[34]) STA_RES03: Often forget my duties	3.316	0.689	60 639
SSES	STQ	20	STA_RES04: Avoid responsibilities STA_RES06: A responsible person	3.077	0.572	60 445
SSES	STQ	21	STA_RES05: Keep my promises STA_RES07: *Item removed from scaling, see* OECD (2021[34])	3.323	0.585	60 623
SSES	STQ	22	STA_SEL01: Careful with what I say to others STA_SEL05: *Item removed from scaling, see* OECD (2021[34])	3.289	0.706	60 604
SSES	STQ	23	STA_SEL03: Think carefully before doing something STA_SEL08: Often rush into action without thinking	3.150	0.591	60 293
SSES	STQ	24	STA_SOC04: Like to be alone STA_SOC06: *Item removed from scaling, see* OECD (2021[34])	3.409	0.700	60 494
SSES	STQ	25	STA_SOC07: Make friends easily STA_SOC08: *Item removed from scaling, see* OECD (2021[34])	2.987	0.504	60 545
SSES	STQ	26	STA_STR01: Relaxed and handle stress well STA_STR02: Get nervous easily	3.297	0.710	60 677
SSES	STQ	27	STA_TOL06: Not interested in other countries and cultures STA_TOL08: *Item removed from scaling, see* OECD (2021[34])	2.982	0.576	60 384
SSES	STQ	28	STA_TRU05: Distrust people STA_TRU08: Trust others	3.066	0.572	60 477
SSES	PAQ	1	PAA_ASS05: Dislike leading a team PAA_ASS07: *Item removed from scaling, see* OECD (2021[34])	3.016	0.552	33 778
SSES	PAQ	2	PAA_COO02: Get along well with others PAA_COO04: Start arguments with others	3.086	0.526	34 407
SSES	PAQ	3	PAA_CRE01: Find new ways to do things PAA_CRE07: Find it difficult to create new things	3.103	0.553	33 903
SSES	PAQ	4	PAA_CRE02: Original, come up with new ideas PAA_CRE08: *Item removed from scaling, see* OECD (2021[34])	3.014	0.516	33 873
SSES	PAQ	5	PAA_CRE03: *Item removed from scaling, see* OECD (2021[34]) PAA_CRE06: Have a good imagination	3.120	0.495	34 068
SSES	PAQ	6	PAA_CUR05: Like learning new things PAA_CUR06: Don't like learning	3.133	0.534	34 080
SSES	PAQ	7	PAA_EMO01: Not easily upset PAA_EMO06: Have unpredictable emotions and moods	2.851	0.703	34 199
SSES	PAQ	8	PAA_EMO02: Keep my emotions under control PAA_EMO05: *Item removed from scaling, see* OECD (2021[34])	3.094	0.594	34 249
SSES	PAQ	9	PAA_EMO03: Get mad easily PAA_EMO07: Stay calm even in tense situations	3.072	0.635	34 092
SSES	PAQ	10	PAA_EMO04: Know how to control my anger	2.857	0.537	33 867

			PAA_EMO08: Often feel angry			
SSES	PAQ	11	PAA_EMP02: Important to me that my friends are okay PAA_EMP08: *Item removed from scaling, see* OECD (2021[34])	3.421	0.582	33 873
SSES	PAQ	12	PAA_ENE01: Full of energy PAA_ENE04: Have less energy than my classmates	3.106	0.53	34 309
SSES	PAQ	13	PAA_ENE06: Tire out quickly PAA_ENE08: Maintain high energy throughout the day	3.090	0.563	33 883
SSES	PAQ	14	PAA_RES02: *Item removed from scaling, see* OECD (2021[34]) PAA_RES03: Often forget my duties	3.426	0.597	34 318
SSES	PAQ	15	PAA_RES04: Avoid responsibilities PAA_RES06: A responsible person	3.198	0.484	34 043
SSES	PAQ	16	PAA_RES05: Keep my promises PAA_RES07: *Item removed from scaling, see* OECD (2021[34])	3.23	0.490	34 040
SSES	PAQ	17	PAA_OPT01: Often feel sad PAA_OPT07: A happy person	3.291	0.537	33 900
SSES	PAQ	18	PAA_OPT02: Believe good things will happen to me PAA_OPT08: Expect bad things to happen	3.163	0.555	33 883
SSES	PAQ	19	PAA_PER02: Make sure that I finish tasks PAA_PER03: Give up easily	3.054	0.567	34 167
SSES	PAQ	20	PAA_PER04: Finish what I start PAA_PER05: Leave things unfinished	3.103	0.451	34 216
SSES	PAQ	21	PAA_PER06: *Item removed from scaling, see* OECD (2021[34]) PAA_PER08: Finish things despite difficulties in the way	3.174	0.512	33 836
SSES	PAQ	22	PAA_SEL01: Careful with what I say to others PAA_SEL05: *Item removed from scaling, see* OECD (2021[34])	3.286	0.651	34 196
SSES	PAQ	23	PAA_SEL03: Think carefully before doing something PAA_SEL08: Often rush into action without thinking	3.023	0.506	33 837
SSES	PAQ	24	PAA_SOC04: Like to be alone PAA_SOC06: *Item removed from scaling, see* OECD (2021[34])	3.260	0.597	34 093
SSES	PAQ	25	PAA_SOC07: Make friends easily PAA_SOC08: *Item removed from scaling, see* OECD (2021[34])	3.012	0.442	33 925
SSES	PAQ	26	PAA_STR01: Relaxed and handle stress well PAA_STR02: Get nervous easily	3.185	0.612	34 271
SSES	PAQ	27	PAA_TOL06: Not interested in other countries and cultures PAA_TOL08: *Item removed from scaling, see* OECD (2021[34])	3.076	0.561	33 903
SSES	PAQ	28	PAA_TRU05: Distrust people PAA_TRU08: *Item removed from scaling, see* OECD (2021[34])	3.094	0.52	33 836
SSES	TCQ	1	TCA_COO02: Get along well with others TCA_COO04: Start arguments with others	3.004	0.578	50 872
SSES	TCQ	2	TCA_EMO02: Keep my emotions under control TCA_EMO06: Have unpredictable emotions and moods	3.081	0.504	50 976
SSES	TCQ	3	TCA_ENE01: Full of energy TCA_ENE04: Have less energy than my classmates	3.031	0.481	50 940
SSES	TCQ	4	TCA_RES05: Keep my promises TCA_RES07: *Item removed from scaling, see* OECD (2021[34])	3.128	0.446	50 915
SSES	TCQ	5	TCA_SEL03: Think carefully before doing something TCA_SEL05: *Item removed from scaling, see* OECD (2021[34])	3.008	0.547	50 956
IELS	PAQ	1	ELPAQ0702: Is helpful to other children (e.g. if someone is hurt or upset) ELPAQ0721: Fights with other children	3.094	0.487	4 943
IELS	PAQ	2	ELPAQ0704: Dislikes it when asked to play in a different way (e.g. frowns, stamps foot) ELPAQ0710: Is curious, likes to explore or try new things	3.564	0.590	4 849
IELS	PAQ	3	ELPAQ0705: Prevents other children from doing their own activities ELPAQ0713: Joins in with other children playing	3.144	0.525	4 905
IELS	PAQ	4	ELPAQ0707: Tries to comfort others when they are upset ELPAQ0728: Is unresponsive to people feeling sad	2.836	0.506	4 845
IELS	PAQ	5	ELPAQ0709: Is confident around adults ELPAQ0727: Is initially cautious with unfamiliar adults	3.761	0.596	4 828
IELS	PAQ	6	ELPAQ0711: Considers other people's feelings ELPAQ0714: Is unaware of other people's emotions	2.967	0.540	4 907

IELS	PAQ	7	ELPAQ0712: Says nice or friendly things to other children ELPAQ0719: Teases other children	2.822	0.490	4 849
IELS	PAQ	8	ELPAQ0717: Is hesitant when making requests ELPAQ0725: Openly approaches familiar adults when she/he needs help	3.303	0.552	4 859
IELS	TCQ	1	ELCHQ0602: Is helpful to other children (e.g. if someone is hurt or upset) ELCHQ0621: Fights with other children	2.826	0.529	6 378
IELS	TCQ	2	ELCHQ0604: Dislikes it when asked to play in a different way (e.g. frowns, stamps foot) ELCHQ0610: Is curious, likes to explore or try new things	2.974	0.662	6 233
IELS	TCQ	3	ELCHQ0605: Prevents other children from doing their own activities ELCHQ0613: Joins in with other children playing	2.957	0.549	6 379
IELS	TCQ	4	ELCHQ0607: Tries to comfort others when they are upset ELCHQ0628: Is unresponsive to people feeling sad	2.668	0.494	6 217
IELS	TCQ	5	ELCHQ0609: Is confident around adults ELCHQ0627: Is initially cautious with unfamiliar adults	3.323	0.524	6 178
IELS	TCQ	6	ELCHQ0611: Considers other people's feelings ELCHQ0614: Is unaware of other people's emotions	2.870	0.465	6 343
IELS	TCQ	7	ELCHQ0612: Says nice or friendly things to other children ELCHQ0619: Teases other children	2.693	0.493	6 282
IELS	TCQ	8	ELCHQ0617: Is hesitant when making requests ELCHQ0625: Openly approaches familiar adults when she/he needs help	3.163	0.513	6 317

Annex C. Details on datasets: Version, Grouping and Weight variables

Table C.1. Details on datasets: Version, Grouping and Weight variables

Study	QQQ	Version variable	Version	Group variable	Individual-level weight variable	Senate weight variable
PISA 2018	STQ	VER_DAT	2019-08-14	CNT	W_FSTUWT	SENWT
	PAQ	VER_DAT	2019-08-14	CNT	W_FSTUWT	SENWT
	TCQ	VER_DAT	2019-06-24	CNT	<NA>	*Simple senate weight (own construction)*
PISA 2015	STQ	VER_DAT	2016-11-17	CNT	W_FSTUWT	SENWT
	TCQ	VER_DAT	2016-11-15	CNT	<NA>	*Simple senate weight (own construction)*
PISA 2012	STQ	VER_STU	2014-06-11	CNT	W_FSTUWT	senwgt_STU
	SCQ	VER_SCQ	2013-11-28	CNT	W_FSCHWT	SENWGT_SCQ
TALIS 2018	TCQ	VERSION	310	CNTRY	TCHWGT	*constructed based on TCHWGT*
	SCQ	VERSION	310	CNTRY	SCHWGT	*own construction based on SCHWGT*
TALIS 2013	TCQ	VERSION	31	CNTRY	TCHWGT	*own construction based on TCHWGT*
	SCQ	VERSION	31	CNTRY	SCHWGT	*own construction based on SCHWGT*
SSES	STQ	<NA>	<NA>	SiteID	WT2019	*own construction based on WT2019*
	PAQ	<NA>	<NA>	SiteID	WT2019_PA	*own construction based on WT2019_PA*
	TCQ	<NA>	<NA>	SiteID	WT2019_TC	*own construction based on WT2019_TC*
IELS	PAQ	VERSIONDATE	2020-04-24	CNTRY	CHILDWGT	SENWGT
	TCQ	VERSIONDATE	2020-04-24	CNTRY	CHILDWGT	SENWGT