OECD Education Working Papers No. 133

Test-taking engagement in PIAAC

Frank Goldhammer,
Thomas Martens,
Gabriela Christoph,
Oliver Lüdtke

https://dx.doi.org/10.1787/5jlzfl6fhxs2-en

OECD

Unclassified **EDU/WKP(2016)7**

Organisation de Coopération et de Développement Économiques
Organisation for Economic Co-operation and Development **20-Apr-2016**

English - Or. English

DIRECTORATE FOR EDUCATION AND SKILLS

EDU/WKP(2016)7
Unclassified

TEST-TAKING ENGAGEMENT IN PIAAC

Education Working Paper No. 133

By Frank Goldhammer (DIPF and ZIB), Thomas Martens (MSH), Gabriela Christoph (DIPF), Oliver Lüdtke (IPN and ZIB)

DIPF – German Institute for International Educational Research
MSH – Medical School Hamburg
IPN – Leibniz Institute for Science and Mathematics Education
ZIB – Centre for International Student Assessment

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

Frank Goldhammer, DIPF (goldhammer@dipf.de)

JT03394459

Complete document available on OLIS in its original format

English - Or. English

OECD EDUCATION WORKING PAPERS SERIES

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

---

www.oecd.org/edu/workingpapers

TEST-TAKING ENGAGEMENT IN PIAAC

Abstract

In this study, we investigated how empirical indicators of test-taking engagement can be defined, empirically validated, and used to describe group differences in the context of the Programme of International Assessment of Adult Competences (PIAAC). The approach was to distinguish between disengaged and engaged response behavior by means of response time thresholds. Constant thresholds of 3000 ms and 5000 ms were considered, as well as item-specific thresholds based on the visual inspection of (bimodal) response time distributions (VI method) and the proportion correct conditional on response time (P+>0% method). Overall, the validity checks comparing the proportion correct of engaged and disengaged response behavior by domain and by item showed that the P+>0% method performed slightly better than the VI method and the methods assuming constant thresholds. The results for Literacy and Numeracy by module revealed that there was an increase from Module 1 to Module 2 in the proportion of disengaged responses, suggesting a drop in test-taking engagement. The investigation of country differences in test-taking engagement by domain using the P+>0% method showed that the proportion of responses classified as disengaged was quite low. For Literacy, the proportion was well below 5% for the majority of countries; in Numeracy, the proportion was even smaller than 1% for almost all countries; while for Problem solving, the proportion of disengaged responses was more than 5% but usually well below 10%. There were significant differences in test-taking engagement between countries; the obtained effect sizes were small to medium. Population differences in test-taking engagement were highly correlated between the three domains, suggesting that test-taking engagement can be conceived as a consistent characteristic. Furthermore, there was a clear negative association between test-taking disengagement and proficiency in Literacy, Numeracy and Problem solving, respectively. Finally, subgroup differences for gender, age, educational attainment, and language proved to be insignificant or very small. Results suggest that males tend to be more disengaged, that disengagement increases with age in Problem solving, with lower educational attainment and when the test language is not the same as a testee's native language.

Résumé

Dans cette étude, nous analysons comment des indicateurs empiriques de l'engagement dans les réponses à un test peuvent être définis, validés empiriquement et utilisés pour décrire des différences entre les groupes dans le cadre du PIAAC. L'approche utilisée consiste à établir une distinction entre un comportement de réponse « désengagé » et un comportement de réponse « engagé » sur la base de seuils de temps de réponse. Des seuils constants de 3000 ms et 5000 ms ont été retenus, ainsi que des seuils spécifiques en fonction des items, sur la base de l'inspection visuelle des distributions (bimodales) du temps de réponse (méthode VI) et du pourcentage de réponses correctes en fonction du temps de réponse (méthode P+>0%). Dans l'ensemble, les contrôles de validité comparant le pourcentage de réponses correctes des comportements « engagés » et « désengagés » par domaine et par item montrent que la méthode P+>0% donne des résultats légèrement meilleurs que la méthode VI et les méthodes fondées sur des seuils constants. Les résultats en littératie et en numératie par module indiquent une augmentation du pourcentage de réponses désengagées entre le module 1 et le module 2, suggérant une baisse de l'engagement dans les réponses au test. L'étude par domaine des différences d'engagement dans les réponses au test entre les pays, sur la base de la méthode P+>0%, met au jour un pourcentage assez faible de réponses pouvant être qualifiées de « désengagées ». En littératie, ce pourcentage est nettement inférieur à 5 % dans la majorité des pays ; en numératie, il est même inférieur à 1 % dans la quasi-totalité des pays ; en résolution de problèmes, en revanche, il est supérieur à 5 %, mais reste en général nettement inférieur à 10 %. Des différences significatives d'engagement dans les réponses au test s'observent entre les pays ; l'ampleur de l'effet va de faible à moyenne. Les différences d'engagement dans les réponses au test entre

les différents groupes sont fortement corrélées entre les trois domaines, semblant indiquer que l'engagement dans les réponses au test peut être perçu comme une caractéristique constante. En outre, une nette corrélation négative s'observe entre d'un côté, le désengagement dans les réponses au test, et de l'autre, les compétences en littératie, en numératie et en résolution de problèmes, respectivement. Enfin, les différences entre les sous-groupes en fonction du sexe, de l'âge, du niveau de formation et de la langue s'avèrent négligeables ou très faibles. D'après les résultats, les hommes tendent à être plus désengagés, et le désengagement augmente avec l'âge en résolution de problèmes, lorsque le niveau de formation des répondants est plus faible et lorsque leur langue maternelle n'est pas celle du test.

Reader's guide

A note regarding the Russian Federation:

Readers should note that the sample for the Russian Federation does not include the population of the Moscow municipal area. The data published, therefore, do not represent the entire resident population aged 16-65 in Russia but rather the population of Russia *excluding* the population residing in the Moscow municipal area. More detailed information regarding the data from the Russian Federation as well as that of other countries can be found in the *Technical Report of the Survey of Adult Skills*.

Country coverage:

Readers should also note that data presented in the tables for United Kingdom are referring to England and Northern Ireland, and data presented for Belgium are referring to Flanders.

TABLE OF CONTENTS

Tables

Figures

INTRODUCTION

This research study addresses the question of how empirical indicators of test-taking engagement can be developed and defined for the Programme for the International Assessment of Adult Competencies (PIAAC). The obtained results provide the basis for describing population differences in test-taking engagement among countries that participated in PIAAC (Round 1).

Since PIAAC is a computer-based large scale assessment, new methods and indicators of test-taking engagement can be derived utilising behavioral process data, in particular response time data. In the present study, disengaged responses were defined as responses with response times below a specific response time threshold. Based on this definition, responses can be evaluated as engaged or disengaged. A disengaged response means that the testee is not willing to spend the time to show what he or she knows and can do.

We used various methods for determining such thresholds, including both constant and item-specific thresholds, and validated them. For each of the methods, proportions of test-taking disengagement are reported by domain across PIAAC populations. In doing so, item positions within the PIAAC assessment design were taken into account, because testing time can be expected to be a source of test-taking disengagement. Furthermore, country differences in test-taking engagement were investigated by domain. In this analysis as well, the correlation between domain-specific test-taking engagement indicators was assessed as well as their relation to the respective proficiency measure. Finally, to shed some light on potential sources of disengagement, subgroup differences were estimated based on the background variables gender, age, educational attainment, and language.

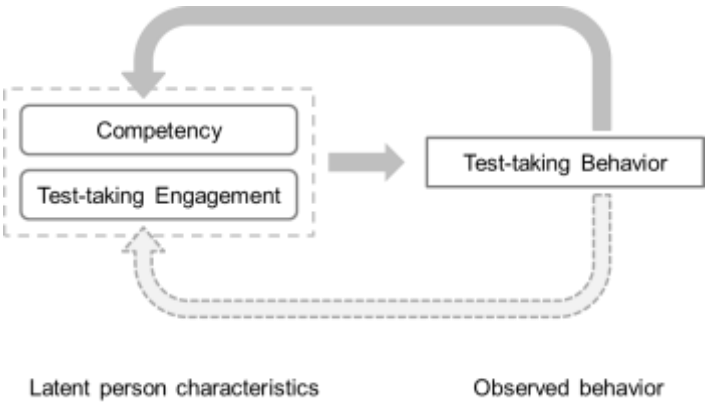
Test-taking engagement

The PIAAC study aims at measuring competencies in the domains of Numeracy, Literacy and Problem solving in Technology-Rich Environments. The measurement of such competencies needs to deal with various challenges due to the requirements of measurement theory (Brennan, 2006). The steps taken to meet these requirements are described in the technical report of PIAAC (OECD, 2013a). A fundamental threat to the validity of test score interpretation that cannot be avoided by careful test design is a lack of test-taking engagement (Wise & DeMars, 2005), especially in low-stakes assessments such as PIAAC. More specifically, it can be assumed that testing-taking behavior is not only influenced by the testee's level of competency but also by his or her actual test-taking engagement (see Figure 1).

Thus, test scores that are supposed to reflect the level of competency may be confounded with the level of test-taking engagement (Braun, Kirsch, Yamamoto, Park, & Eagan, 2011). In particular, test scores may underestimate the true proficiency level. This poses a serious challenge if test-taking engagement varies across testees as it results in a biasing effect varying across observational units (individuals, groups). Put differently, construct-irrelevant variance is introduced into the measure, which affects the validity of inferences based on test scores.

The problem of disengagement can be addressed by avoiding low test-taking engagement. This requires improving our understanding of the response process in low-stakes assessments with respect to test-taking engagement. For this, experimental (small-scale) studies could be conducted which try to influence the level of test-taking engagement, for instance, by means of incentives (cf. Braun et al., 2011). Alternatively, invalid responses could be identified afterwards, and this information could be taken into account when deriving test scores and estimating population parameters. In the present study, we considered various approaches for identifying disengaged responses in PIAAC, as described in the following section.

Figure 1. Competency and test-taking engagement



Test-taking behavior is influenced by both the to-be assessed competency and by the individual test-taking engagement. Test-taking behavior is used to draw inferences on competency (response data) and can also be used to judge test-taking engagement (response time data).

Indicators of test-taking engagement

One approach to identifying the influence of disengaged test-taking behavior is to employ self-report questionnaires assessing test-taking engagement (e.g. the effort thermometer in PISA). However, these are retrospective measures, and they might be influenced by social desirability, representing the testee's desire to fulfill the test administrator's expectations (Lee and Jia, 2014). Thus, there are doubts that testees would truthfully answer that they are not motivated. Furthermore, in PIAAC no measure of self-reported test-taking engagement was used.

Another promising and common approach for detecting disengaged test-taking behavior assumes that engaged task completion is indicated by taking at least a certain minimum amount of time required, for instance, to read and understand the test instructions, process the stimulus' content, etc., whereas disengaged test-taking behavior means taking less time or guessing rapidly. Thus, response time distribution can be used to derive a threshold that separates engaged and disengaged response behavior at the level of individual items.

Different methods for identifying response time thresholds have been proposed: The three-second rule is commonly used to define a constant threshold (cf. Lee and Jia, 2014; Kong, Wise, and Bhola, 2007). The idea of item-specific thresholds elaborates on the assumption that engaged test-taking behavior is associated with taking a minimum amount of time to be able to give a correct response. This amount of time may differ between items. For instance, testees might be able to quickly process the necessary information provided for an easy Numeracy item, whereas solving a hard Literacy item presenting a long text would require much more time. Thus, processes of test-taking disengagement in the Numeracy item would be expected below a shorter initial time period than in the Literacy item.

One approach to determine item-specific thresholds is to inspect the response time distribution visually (DeMars, 2007; Setzer, Wise, van den Heuvel, and Ling, 2013). The idea is to identify the threshold as the response time marking a distinctive gap in a bimodal response time distribution. An automated way of determining a threshold was proposed by Wise and Ma (2012). Their normative threshold method defines a certain percent (e.g. 10%) of the average item response time as the threshold and assumes a maximum threshold value of, for instance, 10 seconds. As another approach to obtain item-specific thresholds, Lee and Jia (2014) applied a method to MC items already considered by Ma, Wise, Thum, and Kingsbury (2011). For each item, the proportion correct conditional on the response time (i.e.

for every possible response time value) was computed. The threshold was defined as the response time which is clearly associated with a proportion correct greater than the chance level for success (e.g. 0.25 for a four-choice item). In their study, Lee and Jia (2014) concluded that the visual inspection of response time distributions combined with conditional proportion correct information (VITP approach) outperformed the normative threshold method, as the latter produces smaller thresholds.

To compare the performance of alternative test-taking engagement indicators, validity checks need to be conducted. An indicator of disengaged response behavior is considered valid if it identifies responses with no chance for success. Thus, the method should be able to detect timing behavior which does not enable the correct completion of the item, for instance, because of rapid guessing or due to the fact that the time on the item is below the minimum allowing for success above chance level. In the present study, two validity checks as described by Lee and Jia (2014) were performed.

Research goals

In the present study, the following three research goals were pursued:

1. Deriving indicators of test-taking engagement for PIAAC by means of response time thresholds (constant and item-specific thresholds were considered).

2. Validating the obtained indicators of test-taking engagement by inspecting the proportion correct for engaged and disengaged behavior.

3. Using these indicators to compute the proportion of disengaged responses by country and by domain as well as to investigate subgroup differences in PIAAC.

METHOD

Sample

The target population for PIAAC consisted of all non-institutionalised adults between age 16 and 65 (inclusive) who reside in the country (meaning their usual place of residency is in the country) at the time of data collection. For the following analyses, we merged the public use files from all countries that participated in the PIAAC main study (Round 1), resulting in a sample of 152 514 participants from 22 countries (data from two more countries, Australia and Cyprus[1], was not published). The public use files were downloaded from the OECD webpage on 18 July, 2014. The average age was 40.05 years (SD = 14.50), while 27.70% of all cases had missing information on age ("Not stated or inferred"). Across all participants, 47.40% were male and 52.60% female.

[1] 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".

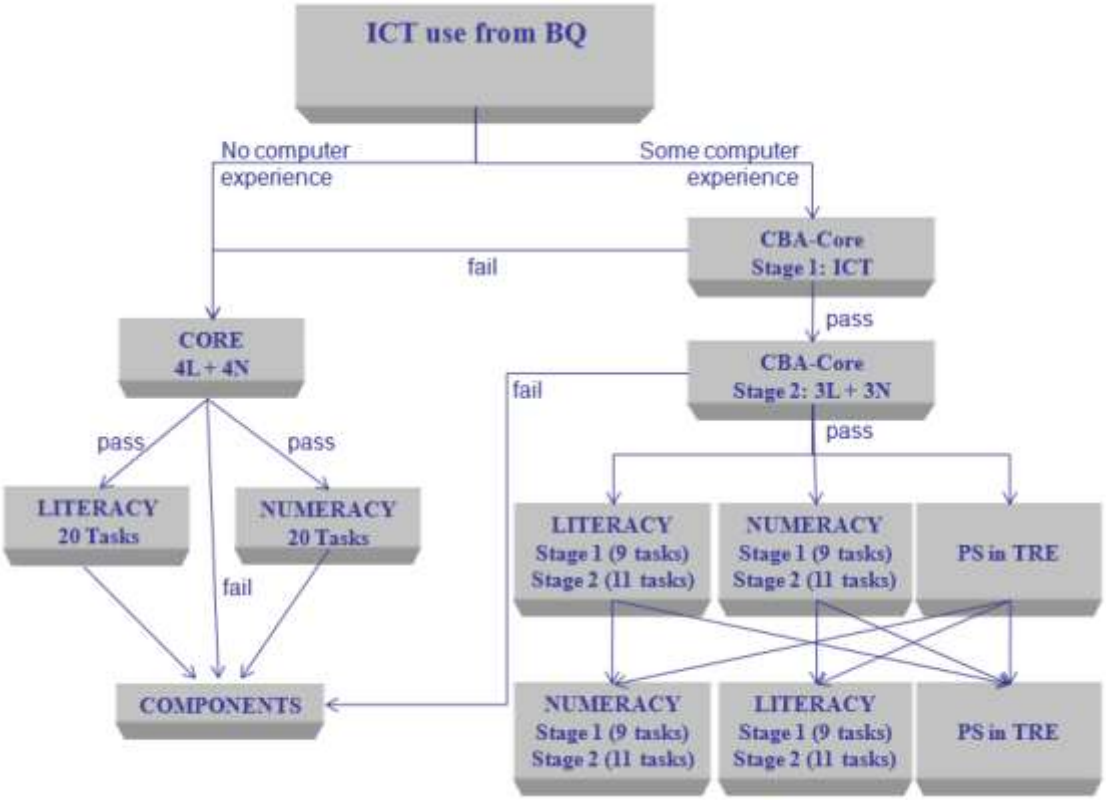
Note by all the European Union Member States of the OECD and the European Union: The Republic of Cyprus is recognised by all members of the United Nations with the exception of Turkey. The information in this document relates to the area under the effective control of the Government of the Republic of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

PIAAC assessment design

Figure 2 below illustrates the PIAAC assessment design. In the technical report, it is described as follows (OECD, 2013a, Chapter 1, pp. 8-9):

> The PIAAC assessment design for the Main Study was based on an assumption of 60 minutes of testing time, on average, for the direct assessment. As PIAAC was not a timed assessment, some respondents were expected to take longer to complete the survey. The Main Study design was implemented using the design illustrated below, where L represents literacy tasks, N represents numeracy tasks and PSTRE represents tasks involving problem solving in technology-rich environments. Among other things, the background questionnaire (BQ) asked about the respondent's computer experiences, which was essential to branch respondents to either the paper-and-pencil or CBAs at the conclusion of the BQ. Respondents with no computer experience based on BQ questions G_04 and/or the H_04a were routed to the paper branch, as were respondents refusing to take the test on the computer. The remainder of respondents were routed to the computer branch of the survey.
> […]
> The computer-delivered branch of the assessment first directed respondents to the CBA Core section, which was composed of two stages taking approximately five minutes each. Poor performance on either stage of the computer-based CBA Core section resulted in switching over to the appropriate sections of the paper-and-pencil instruments. Respondents who failed CBA Core Stage 1 (which contained Information and Communication Technology (ICT) related tasks) were directed to begin the paper-based Core section and proceed with the process outlined in the above bullet. Respondents who passed CBA Core Stage 1 but failed CBA Core Stage 2 (which contains 6 cognitive items) were then administered only the reading components tasks. Respondents who performed well on the both CBA Core sections were routed to one of three possible outcomes (each taking approximately 50 minutes): 50% of respondents received a combination of literacy and numeracy tasks, 33% received problem solving combined with either literacy or numeracy, and 17% received only problem-solving sections.

Figure 2. Integrated Main Study assessment design



The right side shows the computer-based part of the assessment design. Participants completed CBA Module 1 and Module 2 with a combination of Literacy and Numeracy, a combination of Problem solving along with Literacy or Numeracy, or only Problem solving.

Source: Technical Report of the Survey of Adult Skills (PIAAC), OECD, 2013a Chapter 1, p. 10).

For the present analyses, it is important to note that item content for Literacy and Numeracy was administered in both Modules 1 and 2; for Problem solving, however, the order of item content was fixed. This means that random equivalent groups completed the Literacy items followed by Numeracy items and vice versa. This design feature allows for an investigation of whether the rate of disengaged response behavior depends on items' position in the first or the second half of the cognitive assessment.

PIAAC data at the item level

The PIAAC public use file provides four variables for each item in the computer-based cognitive assessment, that is, the item score, the total time (response time), the time to the first action, and the number of actions. Since there was no global time limit, each testee should have visited each item. There was no way to skip items since testees could navigate only successively from one item to the next item (backward navigation was limited to Literacy and Numeracy units).

The item score includes the categories correct, incorrect (for Problem solving also partial credit), no response, and not reached/not attempted. Regarding missing responses, "No response" means that at least five seconds were spent on the item without any interaction, whereas "Not reached/not attempted" means that less than five seconds were spent on the item by the testee without any interaction. The latter coding is

in line with the idea that the cofounding of the competency assessment with test-taking engagement can be avoided by not coding the missing response as an incorrect response (cf. OECD, 2013a, Chapter 17).

For the present analyses, Problem solving scores were dichotomized by turning partial credit responses into incorrect responses. Moreover, only response times and item scores were considered. For validating the indicators of test-taking engagement, the proportion correct was computed as the ratio of the number of correct responses and the total number of observations (including correct, incorrect, no response, not attempted/not reached).

Deriving indicators of test-taking engagement

Response time thresholds

Items were judged to be completed without sufficient test-taking engagement (non-solution behavior) if the response time fell below a certain threshold. Other aspects of performance such as number of interactions and, if there was an interaction, the correctness of the response, were not taken into account for this assessment.

Two response-time thresholds that were identical across items were used, 3000 ms and 5000 ms. The three-second rule is a commonly used approach (cf. Lee and Jia, 2014; Kong, Wise, and Bhola, 2007). The second threshold of 5000 ms (e.g. Wise and Kong, 2005) was included as well since in PIAAC missing responses were coded depending on whether the time taken was less than five seconds ("Not reached/Not attempted") or more than five seconds ("No response") (see public use file and OECD, 2013a, Chapter 17).

As an approach to obtain item-specific thresholds, the visual inspection (VI) method was applied (cf. Kong, Wise, and Bhola, 2007; Lee and Jia, 2014). This method requires disengaged test-taking behavior to give rise to a bimodal item response time distribution (cf. Figure 4). Typically, there is a high-frequency spike during the initial seconds after the item is administered, followed by a region of low frequency, followed by another strong increase in frequency that finally decreases (cf. Kong et al., 2007). This bimodal distribution is assumed to be a mixture of two time distributions: one distribution for participants responding quickly and another time distribution for those responding slowly. The threshold separating disengaged behavior and engaged behavior is obtained as the judged end point of the short time spike (Wise, 2006).

As another approach for generating item-specific thresholds, the proportion correct greater than zero (P+>0%) method was applied. It is an adapted version of Lee and Jia's (2014) visual inspection of response time distributions with conditional P+ information (VITP). The assumption is that for PIAAC items the chance level for obtaining a correct response can be assumed to be zero because almost all response formats do not allow for rapid correct guesses. There are only five MC-like Numeracy items; all other items require entering numbers or interacting with the stimulus, for instance, by highlighting text or clicking a graphical element. For determining the P+>0% threshold, as a preparatory step, the proportion correct conditional on the response time was computed at one second intervals. The threshold was defined as the shortest response time associated with proportion correct greater than zero.

Level of aggregation

Since PIAAC items show great variability in complexity and difficulty and consequentially in time intensity, we favored item-specific response time thresholds. This is also in line with previous research (e.g. Hauser and Kingsbury, 2009; Lee and Jia, 2014; Wise, 2006). Furthermore, a particular item can be completed in various contexts. Accordingly, even for a single item response time thresholds can be determined in multiple ways. In particular, it could be assumed that the position of an item (Module 1 vs. Module 2) within the assessment design has an impact on the location of the threshold separating engaged

and disengaged response behavior (e.g. due to fatigue effects). Therefore, for Literacy and Numeracy, we determined response time thresholds for Module 1 and Module 2 separately.

Furthermore, some (Literacy and Numeracy) items can occur in different testlets at a particular stage of a module. However, we did not have any hypothesis on the effect of the testlet on the response time threshold. Neither did we expect the item response time threshold to vary across countries. From an information processing perspective, there seems to be no reason to assume that the minimum amount of time needed to be able to complete a particular item above chance level varies across countries.

Validating indicators of disengaged response behavior

Proportion correct

The first validity check compares the proportion correct of engaged response behavior with the proportion correct of disengaged response behavior (cf. Wise and Kong, 2005; Wise and Ma, 2012). For a valid indicator, it is expected that in the case of engaged response behavior, the probability of obtaining a correct response is much higher than chance level, whereas for disengaged response behavior the probability of obtaining a correct response should be at chance level. Given the response formats used in the PIAAC for Literacy, Numeracy, and Problem solving, which usually include many response options, chance level is assumed to be zero or close to zero for almost all items. In the present study, we compared the proportions correct at both the domain and item levels.

Correlation of score group and proportion correct

As another validity check, we investigated for each item the relationship between the participants' overall performance (score group) and the probability of success on the item conditional on the response behavior, that is, engaged and disengaged response behavior (cf. Lee and Jia, 2014; Wise and DeMars, 2006). For each item, persons completing the item were divided into score groups, and the proportion correct was computed for each score group. For persons with engaged response behavior for a specific item, their overall scores should be positively associated with the proportion correct in the item; however, no such relationship was expected for the group of persons evaluated as disengaged on that item.

In the present study, participants were divided into score groups as defined by PIAAC competence levels (see OECD, 2013a, Chapter 18). Thus, there were six score groups for Literacy and Numeracy and four score groups for Problem solving. In assigning persons to score groups, the first plausible value of the respective domain was used.

Data analysis

Computations and plots for descriptive statistics were done using SPSS Version 21 and R (R Core Team, 2014). The procedures for generating tables and graphics were documented as syntax files and can be obtained from the authors upon request.

Thresholds based on the VI method were determined independently by two raters. Inter-rater agreement was assessed by means of the ICC2 statistic (Shrout and Fleiss, 1979), which assumes a random sample of raters rating each object and represents a measure of absolute agreement in the ratings. ICC analyses were conducted using the R package psych (Revelle, 2015).

The average proportion of disengaged responses at the population or subgroup level and related standard error were computed by means of the R package BIFIEsurvey (Robitzsch and Oberwimmer, 2015), taking weights and replicate weights into account as provided in the public use file. Using the BIFIEsurvey package, the $D1$ test statistic (cf. Enders, 2010, Chapter 8) was computed to test the equality

of means between countries. Furthermore, given the big sample sizes, the effect size, η^2, was also considered to judge the relevance of significant country differences (small: 0.01, medium: 0.06, large: 0.14, cf. Cohen, 1988).

RESULTS

Response time thresholds as indicators of test-taking engagement

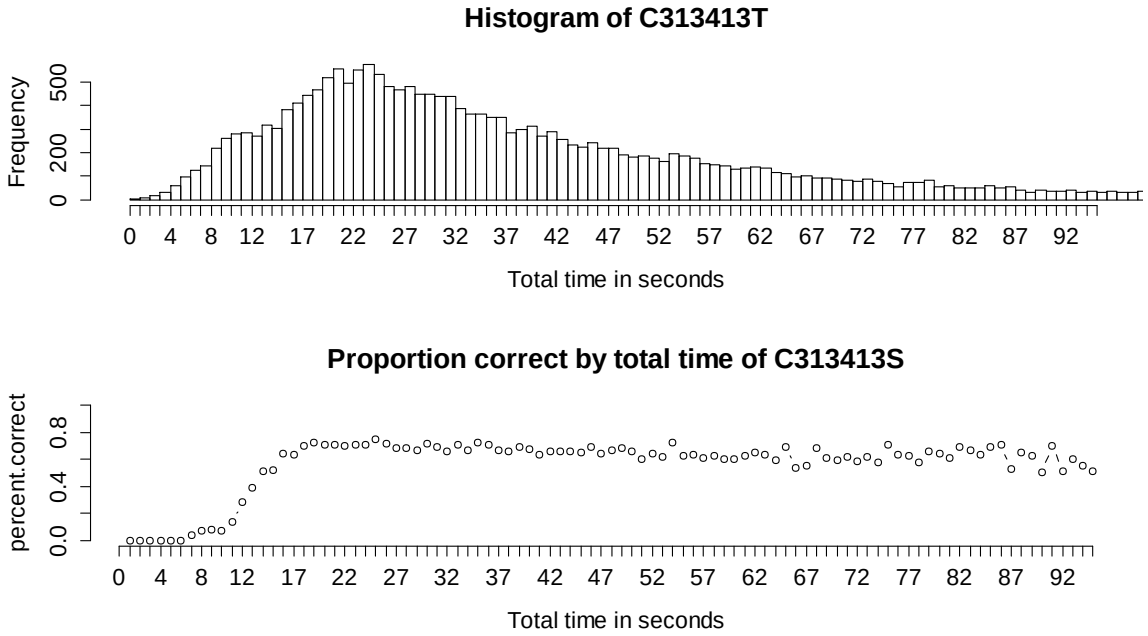
Item-specific RT thresholds were defined using the visual inspection (VI) method and the proportion of correct responses greater than zero (P+>0%) method.

Visual inspection (VI) threshold

The upper part of Figure 3 shows an example of a response time distribution which did not exhibit bimodality. In such cases, no threshold could be obtained using the VI method.

The upper part of Figure 4 shows a response time distribution which clearly has two modes and a clear-cut gap in between, which can be interpreted as a threshold separating disengaged and engaged response behavior. For the sample item Literacy C313410, the threshold was rated to be 12 s.

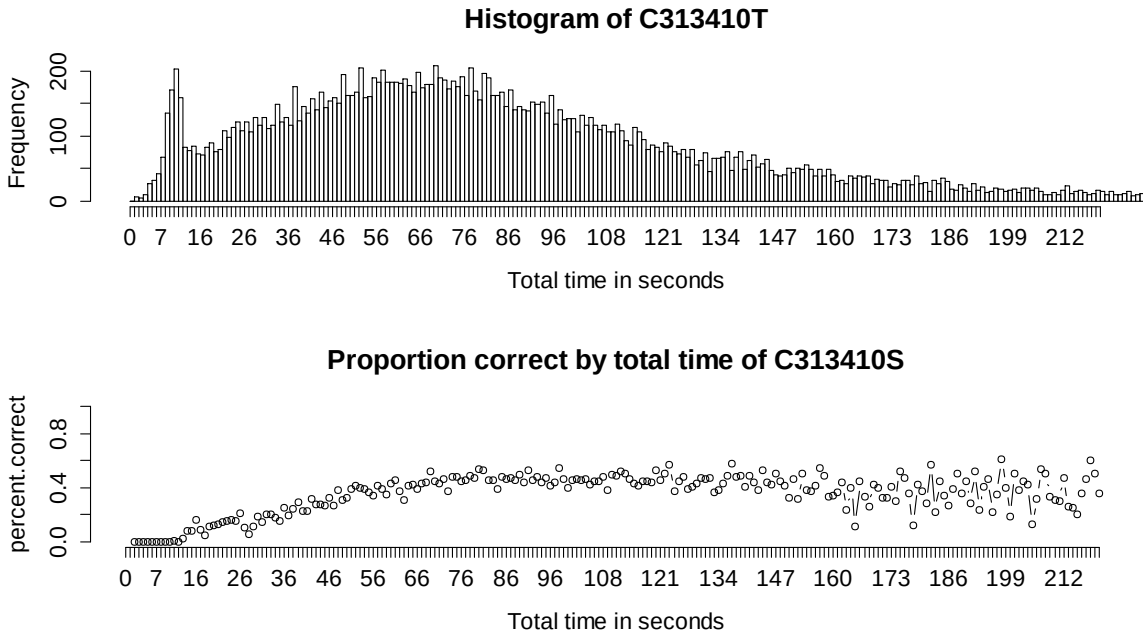
Figure 3. Response time distribution and proportion correct by response time.



Upper part: The histogram shows the response time distribution (truncated at the 95th percentile) for a selected Literacy item completed in Module 2. Lower part: The plot shows the proportion correct conditional on the response time (total time) for the item.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Figure 4. Response time distribution and proportion correct by response time



Upper part: The histogram shows the response time distribution (truncated at the 95th percentile) for a selected Literacy item completed in Module 2. Lower part: The plot shows the proportion correct conditional on the response time (total time) for the item.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

The raters (Rater 1 and Rater 2) of the distributions were instructed to first judge whether the item response time distribution showed a bimodal shape, that is, whether a spike for short response times was visible. If there was a spike, the threshold was determined as the response time where the spike ends. There were many bimodal distributions in which there was no clear-cut local minimum between the two modes (i.e. deep and narrow "valley", cf. Figure 3, upper part) but a wide range of about equally low values between the two modes (i.e. broad "valley"). In these cases the threshold was determined as the response time where the plateau of equally low values began.

The analysis of inter-rater agreement revealed that the two raters showed very high agreement in setting response time thresholds for Literacy (ICC2=0.95) and Numeracy (ICC2=0.93), while agreement for Problem solving was much lower (ICC2=0.40). For the following analyses, the thresholds from Rater 1 are used.

Table 1 presents the proportion of items that exhibited a bimodal response time distribution for each of the three domains, and for Literacy and Numeracy separately for Module 1 and Module 2. The results consistently indicate that about two third of the items in Literacy and Numeracy (Module 1) showed a bimodal response time distribution. Interestingly, in Module 2, this ratio was substantially increased to almost 100% for both Literacy and Numeracy. For Problem solving, almost 80% of items showed a bimodal shape (since the order of Problem solving items included in Module 1 and 2 was not balanced, results by module are not available). The increase in items for which bimodality could be more clearly defined was due to a more distinguished, that is, higher, peak of short response times. This suggests that the proportion of disengaged response behavior increased in Module 2.

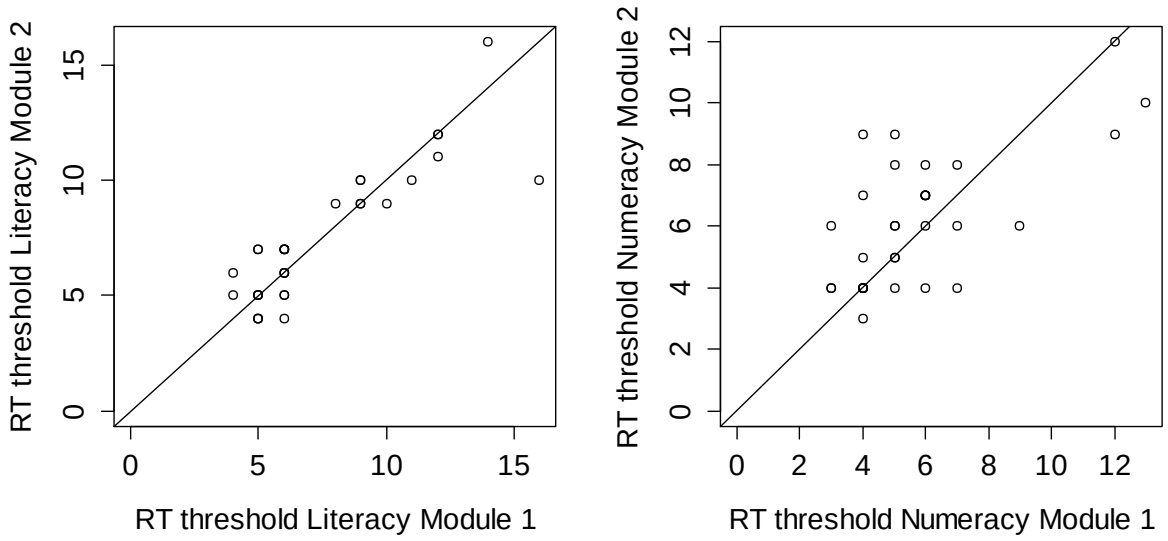
Table 1. Items exhibiting a bimodal response time distribution

	Total number of items	Module	Proportion of items with bimodal response time distribution
Literacy	49	1	0.65
		2	0.98
Numeracy	49	1	0.67
		2	0.98
Problem solving	14	n/a	0.79

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

For a substantial proportion of Literacy and Numeracy items presented in Module 1, a response time threshold could not be obtained using the VI method. This problem could be remedied by using thresholds that were obtained by collapsing Module 1 and Module 2 data. This increased the number of response time observations for each item, and the bimodal distribution usually found in Module 2 items assured that a bimodal distribution could be found across Module 1 and Module 2 as well. In such cases, item thresholds can definitely not be understood as module-specific thresholds anymore. In the end, only for one Literacy item (Item 43: C313413), two Numeracy items (Item 1: C615602, Item 5: C604505), and three Problem solving items (Item 1: U01a000, Item 3: U03a000, Item 7: U04a000) could no threshold be determined.

Figure 5. Associations of response time (RT) thresholds (VI method) between Module 1 and Module 2.



Left part: Literacy items. Right part: Numeracy items. The RT thresholds are in seconds.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

For Literacy and Numeracy, the response time thresholds were determined separately for Module 1 and Module 2 under the assumption that the location of the response time threshold might change due to a position effect. Figure 5 shows how the thresholds of Module 1 relate to those of Module 2 (note that the thresholds were determined at one second intervals, i.e. in Figure 5 one dot may represent more than one item). The variation around the diagonal line means that there were some differences between the two modules, but overall there was still high consistency. This was also indicated by the correlation between the modules' thresholds, which was r=0.87 (p<0.001) for Literacy and r=0.66 (p<0.001) for Numeracy.

Also, differences in the average response time thresholds were only small (Module 1 vs. Module 2: 7.469 s vs. 7.042 s for Literacy, for Numeracy 5.727 s vs. 5.458 s). These average response time thresholds indicate that the response time associated with the VI method was longer for Literacy than for Numeracy (see also Figure 5).

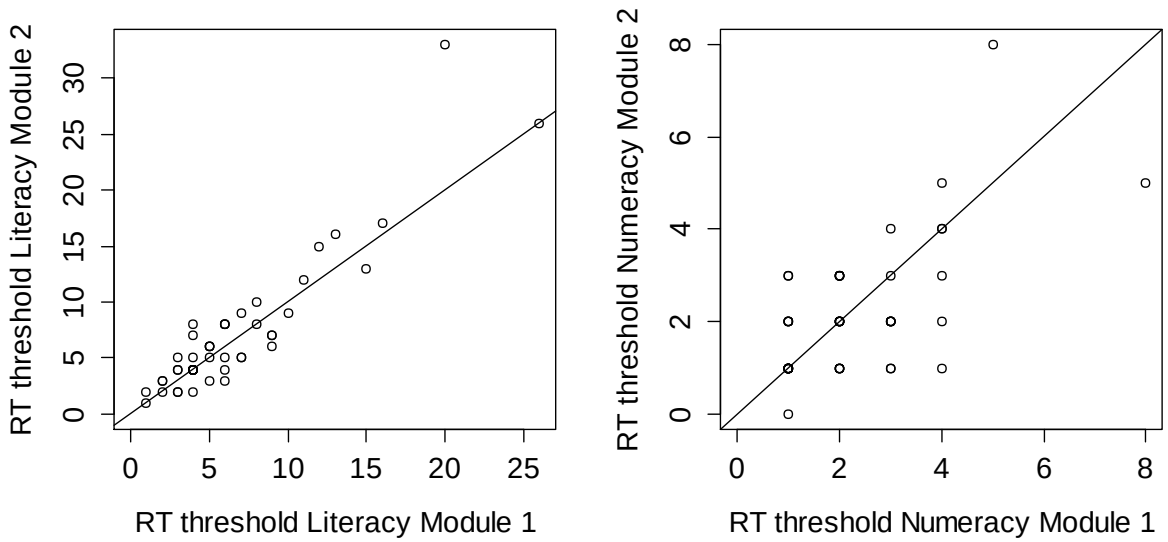
The variability in the response time thresholds defined by means of the VI method confirms the need for item-specific thresholds. For Literacy, the thresholds varied between 4 s and 16 s for Module 1 and between 3 s and 16 s for Module 2; for Numeracy, between 3 s and 13 s for Module 1 and 3 s and 12 s for Module 2; and for Problem solving between 8 s and 25 s.

Proportion correct greater than zero (P+>0%) threshold

The thresholds as defined by the P+>0% method could be determined automatically as the shortest response time associated with a proportion correct than was greater than zero. There were items for which the proportion correct was greater than 0% for all empirical response time intervals. This means that all responses were considered to be engaged responses. This was the case for several Numeracy items (Module 1: 28.57%, Module 2: 24.49%), a few Literacy items (Module 1: 4.17%, Module 2: 2.04%) and no Problems solving items.

As for the VI method, response time thresholds for Literacy and Numeracy were determined separately for Module 1 and Module 2. Figure 6 shows how the thresholds of Module 1 relate to those of Module 2. The variation around the diagonal line suggests that there were some differences between the two modules, but overall there was still high consistency. The correlation of the modules' thresholds was r=0.92 (p<0.001) for Literacy and r=0.63 (p<0.001) for Numeracy. As before, the difference in the average response time threshold was only small (Module 1 vs. Module 2: 6.510 s vs. 6.980 s for Literacy, 2.265 s vs. 2.163 s for Numeracy). These average response time thresholds indicate that the response time associated with the P+>0% method was much longer for Literacy than for Numeracy (see also Figure 6).

Figure 6. Associations of response time (RT) thresholds (P+>0% method) between Module 1 and Module 2.



Left part: Literacy items. Right part: Numeracy items. The RT thresholds are in seconds.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

For Literacy, the thresholds varied between 1 s and 26 s for Module 1 and between 1 s and 33 s for Module 2; for Numeracy, between 1 s and 8 s for Module 1 and 0 ms and 8 s for Module 2; and for Problem solving between 3 s and 76 s.

Validation of indicators of test-taking engagement

Proportion correct

A valid indicator should separate disengaged responses which are associated with a low probability of success from engaged responses which are associated with a higher probability of success. Table 2 shows the observed proportion correct for responses classified as engaged and those classified as disengaged in each of the four methods (5000, 3000, VI, P+>0%). For each domain and method, the proportion correct values were obtained by averaging across items and countries.

The proportion correct for disengaged response behavior was close to zero or zero, whereas the proportion correct for engaged response behavior was much higher (depending on the overall difficulty of items). Most importantly, for Literacy and Problem solving the proportion correct for engaged response behavior was highest using the VI method and the P+>0% method (0.56 for Literacy and 0.43 for Problem solving), and the proportion correct for disengaged response behavior was lowest using the P+>0% method (0.00 for both Literacy and Problem solving). The latter finding was expected a priori, since the P+>0% method by definition is associated with a proportion correct of zero for disengaged responses. Regarding Numeracy, the proportion correct for engaged response behavior was highest for the 5000 method (0.64); however, this method also exhibited the highest proportion correct for disengaged response behavior. Thus, the difference between proportion correct for engaged and disengaged response behavior (see last column of Table 2) favored the P+>0% method for Numeracy as well.

Table 2. Average proportion correct for engaged and disengaged response behavior.

	Method	Proportion correct - Engaged	Proportion correct - Disengaged	Difference
Literacy	5000	0.55	0.02	0.53
	3000	0.55	0.01	0.54
	VI	0.56	0.02	0.54
	P+>0%	0.56	0.00	0.56
Numeracy	5000	0.64	0.09	0.55
	3000	0.63	0.04	0.59
	VI	0.63	0.07	0.56
	P+>0%	0.63	0.00	0.63
Problem solving	5000	0.40	0.00	0.40
	3000	0.40	0.00	0.40
	VI	0.43	0.01	0.42
	P+>0%	0.43	0.00	0.43

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Figure 7 (Literacy), Figure 8 (Numeracy) and Figure 9 (Problem solving) show the difference between the proportion correct for responses classified as engaged and disengaged response behavior at the item level. As before, the validity of the indicator is reflected in the magnitude of the difference between the two measures.

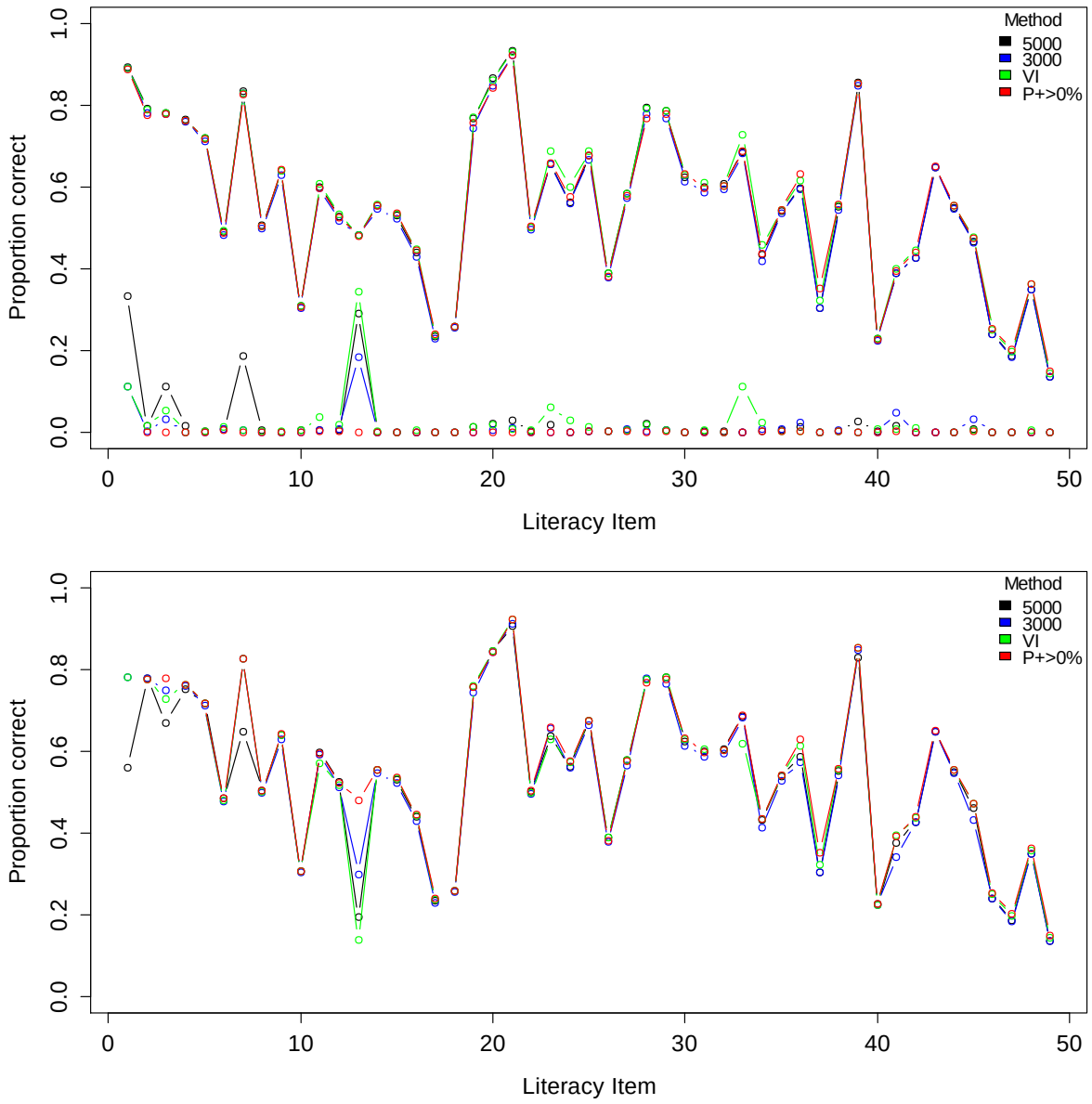
In the upper part of Figure 7 (Literacy), the upper four lines show the proportion correct for engaged response behavior using the methods 5000 (black line), 3000 (blue line), VI (green line) and P+>0% (red line). There were no large or consistent differences among the four methods. For some items, the VI method showed the highest proportion correct (e.g. items 24-26), whereas for other items this was the

P+>0% method (e.g. items 46-50). The lower four lines (in the upper part of Figure 7) represent the corresponding proportion correct for responses classified as disengaged response behavior. There are some peaks for the 5000, 3000 and VI methods (cf. items 1, 7, 13), suggesting that several correct responses were considered as disengaged response behavior. The line for the P+>0% method was zero. This was a consequence of how thresholds were defined, that is, responses classified as disengaged by definition show a proportion correct of zero. Note that for the P+>0% method some data points were missing (e.g. item 1) which was due to the fact that for some items the proportion correct was greater than zero even for the shortest response time. This means that there were no responses classified as disengaged. In the lower part of Figure 7, the four lines show the difference between engaged and disengaged response behavior using the methods 5000 (black line), 3000 (blue line), VI (green line) or P+>0% (red line). For the majority of items, the difference was greatest for the P+>0% method, followed by the VI method.

Figure 8 shows the findings for Numeracy. The proportion correct for engaged behavior was somewhat lower for the P+>0% method compared to the 5000, 3000, and VI methods. However, this was compensated for by the fact that the proportion correct for disengaged behavior was well above zero for the 5000, 3000, and VI methods across most of the items (see upper part of Figure 8). Accordingly, the lower part of Figure 8 demonstrates that the difference between the proportion correct for engaged and disengaged response behavior was greatest for the P+>0% method for almost all Numeracy items.

Finally, Figure 9 shows the findings for Problem solving. The proportion correct for engaged response behavior was mostly the highest for the P+>0% method, followed closely by the VI method, whereas the proportion correct for disengaged response behavior was lowest for the P+>0% method (see upper part of Figure 8). Accordingly, the lower part of Figure 8 shows that the difference between proportion correct for engaged and disengaged response behavior was greatest for the P+>0% method for almost all items (followed by the VI method).

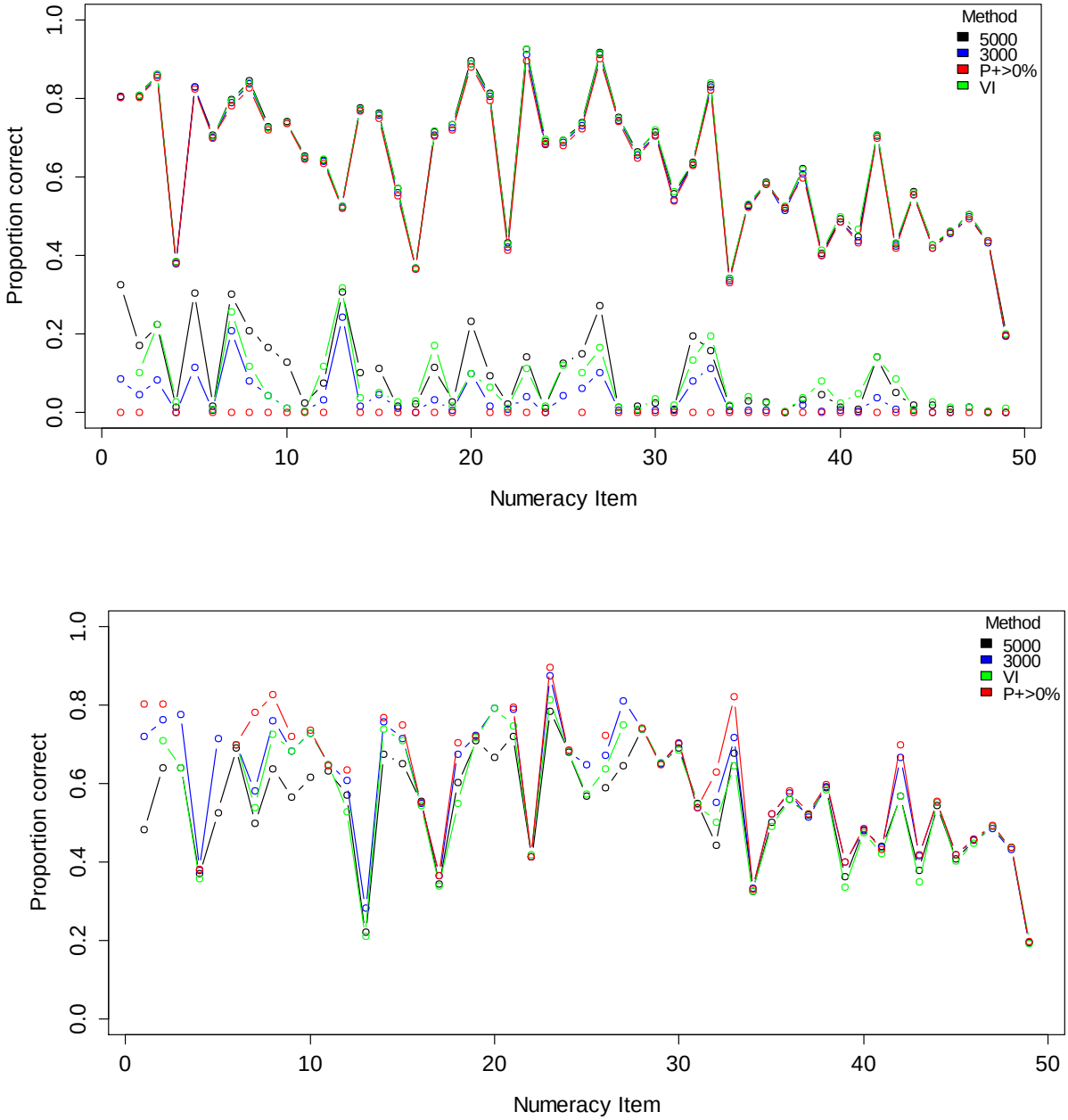
Figure 7. Proportion correct for engaged and disengaged response behavior in Literacy.



Upper part: The upper four lines show the proportion correct for engaged response behavior using the methods 5000 (threshold of 5000 ms, black line), 3000 (threshold of 3000 ms, blue line), VI (item-specific threshold, green line) and P+>0% (item-specific threshold, red line). The lower four lines represent the corresponding proportion correct for disengaged response behavior. Lower part: The four lines show the difference of the proportion correct for engaged and disengaged responses.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

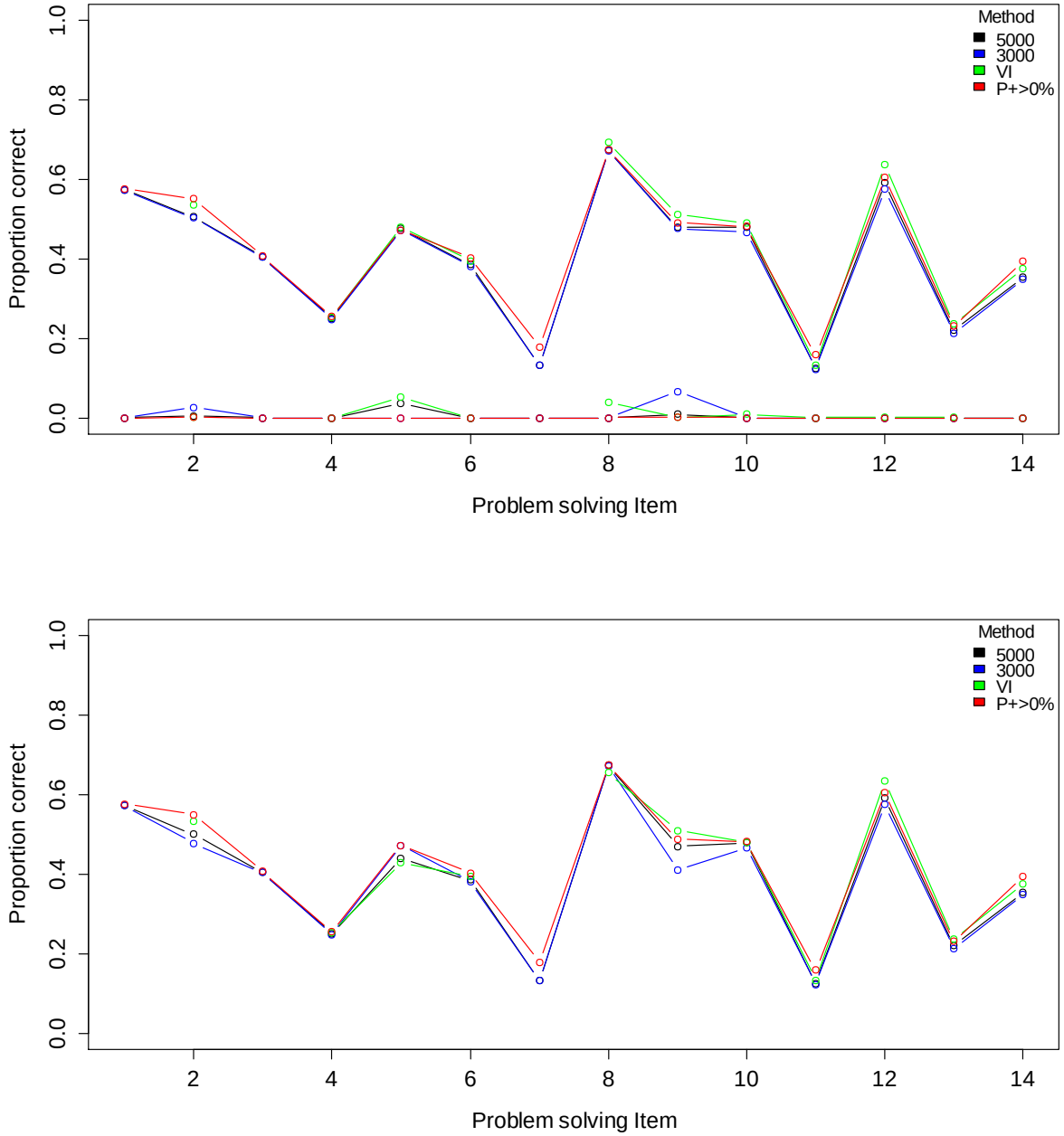
Figure 8. Proportion correct for engaged and disengaged response behavior in Numeracy.



Upper part: The upper four lines show the proportion correct for engaged response behavior using the methods 5000 (threshold of 5000 ms, black line), 3000 (threshold of 3000 ms, blue line), VI (item-specific threshold, green line) and P+>0% (item-specific threshold, red line). The lower four lines represent the corresponding proportion correct for disengaged response behavior. Lower part: The four lines show the difference of the proportion correct for engaged and disengaged responses.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Figure 9. Proportion correct for engaged and disengaged response behavior in Problem solving.



Upper part: The upper four lines show the proportion correct for engaged response behavior using the methods 5000 (threshold of 5000 ms, black line), 3000 (threshold of 3000 ms, blue line), VI (item-specific threshold, green line) and P+>0% (item-specific threshold, red line). The lower four lines represent the corresponding proportion correct for disengaged response behavior. Lower part: The four lines show the difference of the proportion correct for engaged and disengaged responses.

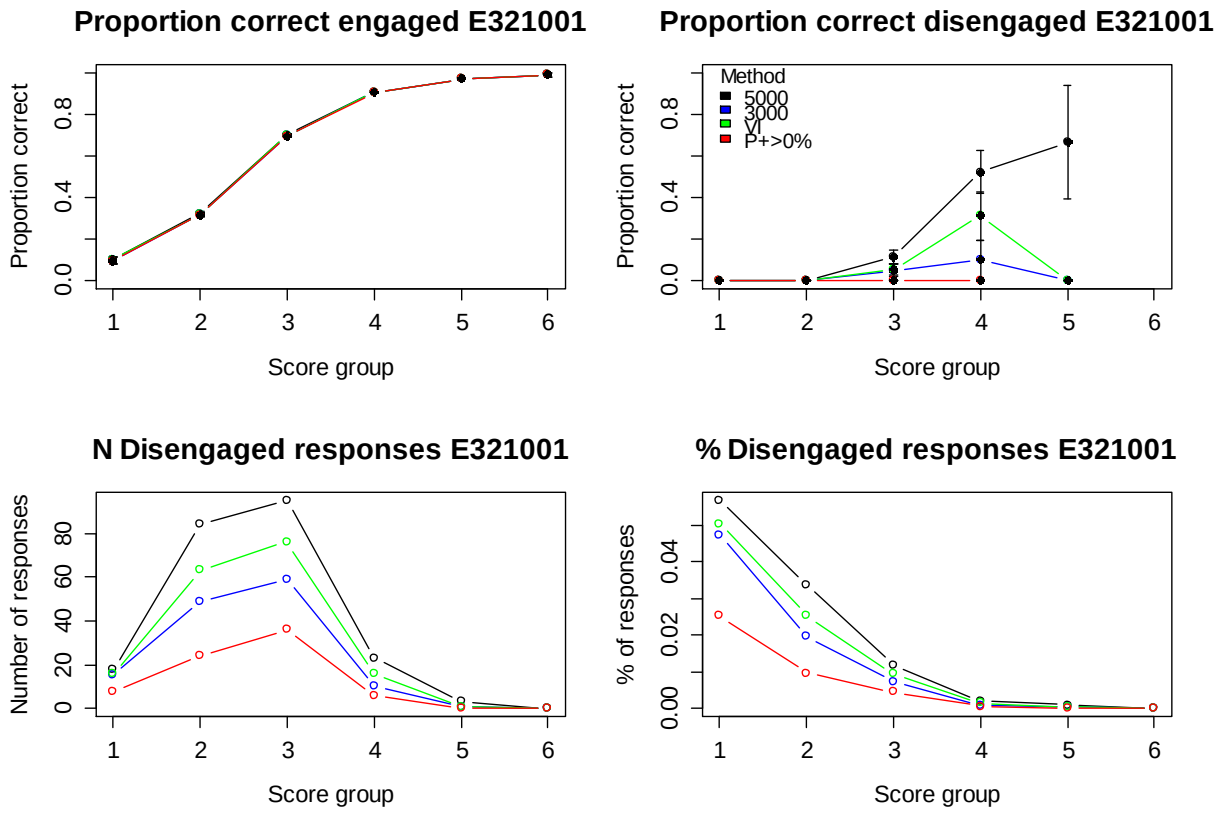
Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Correlation of score group and proportion correct

As another validity check, we investigated for each item the association between participants' overall scores (i.e. a selected plausible value) and the proportion correct for responses classified as engaged response behavior and those classified as disengaged response behavior.

Figures 10, 11 and 12 show the findings for selected Literacy, Numeracy and Problem solving items. The upper left part shows the association between score group (plausible value) and proportion correct for engaged response behavior, and the upper right part the association between score group and proportion correct for disengaged response behavior (error bars indicate +/- 1 SE). The lower left part provides information on the absolute number of disengaged responses by score group, and the lower right part the corresponding proportion of disengaged responses by score group.

Figure 10. Relationship between score group and proportion correct in Literacy (selected item).



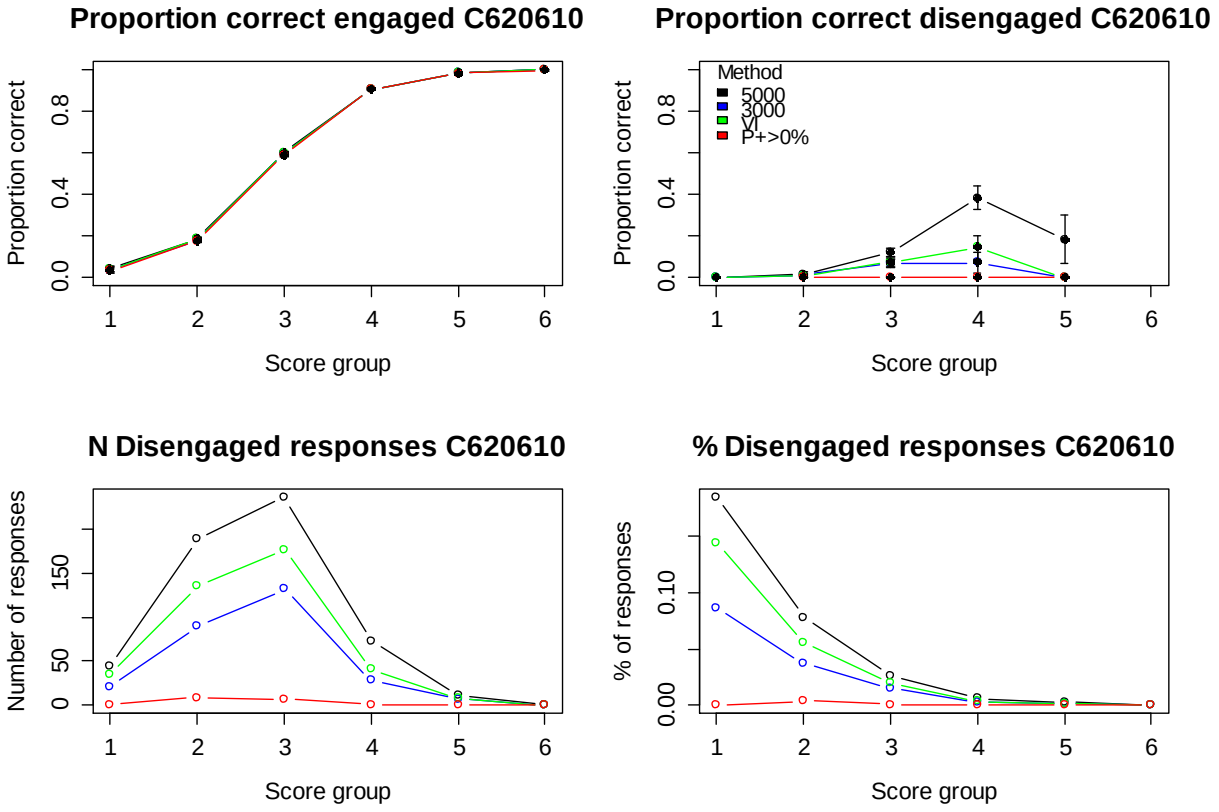
Upper left part: Association between score group (plausible value) and proportion correct for engaged response behavior (error bars indicate +/- 1 SE; given the sample size they are very small). Upper right part: Association between score group and proportion correct for disengaged response behavior (error bars indicate +/- 1 SE). Lower left part: Number of disengaged responses by score group. Lower right part: Proportion of disengaged responses by score group.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

As expected, the association between score group (plausible values) and proportion correct for engaged response behavior was positive for all items (see upper left part in Figures 9, 10, and 11), and, given the large proportion of engaged responses, few differences were revealed among the four methods

for deriving thresholds. However, the upper right part reveals differences in the association between score group and proportion correct. The P+>0% method exhibited a proportion correct of zero across all score groups by definition, whereas the 5000, 3000, and VI methods sometimes showed a positive association, suggesting that correct responses had been falsely classified as disengaged response behavior. This association was expected to be weaker on average for the 3000 method compared to the 5000 method since shifting the threshold to shorter response times should decrease the proportion correct. In principle, this should have also attenuated the association between score group and proportion correct for engaged response behavior by including incorrect responses based on disengaged response behavior. Note that for the Problem solving item (see Figure 12), the proportion correct for disengaged response behavior was zero across score groups for the 5000, 3000 and VI methods as well, proving that Problem solving items typically required more than 5000 ms to obtain a correct response.

Figure 11. Relationship between score group and proportion correct in Numeracy (selected item).



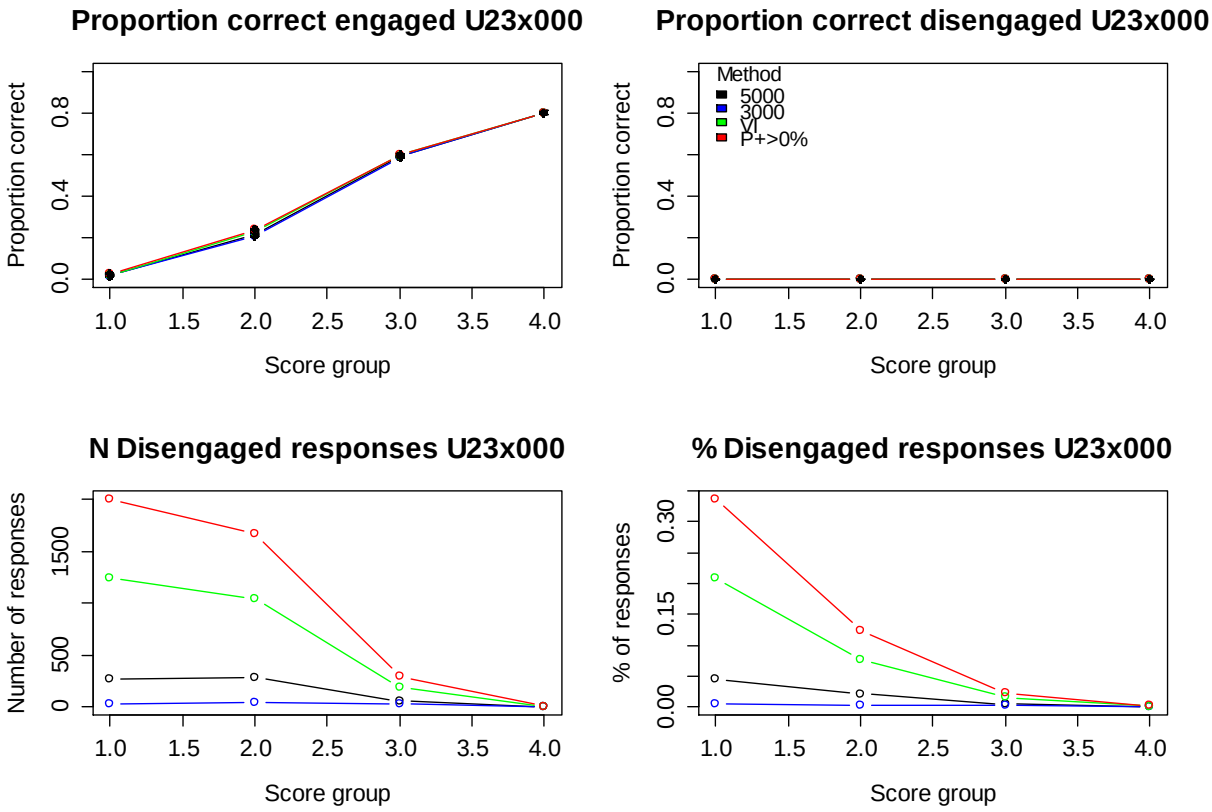
Upper left part: Association between score group (plausible value) and proportion correct for engaged response behavior (error bars indicate +/- 1 SE; given the sample size they are very small). Upper right part: Association between score group and proportion correct for disengaged response behavior (error bars indicate +/- 1 SE). Lower left part: Number of disengaged responses by score group. Lower right part: Proportion of disengaged responses by score group.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

The distribution of the number of disengaged responses (lower left part) was skewed to the right for all four methods and across Literacy, Numeracy and Problem solving. This becomes even more obvious in the diagram showing the proportion of disengaged responses by score group (lower right part). In a very consistent finding, the proportion of disengaged response behavior was highest for the lowest score group

and gradually decreased to about 0% for the most competent testees. Since the P+>0% method considered all responses with response times below the shortest response time required for a correct response as disengaged response behavior, the proportion of disengaged responses can be much higher for the P+>0% method than for the constant thresholds of 3000 ms or 5000 ms. For instance, in the sample Problem solving item the proportion for the lowest score group was more than 30%. The VI method also showed a proportion of more than 20%, which was much higher than the proportions obtained for the constant threshold methods.

Figure 12. Relationship between score group and proportion correct in Problem solving (selected item).



Upper left part: Association between score group (plausible value) and proportion correct for engaged response behavior (error bars indicate +/- 1 SE; given the sample size they are very small). Upper right part: Association between score group and proportion correct for disengaged response behavior (error bars indicate +/- 1 SE). Lower left part: Number of disengaged responses by score group. Lower right part: Proportion of disengaged responses by score group.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Taken together, the validity checks comparing the proportion correct for engaged and disengaged response behavior showed that the P+>0% method defining item-specific response time thresholds performs better than item-specific thresholds based on the VI method, followed by the methods assuming constant thresholds of 3000 ms and 5000 ms, respectively.

Proportion of disengaged response behavior by module

Table 3 shows the proportion of disengaged response behavior by domain obtained using the two constant threshold methods, 5000 ms and 3000 ms, as well as the two item-specific methods VI and P+>0%. The columns indicate whether the respective domain was assessed in Module 1 or Module 2 (e.g. column LIT-NUM presents the results for Literacy administered in Module 1 and Numeracy administered in Module 2). Note that the proportions of disengaged response behavior presented in Table 1 cannot be understood as characteristic across the entire PIAAC population as no sample weights were taken into account. The numbers are simply based on the observed cases as included in the public use file.

Overall, the amount of disengaged test-taking behavior varied between 0.02% for the 3000 method applied to Problem solving and 8.2% for Problem solving using the P+>0% method.

Table 3. Average percentage (%) of disengaged response behavior by domain, method and module

Domain	Method	LIT-NUM	LIT-PS2	NUM-LIT	NUM-PS2	PS1-LIT	PS1-NUM	PS1-PS2
Literacy	5000	1.10	1.02	3.31	n/a	2.94	n/a	n/a
	3000	0.41	0.43	1.38	n/a	1.30	n/a	n/a
	VI	1.95	1.73	5.05	n/a	4.47	n/a	n/a
	P+>0%	1.41	1.32	3.89	n/a	3.58	n/a	n/a
Numeracy	5000	3.46	n/a	1.29	1.15	n/a	2.61	n/a
	3000	1.75	n/a	0.53	0.51	n/a	1.32	n/a
	VI	3.77	n/a	1.47	1.34	n/a	2.90	n/a
	P+>0%	0.67	n/a	0.30	0.28	n/a	0.53	n/a
Problem solving	5000	n/a	1.79	n/a	1.67	0.67	0.57	1.03
	3000	n/a	0.05	n/a	0.05	0.02	0.02	0.02
	VI	n/a	7.48	n/a	6.93	3.66	3.16	5.59
	P+>0%	n/a	8.20	n/a	7.70	6.57	6.16	6.76

n/a: items of the domain were not administered in the respective module.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

As expected, the 5000 method was consistently associated with higher proportions of disengaged response behavior than the 3000 method. The P+>0% method revealed slightly higher proportions than the 5000 method for Literacy, somewhat lower proportions for Numeracy than the 3000 method and clearly higher values for Problem solving. This pattern of differences basically reflects the idea that the time required to solve items correctly differs across domains. The P+>0% method directly represents these differences, while the constant threshold methods ignore them. The VI method produced higher proportions than the constant threshold methods across all domains. Results were quite similar to those obtained by the P+>0% method, except in Numeracy, where the proportion based on the VI method was clearly higher.

Most interesting, Table 3 reveals that the items' position had a considerable effect on the proportion of disengaged responses: In all domains, the percentage of disengaged responses was greater at the later position. For Literacy and Numeracy, the proportion for Module 2 was twice or three times as much as for Module 1. For the domain of Problem solving, this effect was confounded by the fact that the respective items were included in both Modules 1 and 2. However, the pattern of results clearly suggests that there was also a position effect in Problem solving.

As stated above, the P+>0% method provided the most valid estimate of disengaged test-taking behavior based on the applied validity checks. The high proportion of disengagement in the domain of Problem solving using the method P+>0% points to a more differentiated interpretation of disengagement which goes beyond completing or dropping items rapidly. Disengaged behavior in this sense could also include some kind of "informed" disengagement. The testee might not be able to solve the problem at hand correctly, but might be able to determine that he or she has only a very small chance of solving the item correctly. Since the testee does not know for sure, this kind of informed disengagement can still be labeled as disengaged response behavior.

Country differences in test-taking engagement

To determine country differences in test-taking engagement, the dichotomous disengagement indicators for each combination of item and person were aggregated in two steps. First, a person-level variable was computed by averaging the number of disengaged responses in the respective domain. This variable reflects domain-specific individual differences in the proportion of disengaged responses. Second, this variable was averaged to obtain the population mean for each domain. For this step, sampling weights were used to take characteristics of the sample and the selection procedure into account. The uncertainty of population estimates was determined by using replicate weights.

Note that in the first step, the set of items for which the rate of disengaged responses was computed may differ between testees for two reasons. In Literacy and Numeracy, items were completed in either Module 1 or Module 2, and within each module, items differed due to the adaptive test design (i.e. the testlet a testee received depended on background variables and proficiency). As testees were randomly assigned to modules, differences between Module 1 and Module 2 were not expected to affect the estimation of means at the population level. However, given the adaptive test design, there may be differences between countries in the assembly of testlets within a module depending on country differences in proficiency. Following Lee and Jia (2014), the discrepancy between a testee's proficiency level and the difficulty of items affects the rate of disengaged (rapid guessing) responses. In particular, low-performing testees showed higher rates of rapid guessing responses if they had to complete relatively hard items. Thus, the adaptive test design matches proficiency and test difficulty, which could prevent or reduce test-taking disengagement.

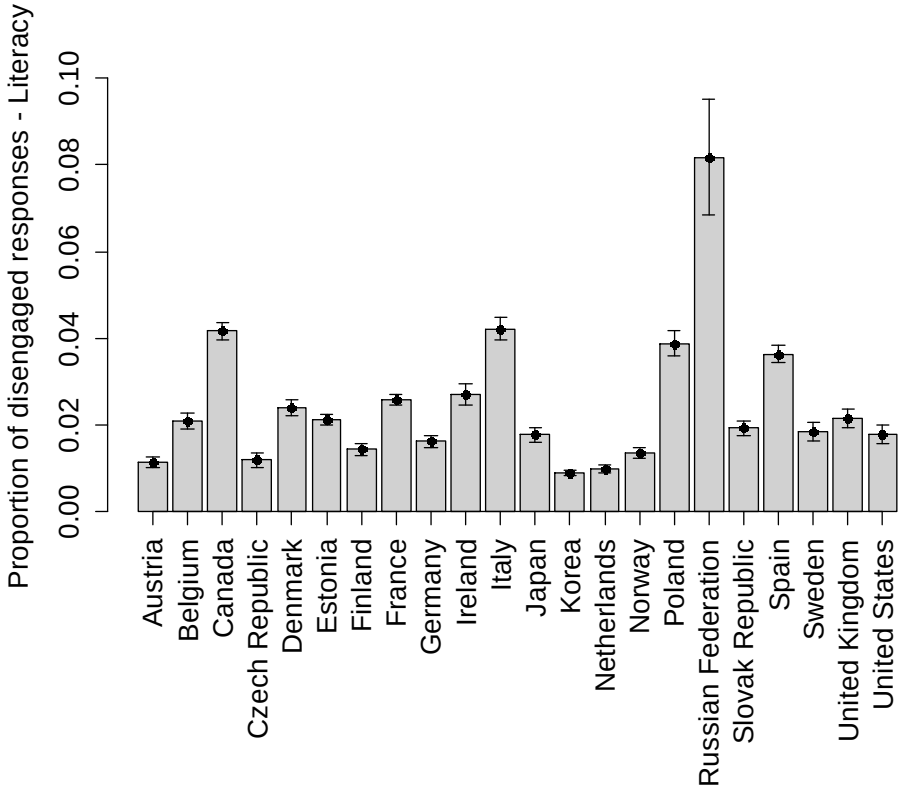
Cross-country differences have to be interpreted cautiously since the results on test-taking engagement are based only on the computer-based version of the PIAAC assessment, and the proportion of testees who opted out of the computer-based assessment differed between countries (cf. OECD, 2013b).

The results presented in the following section are based on the P+>0% method for defining response time thresholds. The tables included in the Appendix (Tables A1, A2, and A3) show the main results for the other methods (5000, 3000, and VI) as well. Overall, findings based on the VI method were very consistent to the ones obtained by means of the P+>0% method, although the absolute level of disengagement may vary between the methods. The constant threshold methods 5000 and 3000 produced quite consistent result patterns for Literacy and Numeracy, but not for Problem solving.

Literacy

Figure 13 shows the average proportion of disengaged responses in Literacy across the 22 countries participating in PIAAC. By far the highest mean was obtained for the Russian Federation, followed by Canada, Italy, Poland, and Spain, whereas the lowest mean was found for Korea. A univariate analysis of variance revealed that the countries differed significantly in the proportion of disengaged responses ($D1$=940.62, $df1$=21, $df2$=1000, p<0.001). The effect size pointed to a medium effect, η^2=0.0479 (η=0.2188, SE_η=0.0352).

Figure 13. Test-taking disengagement in Literacy by country



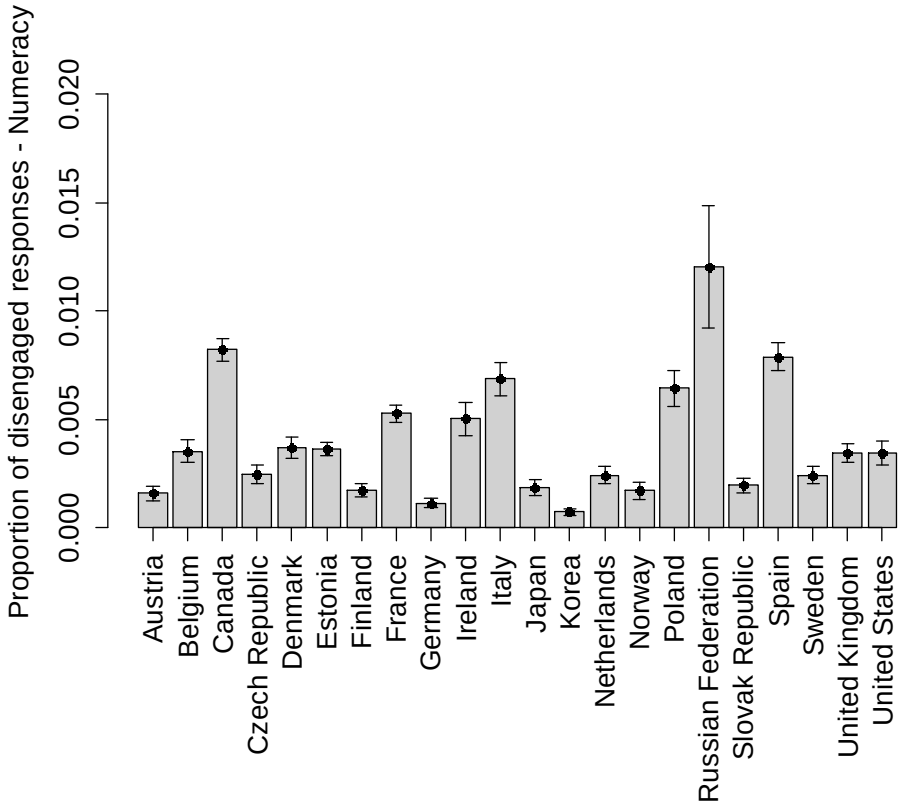
Average proportion of disengaged responses in Literacy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Numeracy

Figure 14 shows the average proportion of disengaged responses in Numeracy. Variation across countries was very consistent with the findings obtained for literacy, although the absolute level of disengagement was much lower for Numeracy. Again, a univariate analysis of variance revealed that countries differed significantly in the proportion of disengaged responses ($D1$=730.14, $df1$=21, $df2$=1000, p<0.001). However, the effect size revealed only a small effect, η^2=0.0144 (η=0.1199, SE_η=0.0225).

Figure 14. Test-taking disengagement in Numeracy by country



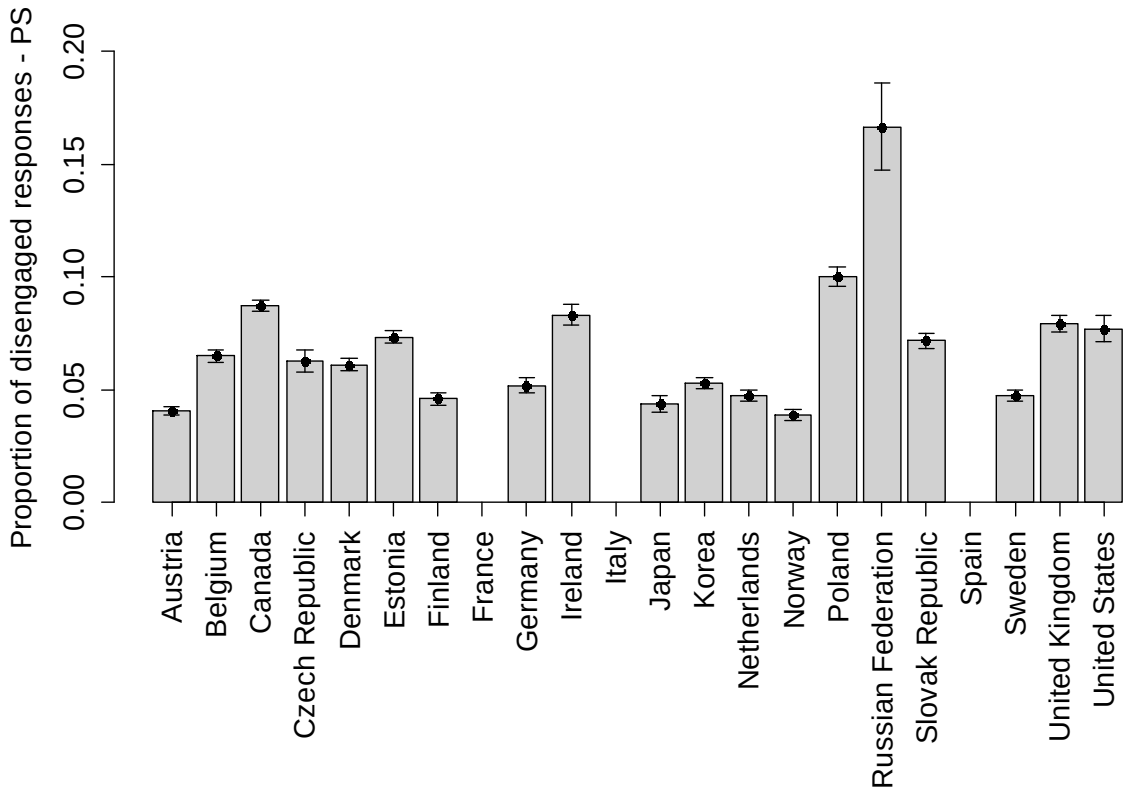
Average proportion of disengaged responses in Numeracy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Problem solving

Figure 15 shows the average proportion of disengaged responses in Problem solving. Variation across countries was very consistent with the findings for Literacy and Numeracy, although the absolute level of disengagement was much higher for Problem solving. As before, a univariate analysis of variance revealed that countries differed significantly in the proportion of disengaged responses ($D1$=880.53, $df1$=18, $df2$=1000, p<0.001). The effect size revealed a medium effect, η^2=0.0596 (η=0.2442, SE_η=0.0374).

Figure 15. Test-taking disengagement in Problem solving (PS) by country



Average proportion of disengaged responses in Problem solving using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included. Note that France, Italy and Spain did not participate in the Problem solving assessment.

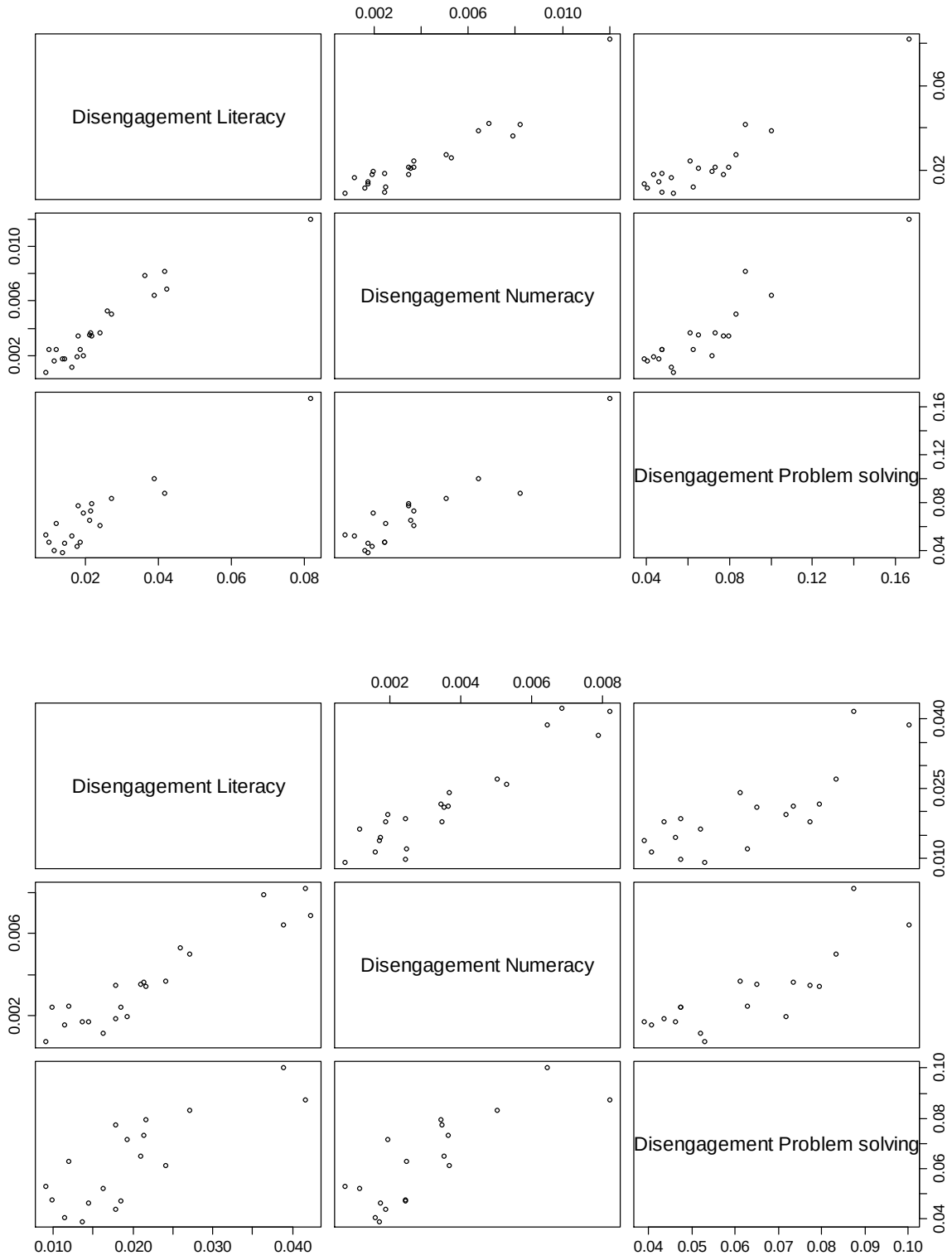
Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Consistency of test-taking engagement across domains

To get an indication of whether test-taking engagement represents a consistent construct across different domains, we determined the association of disengagement across the measurements of Literacy, Numeracy, and Problem solving.

The scatter plots in Figure 16 illustrate that there was high consistency across the three domains - regardless of whether the outlier country the Russian Federation was included (upper part) or not (lower part). The correlations between countries' disengagement in Literacy, Numeracy and Problem solving, corresponding to the upper part of Figure 16, were r=0.95 (p<0.001), r=0.94 (p<0.001), and r=0.92 (p<0.001). When omitting the outlier country (see lower part of Figure 16), the correlations remained very high, r=0.94 (p<0.001), r=0.80 (p<0.001), and r=0.81 (p<0.001).

Figure 16. Association of domain-specific disengagement across countries



Upper part: All countries. Lower part: Without outlier country (the Russian Federation).

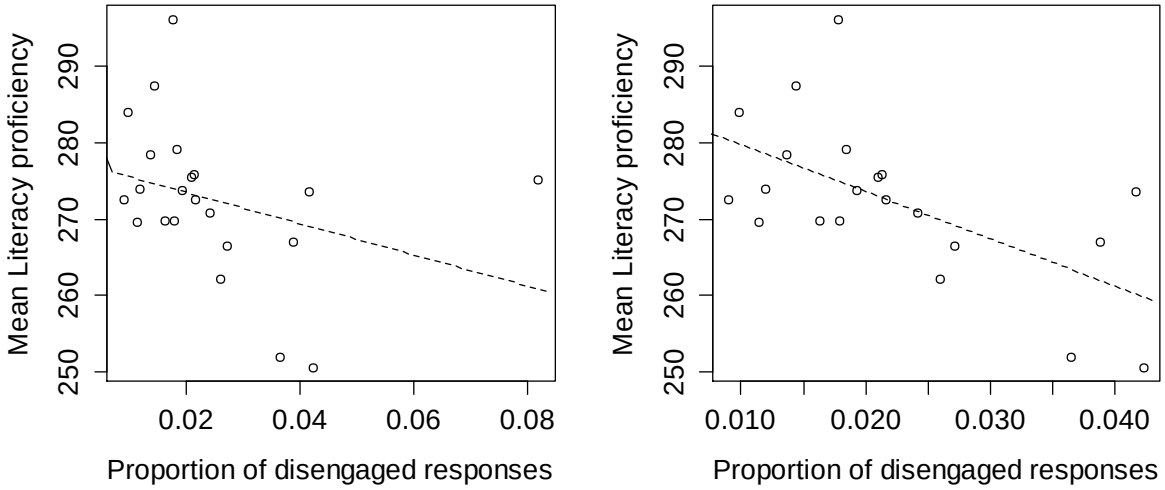
Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Relation of test-taking engagement with proficiency level

The following figures show the association between countries' test-taking disengagement and proficiency level for the domains of Literacy, Numeracy, and Problem solving.

Each respective left part depicts the relation across all 22 countries, whereas in the right part, the outlier country the Russian Federation was omitted. Consistently for Literacy (Figure 17), Numeracy (Figure 18), and Problem solving (Figure 19), a negative relation could be revealed; that is, higher levels of test-taking disengagement were associated with lower levels of proficiency. Put differently, in countries with a higher proportion of response times falling below the response time threshold (defined by means of P+>0%), the proficiency level was lower.

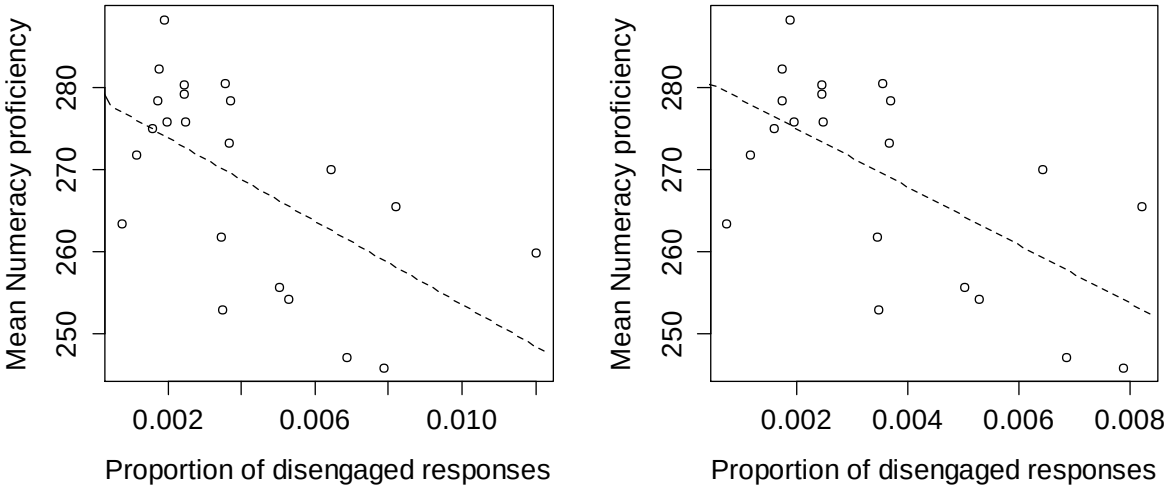
Figure 17. Association between disengagement and proficiency in Literacy across countries



Left part: All countries. Right part: Without outlier country (the Russian Federation).

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

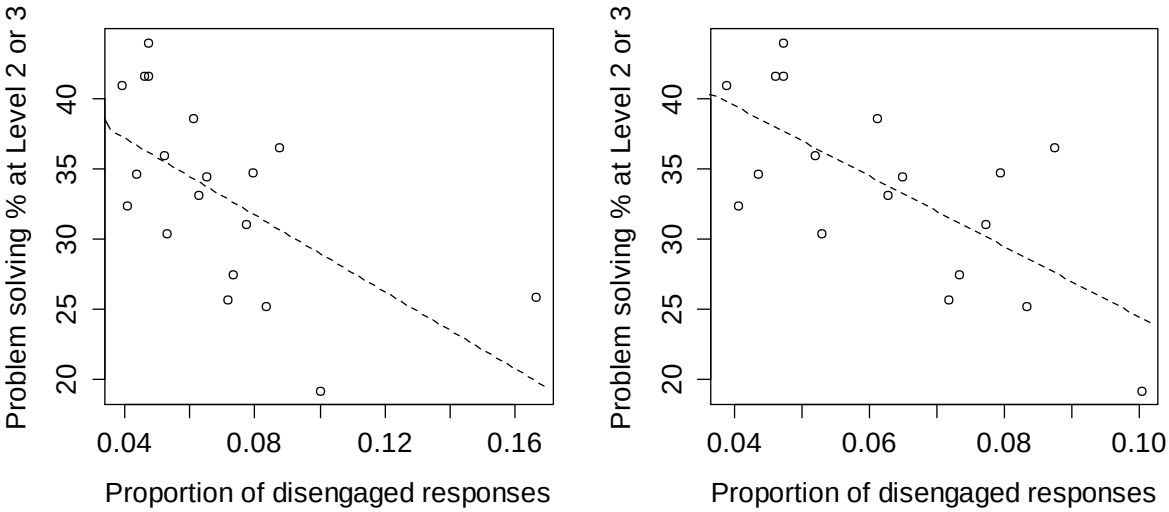
Figure 18. Association between disengagement and proficiency in Numeracy across countries



Left part: All countries. Right part: Without outlier country (the Russian Federation).

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Figure 19. Association between disengagement and proficiency in Problem solving across countries



Left part: All countries. Right part: Without outlier country (the Russian Federation).

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

For Literacy, the correlation of mean disengagement and mean proficiency across all countries was r=-0.33 (p=0.14). When dropping the outlier country of the Russian Federation, it increased substantially

to r=-0.61 (p<0.01). For Numeracy, the correlation of mean disengagement and mean proficiency was r=-0.59 (p<0.01). It increased slightly, r=-0.64 (p<0.01), when omitting the Russian Federation. Finally, for Problem solving, the correlation of mean disengagement and proficiency (represented as % at Level 2 or 3) across all countries was r=-0.61 (p<0.01). It increased slightly, r=-0.69 (p<0.01), when the outlier country the Russian Federation was not included.

Subgroup differences in test-taking engagement

Subgroup differences in test-taking engagement were investigated for gender ("male" and "female"), age ("Aged 24 or less", "Aged 25-34", "Aged 35-54", and "Aged 55 or more"), educational attainment ("Less than high school", "High school", and "Above high school"), and language ("Test language not same as native language", and "Test language same as native language"). The results presented in the following section are based on the P+>0% method for defining response time thresholds.

Gender

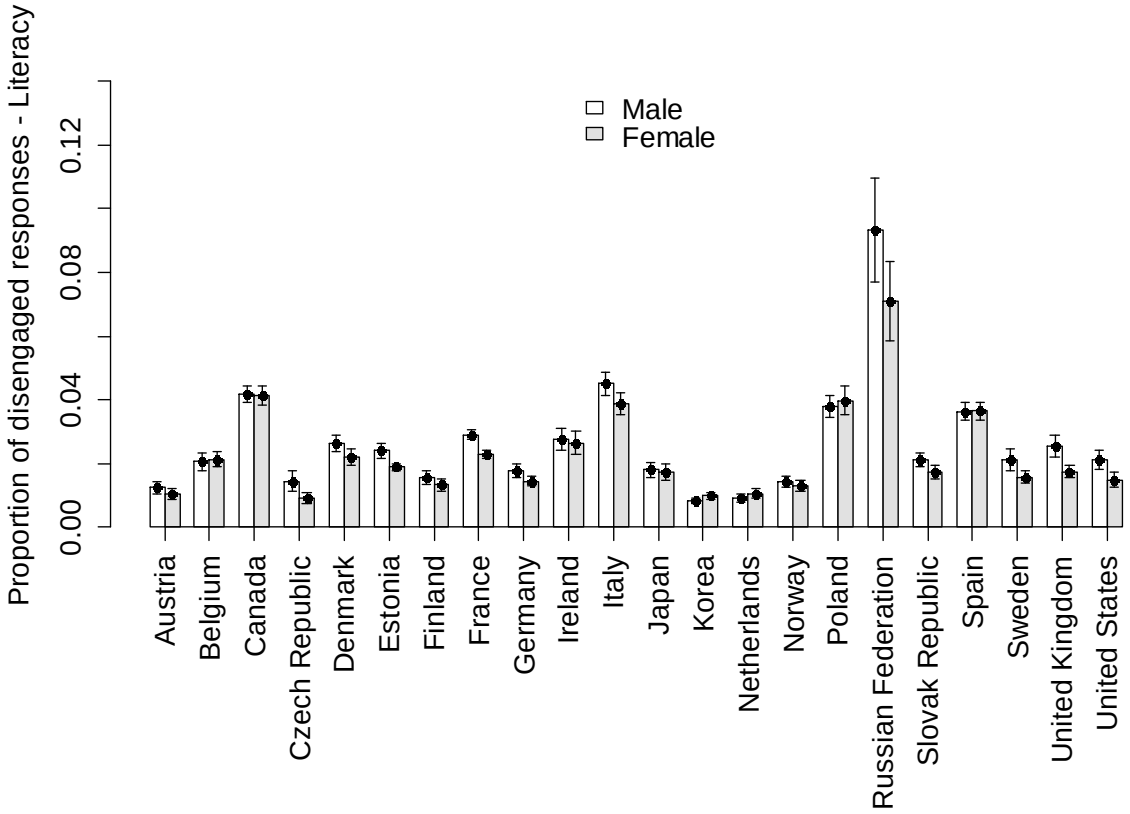
The following results indicate whether test-taking disengagement differs between male and female testees in PIAAC.

Literacy

Across all countries, the proportion of disengaged responses in Literacy was slightly higher for males, M=0.031 (SE=0.002), than for females, M=0.026 (SE=0.002). A univariate analysis of variance revealed that males and females differed significantly in the proportion of disengaged responses ($D1$=9.88, $df1$=1, $df2$=1000, p<0.01). However, the effect size pointed to a negligible effect, η^2=0.0009 (η=0.0294, SE_η=0.009).

Figure 20 shows the difference between males and females in the proportion of disengaged responses in Literacy for each country. The corresponding statistics can be found in Table A4 (see Appendix).

Figure 20. Gender differences in test-taking disengagement in Literacy by country



Average proportion of disengaged responses in Literacy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

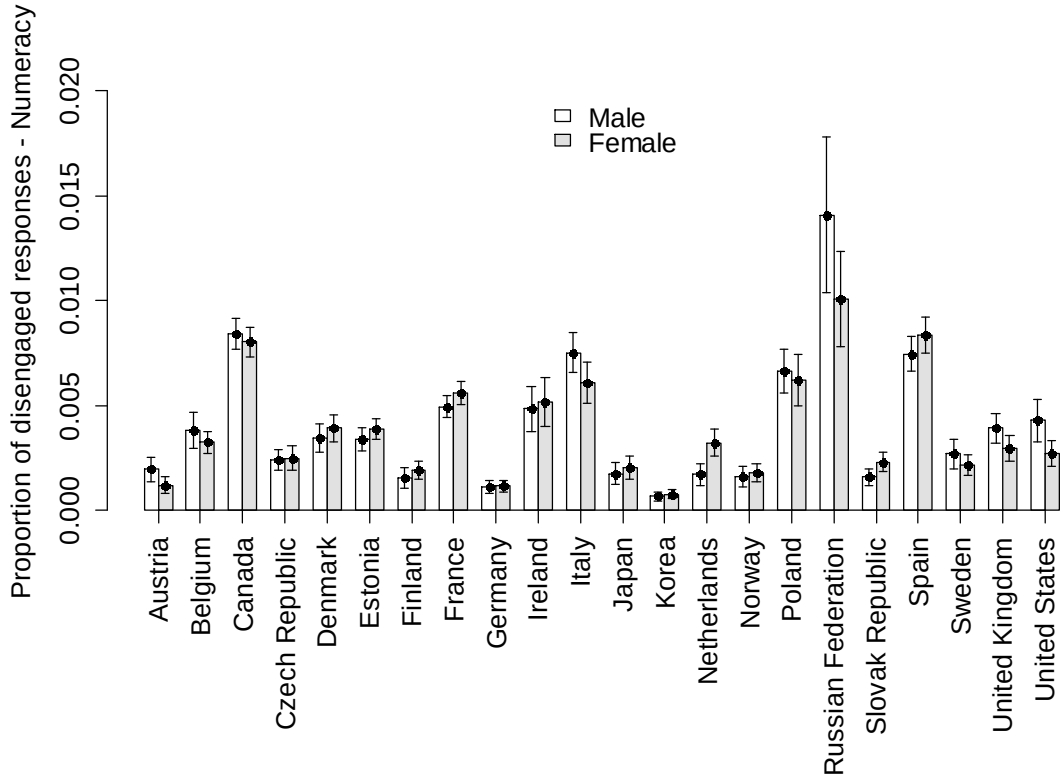
Descriptively, males showed a higher proportion of disengaged responses in most countries. Although in some countries this difference became significant (Estonia, France, the Russian Federation, the United Kingdom, the United States), the effect sizes were very small. Even the greatest effect size of η^2=0.0055 points to only a small effect.

Numeracy

Across all countries, the proportion of disengaged responses in Numeracy was slightly higher for males, *M*=0.005 (*SE*=0.000), than for females, *M*=0.004 (*SE*=0.000). A univariate analysis of variance showed that males and females did not differ significantly in disengagement (*D1*=3.31, *df1*=1, *df2*=1000, *p*=0.07).

Figure 21 shows the difference between males and females in the proportion of disengaged responses in Numeracy for each country. The corresponding statistics can be found in Table A5 (see Appendix).

Figure 21. Gender differences in test-taking disengagement in Numeracy by country



Average proportion of disengaged responses in Numeracy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

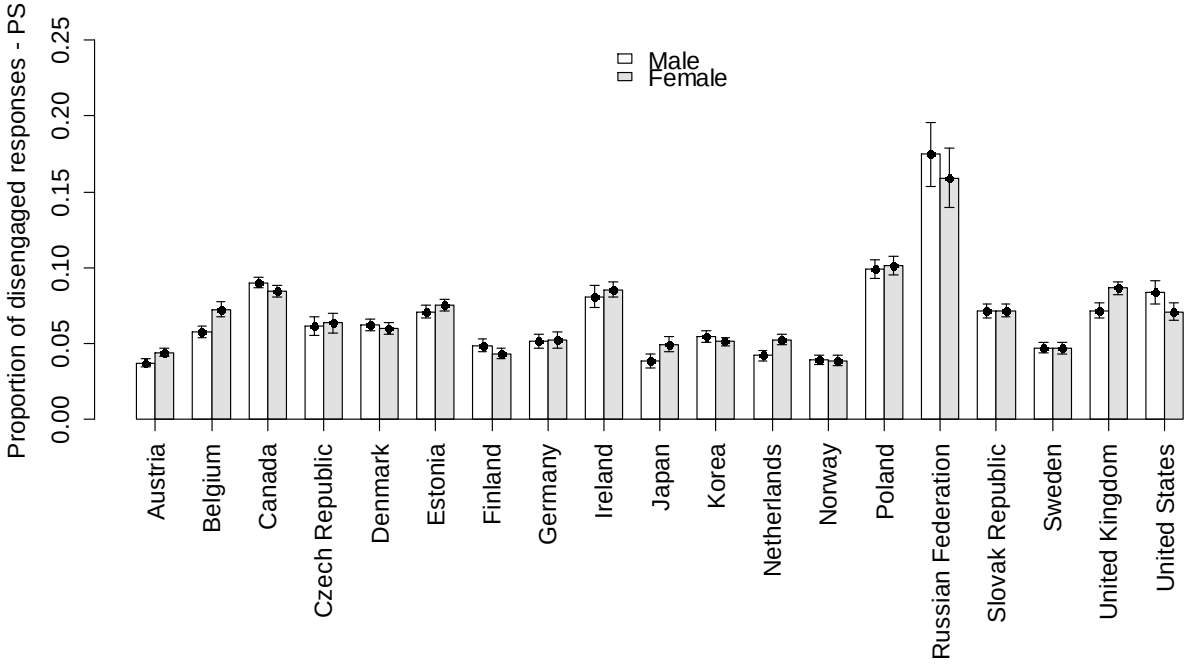
As was the case for Literacy, males showed a slightly higher proportion of disengaged responses at the country level. However, this effect did not reach statistical significance in any country.

Problem Solving

Across all countries, the proportion of disengaged responses in Problem solving was a bit higher for males, M=0.082 (SE=0.004), than for females, M=0.080 (SE=0.003). A univariate analysis of variance revealed that this difference was not significant ($D1$=0.82, $df1$=1, $df2$=1000, p=0.37).

Figure 22 shows the difference between males and females in the proportion of disengaged responses in Problem solving for each country. The corresponding statistics can be found in Table A6 (see Appendix).

Figure 22. Gender differences in test-taking disengagement in Problem solving (PS) by country



Average proportion of disengaged responses in Problem solving using the P+>0% method (error bars indicate +/-1 SE). Only those testee's who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

In some countries, males showed a higher proportion of disengaged responses than females, while in others females were more disengaged than males. Interestingly, statistical significance was achieved for Belgium, the Netherlands, and the United Kingdom, with females being more disengaged than males, whereas in the United States males were significantly more disengaged. However, as for Literacy, effect sizes were very small.

Age

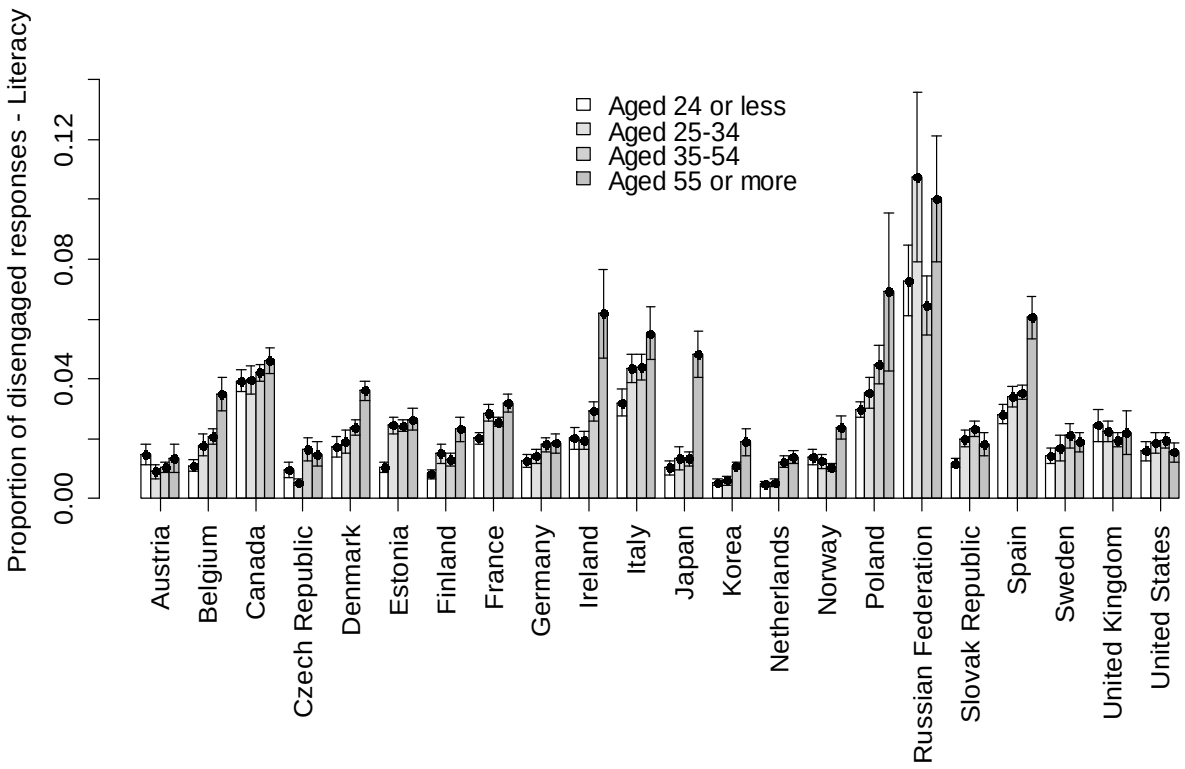
In the following analyses relating testees' age to their disengagement, age was broken up into four groups, that is, "Aged 24 or less", "Aged 25-34", "Aged 35-54", and "Aged 55 or more".

Literacy

Across all countries, disengagement in Literacy varied across age without a clear trend. For the groups "Aged 24 or less", "Aged 25-34", "Aged 35-54", and "Aged 55 or more", the proportions were M=0.025 (SE=0.002), M=0.033 (SE=0.004), M=0.027 (SE=0.001), and M=0.033 (SE=0.002). A univariate analysis of variance showed that the four age groups differed significantly in the proportion of disengaged responses ($D1$=11.85, $df1$=3, $df2$=1000, p<0.001). However, the effect size indicated only a very small effect, η^2=0.0013 (η=0.0355, SE_η=0.0122).

Figure 23 shows differences between age groups in the proportion of disengaged responses in Literacy for each country. The corresponding statistics can be found in Table A7 (see Appendix).

Figure 23. Age differences in test-taking disengagement in Literacy by country



Average proportion of disengaged responses in Literacy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

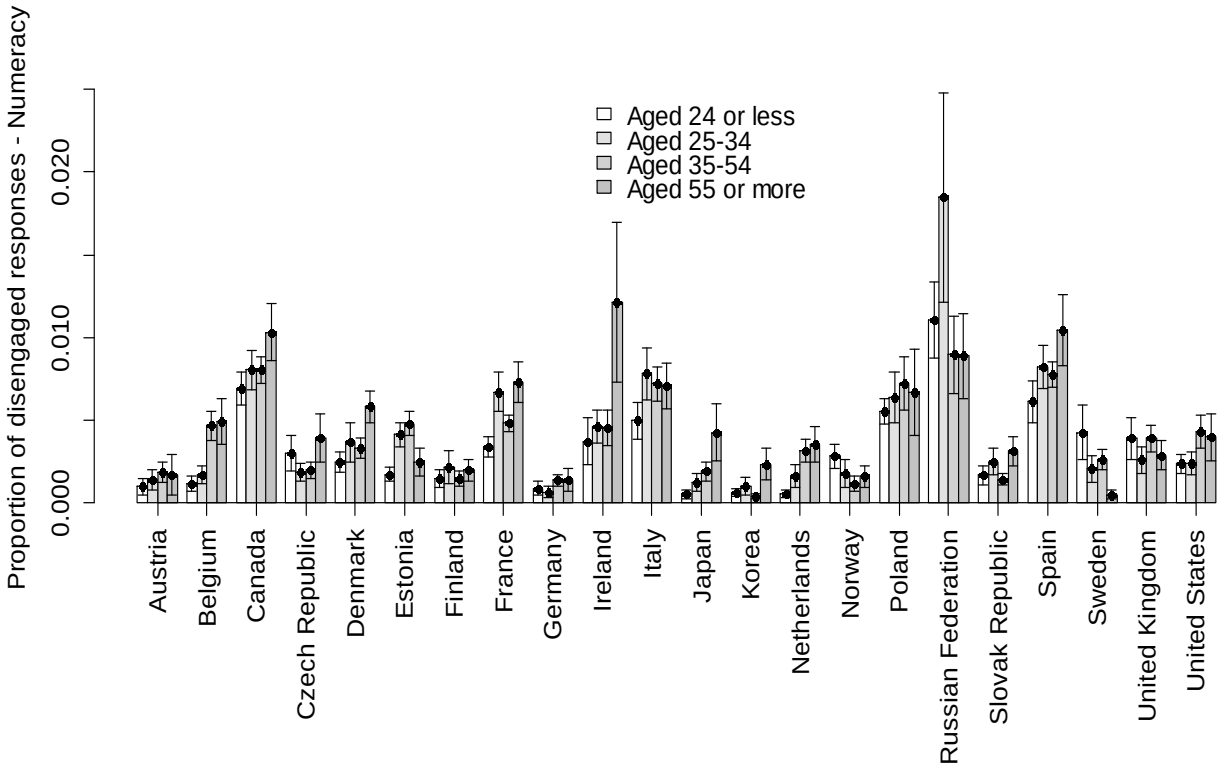
Except for Austria, Canada, the United Kingdom, and the United States, the effect of age was significant in all countries. Interestingly, in some of these countries there was a trend in that disengagement increased across age groups, whereas in other countries no clear trend could be observed. The effect sizes were small at the most. The greatest effect size of η^2=0.0212 was obtained for Japan.

Numeracy

Across all countries, the proportion of disengaged responses in Numeracy hardly varied across age. For the groups "Aged 24 or less", "Aged 25-34", "Aged 35-54", and "Aged 55 or more", the proportions were M=0.004 (SE=0.000), M=0.005(SE=0.001), M=0.005 (SE=0.000), and M=0.005 (SE=0.001). A univariate analysis of variance showed that the four age groups differed significantly in the proportion of disengaged responses ($D1$=7.56, $df1$=3, $df2$=1000, p<0.001). However, the effect size was tiny, η^2=0.0004 (η=0.0199, SE_η=0.0089).

Figure 24 shows differences between age groups in the proportion of disengaged responses in Numeracy for each country. The corresponding statistics can be found in Table A8 (see Appendix).

Figure 24. Age differences in test-taking disengagement in Numeracy by country



Average proportion of disengaged responses in Numeracy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

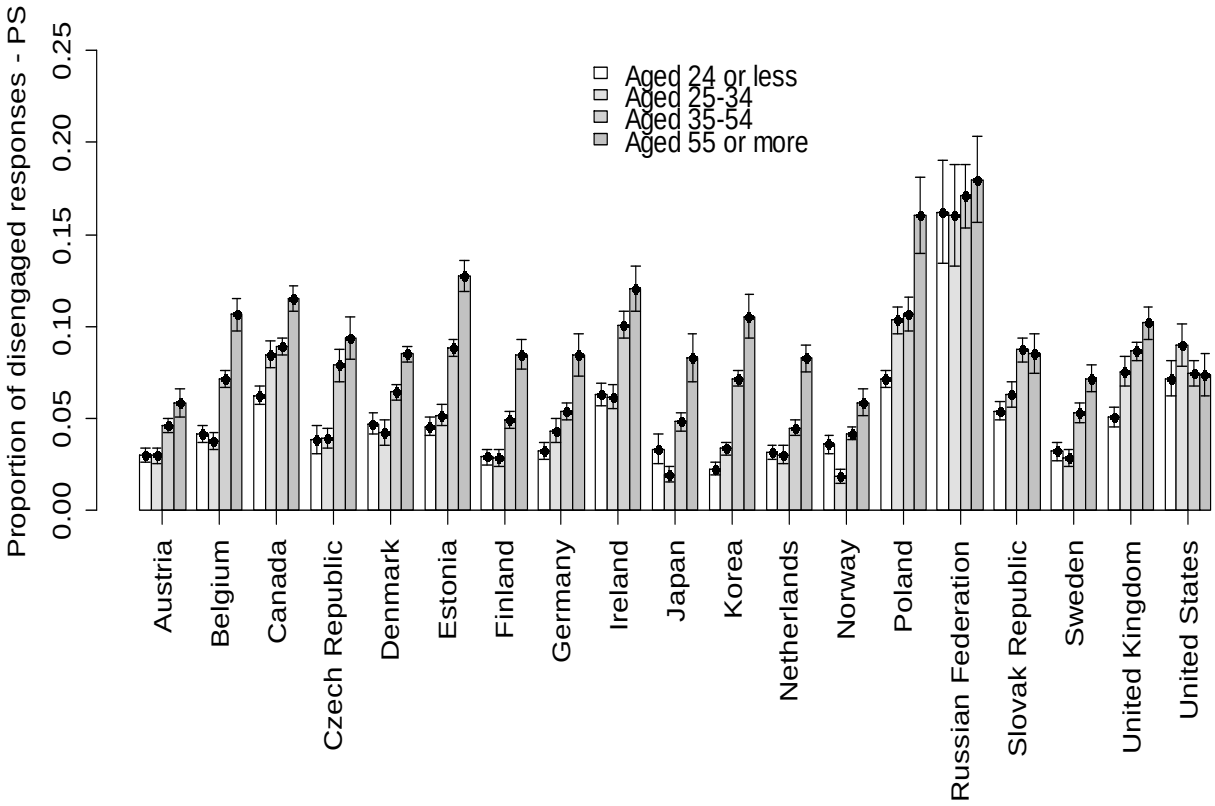
The effect of age was significant in all countries except for Austria, the Czech Republic, Finland, Germany, Poland, and the United Kingdom. In some countries disengagement increased across age groups, though in others it did not. The obtained effect sizes were very small.

Problem Solving

Across all countries, the proportion of disengaged responses in Problem solving seemed to increase with age. For the groups "Aged 24 or less", "Aged 25-34", "Aged 35-54", and "Aged 55 or more", the proportions were M=0.072 (SE=0.005), M=0.082(SE=0.006), M=0.081 (SE=0.003), and M=0.093 (SE=0.007). A univariate analysis of variance showed that the four age groups differed significantly in the proportion of disengaged responses ($D1$=7.33, $df1$=3, $df2$=1000, p<0.001). However, the effect size was very small, η^2=0.0017 (η=0.0413, SE_η=0.0154).

Figure 25 shows differences between age groups in the proportion of disengaged responses in Problem solving for each country. The corresponding statistics can be found in Table A9 (see Appendix).

Figure 25. Age differences in test-taking disengagement in Problem solving (PS) by country



Average proportion of disengaged responses in Problem solving using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

For Problem solving, the effect of age was significant in all countries except for the Russian Federation. Disengagement increased across age groups in most countries, particularly towards the third age group (35-54), and the fourth age group (55 or more). The effect sizes obtained for Problem solving were clearly higher than the ones found for Literacy and Numeracy and ranged from small to medium (Korea).

Educational attainment

In the following analyses on the association between educational attainment and test-taking disengagement, the testees' educational attainment was categorized into "Less than high school", "High school", and "Above high school". In applying these categories, the sample was restricted to those testees who were 25-65 years old, since they can be assumed to have finished their school education.

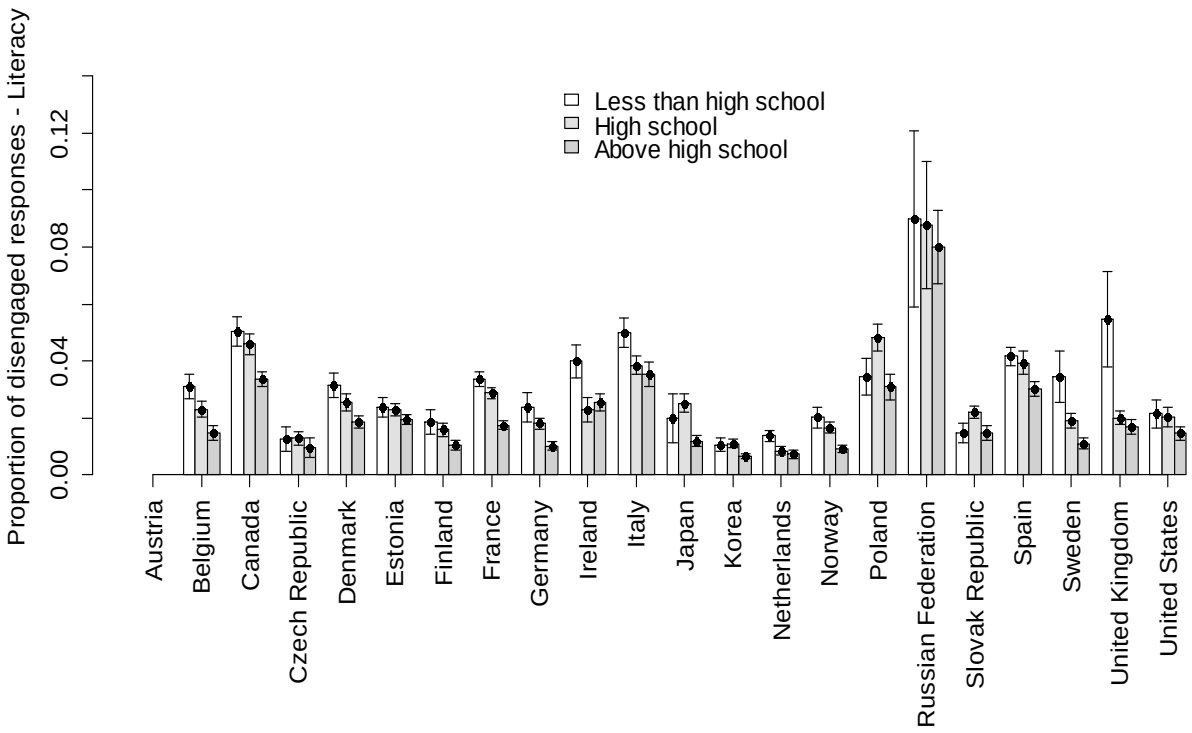
Literacy

Across all countries, the proportion of disengaged responses in Literacy varied across levels of educational attainment. For the group "Less than high school", the proportion was M=0.047 (SE=0.003); for the group "High school", it was M=0.029 (SE=0.002); and for the group "Above high school", it was

M=0.027 (*SE*=0.002). This pattern suggests that with increasing educational attainment, disengagement decreased. A univariate analysis of variance showed that the three educational attainment groups differed significantly in the proportion of disengaged responses (*D1*=20.17, *df1*=2, *df2*=1000, *p*<0.001). However, the effect size indicated only a very small effect, η^2=0.003 (η=0.0552, SE_η=0.0136).

Figure 26 shows differences between educational attainment groups in the proportion of disengaged responses in Literacy for each country.[2] The corresponding statistics can be found in Table A10 (see Appendix).

Figure 26. Educational attainment differences in test-taking disengagement in Literacy by country



Average proportion of disengaged responses in Literacy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Except for the Czech Republic, Estonia, and the Russian Federation, the effect of educational attainment was significant in all countries. In most countries, a trend could be observed suggesting that disengagement decreases with educational attainment. However, obtained effect sizes were very small.

Numeracy

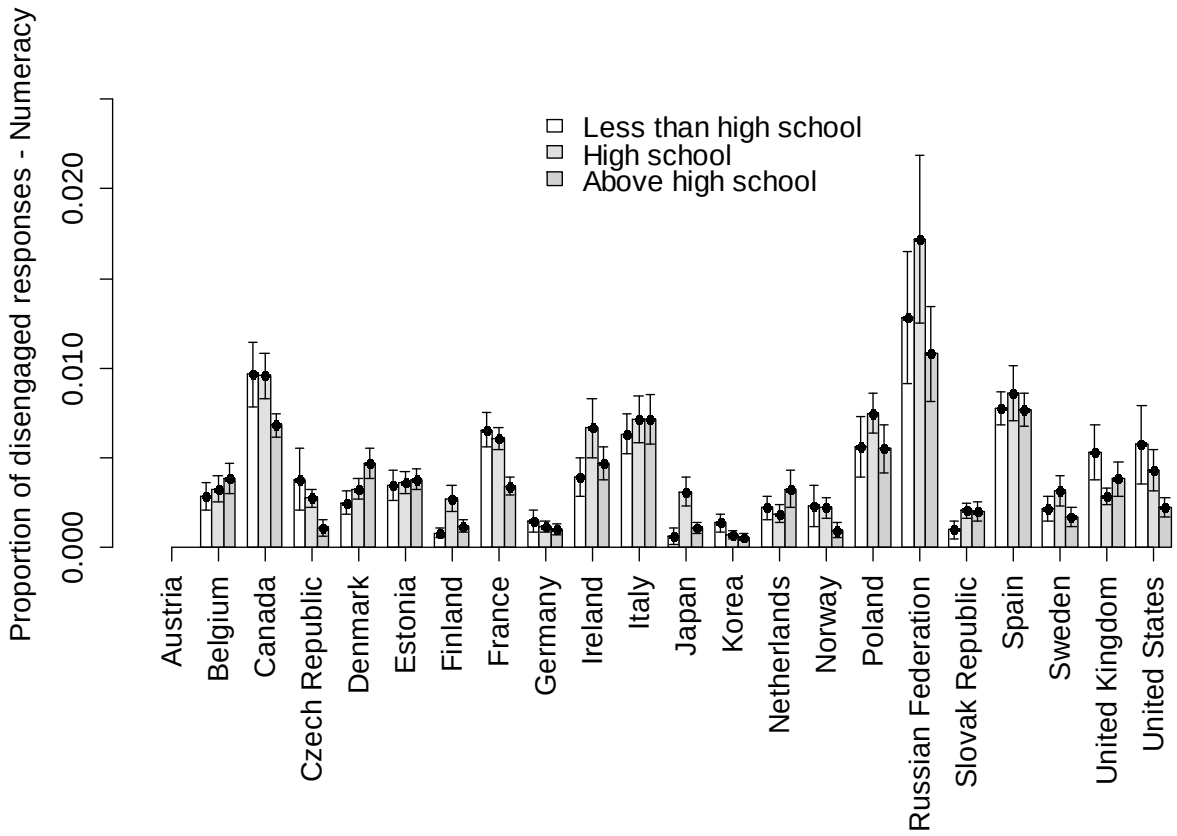
Across all countries, the proportion of disengaged responses in Numeracy varied according to educational attainment. For the group "Less than high school", the proportion was *M*=0.007 (*SE*=0.001); for the group "High school", it was *M*=0.005 (*SE*=0.000); and for the group "Above high school", it was *M*=0.004 (*SE*=0.000). This indicates that disengagement decreased slightly with increasing educational

[2] Note that the public use file version used for the study did not include valid information on educational attainment in Austria. Therefore, no results are reported for Austria.

attainment. A univariate analysis of variance showed that the three educational attainment groups differed significantly in the proportion of disengaged responses (*D1*=6.29, *df1*=2, *df2*=1000, $p<0.001$). However, the effect size revealed only a very small effect, $\eta^2=0.0007$ ($\eta=0.0267$, $SE_\eta=0.0107$).

Figure 27 shows differences between educational attainment groups in the proportion of disengaged responses in Numeracy for each country. The corresponding statistics can be found in Table A11 (see Appendix).

Figure 27. Educational attainment differences in test-taking disengagement in Numeracy by country



Average proportion of disengaged responses in Numeracy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

The effect of educational attainment was significant in about half of the countries. Only in some countries could a trend similar to that of Literacy, suggesting that disengagement decreases with educational attainment, be observed. The obtained effect sizes were very small.

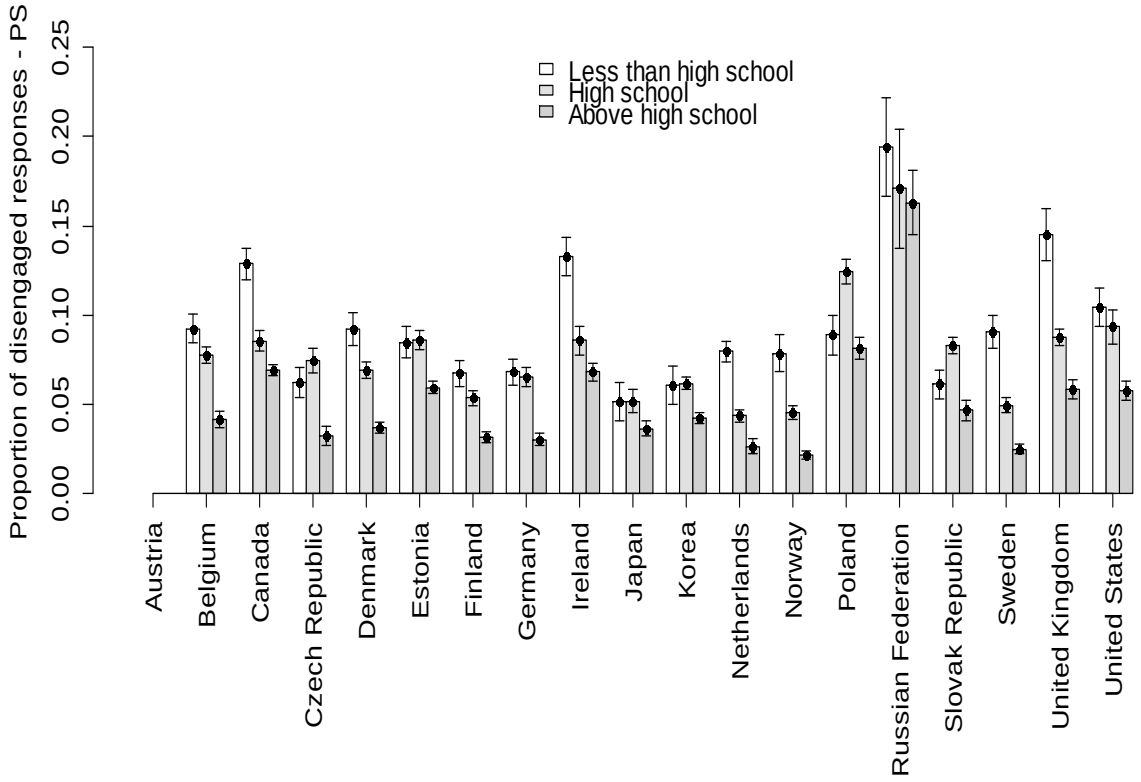
Problem Solving

Across all countries, the proportion of disengaged responses in Problem solving differed across levels of educational attainment. For the group "Less than high school", the proportion was *M*=0.144 (*SE*=0.007); for the group "High school", it was *M*=0.090 (*SE*=0.005); and for the group "Above high school", it was *M*=0.074 (*SE*=0.004). The obtained results indicate that disengagement decreased with increasing educational attainment. A univariate analysis of variance showed that the three educational

attainment groups differed significantly in the proportion of disengaged responses (*D1*=66.20, *df1*=2, *df2*=1000, $p<0.001$). However, the effect size revealed only a small effect, $\eta^2=0.011$ ($\eta=0.1048$, $SE_\eta=0.0145$).

Figure 28 shows differences between educational attainment groups in the proportion of disengaged responses in Problem solving for each country. The corresponding statistics can be found in Table A12 (see Appendix).

Figure 28. Educational attainment differences in test-taking disengagement in Problem solving (PS) by country



Average proportion of disengaged responses in Problem solving using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Except for the Russian Federation, the effect of educational attainment was significant in all countries. Disengagement decreased with greater educational attainment in most countries. The obtained effect sizes were greater than those for Literacy and Numeracy, but still in the range of small effect sizes.

Language

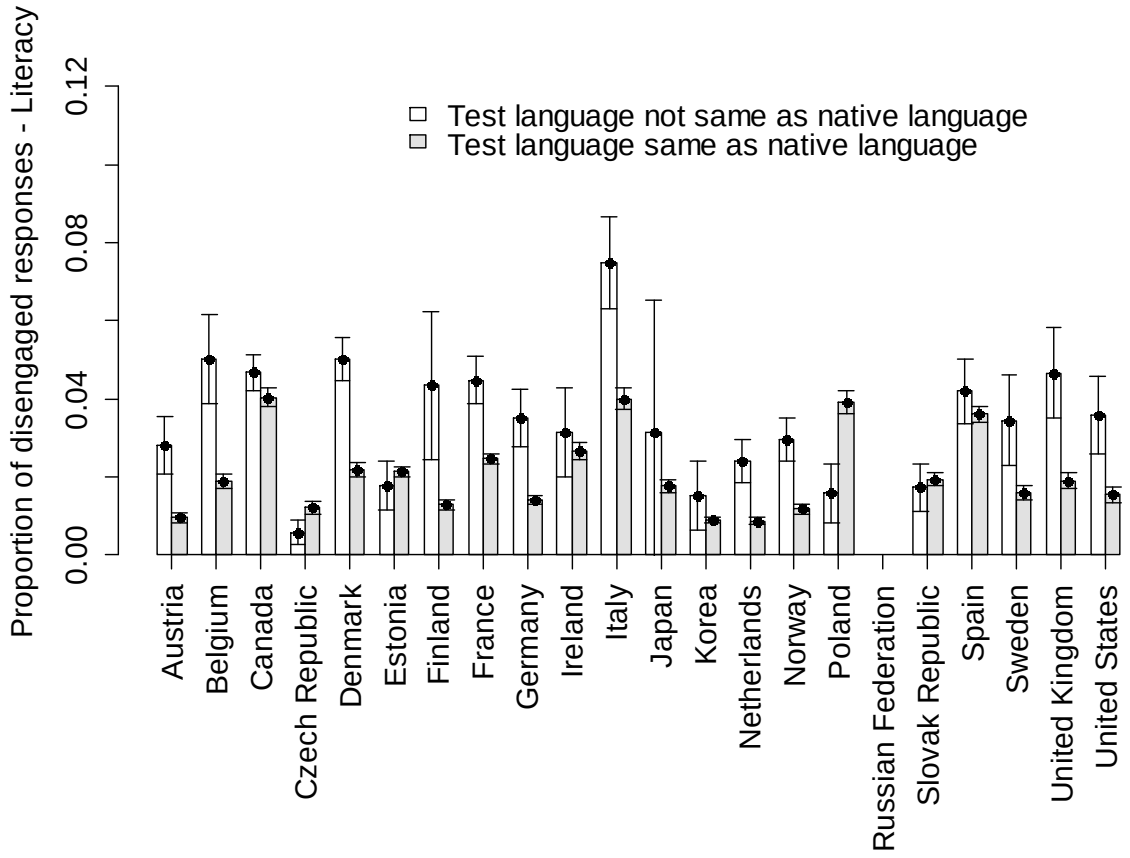
The following analyses investigate whether disengagement differs between testees whose native language was the same as the test language and those whose native language was not the same as the test language.

Literacy

Across all countries, the proportion of disengaged responses in Literacy was higher for testees whose native language was not the same as the test language, M=0.040 (SE=0.005), than for testees whose native language was the same as the test language, M=0.021 (SE=0.001). A univariate analysis of variance showed that this difference was significant ($D1$=18.09, $df1$=1, $df2$=1000, p<0.001). However, the effect size was very small, η^2=0.0042 (η=0.0646, SE_η=0.0147).

Figure 29 shows the difference between "Test language not same as native language" and "Test language same as native language" in the proportion of disengaged responses in Literacy for each country.[3] The corresponding statistics can be found in Table A13 (see Appendix).

Figure 29. Language differences in test-taking disengagement in Literacy by country



Average proportion of disengaged responses in Literacy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

In about half of the countries (i.e. Austria, Belgium, Denmark, France, Germany, Italy, the Netherlands, Norway, the United Kingdom, the United States) testees with native language not the same as test language, showed a significant higher proportion of disengaged responses than testees with

[3] Note that the public use file version used for the study did not include valid information on language for the Russian Federation. Therefore, no results are reported for the Russian Federation.

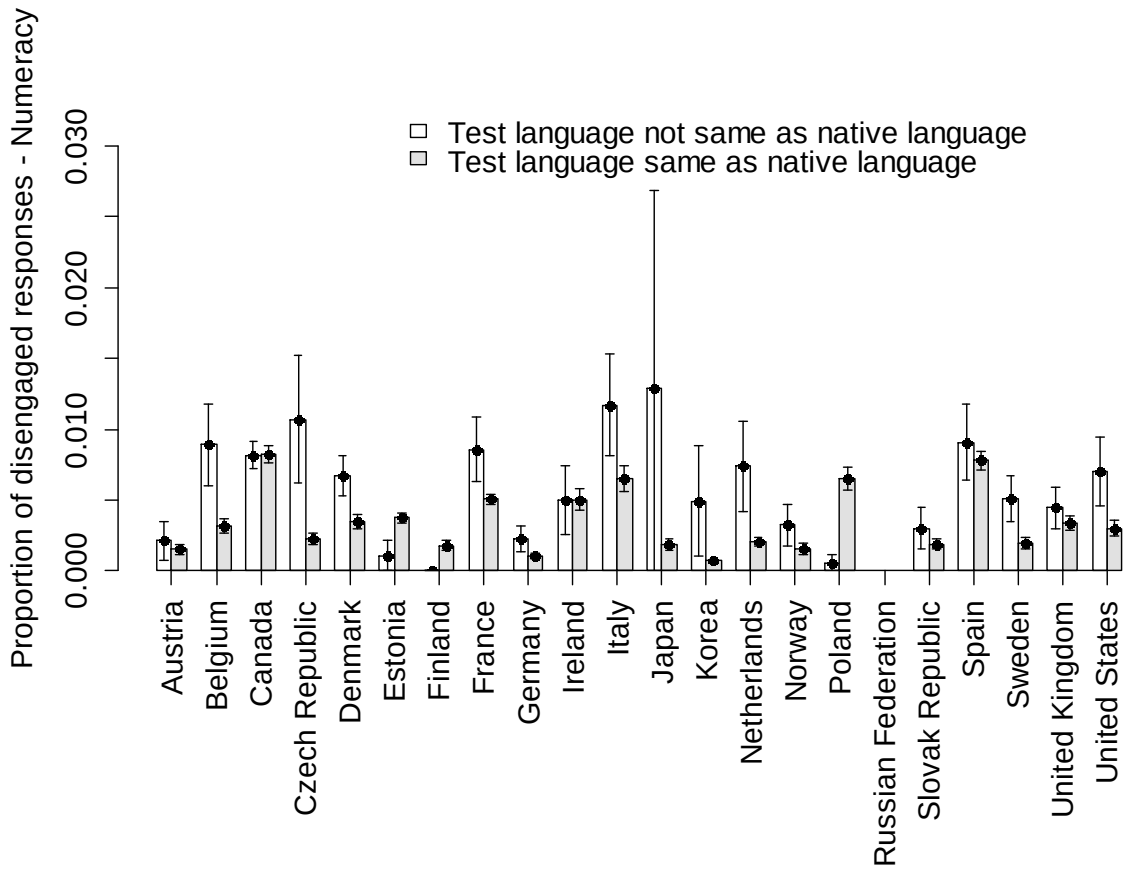
native language the same as test language. Interestingly, in Poland this difference was reversed. Overall, the effect sizes found for significant differences were very small. The greatest effect size of η^2=0.0110 was obtained for the United Kingdom.

Numeracy

Across all countries, the proportion of disengaged responses in Numeracy was slightly higher for testees with native language not the same as test language, *M*=0.007 (*SE*=0.001), than for testees with native language the same as test language, *M*=0.004 (*SE*=0.000). A univariate analysis of variance showed that this difference was significant (*D1*=8.29, *df1*=1, *df2*=1000, *p*<0.01). However, the effect size was very small, η^2=0.0012 (η=0.0348, SE_η=0.0119).

Figure 30 shows for each country the difference between "Test language not same as native language" and "Test language same as native language" in the proportion of disengaged responses in Numeracy. The corresponding statistics can be found in Table A14 (see Appendix).

Figure 30. Language differences in test-taking disengagement in Numeracy by country



Average proportion of disengaged responses in Numeracy using the P+>0% method (error bars indicate +/- 1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

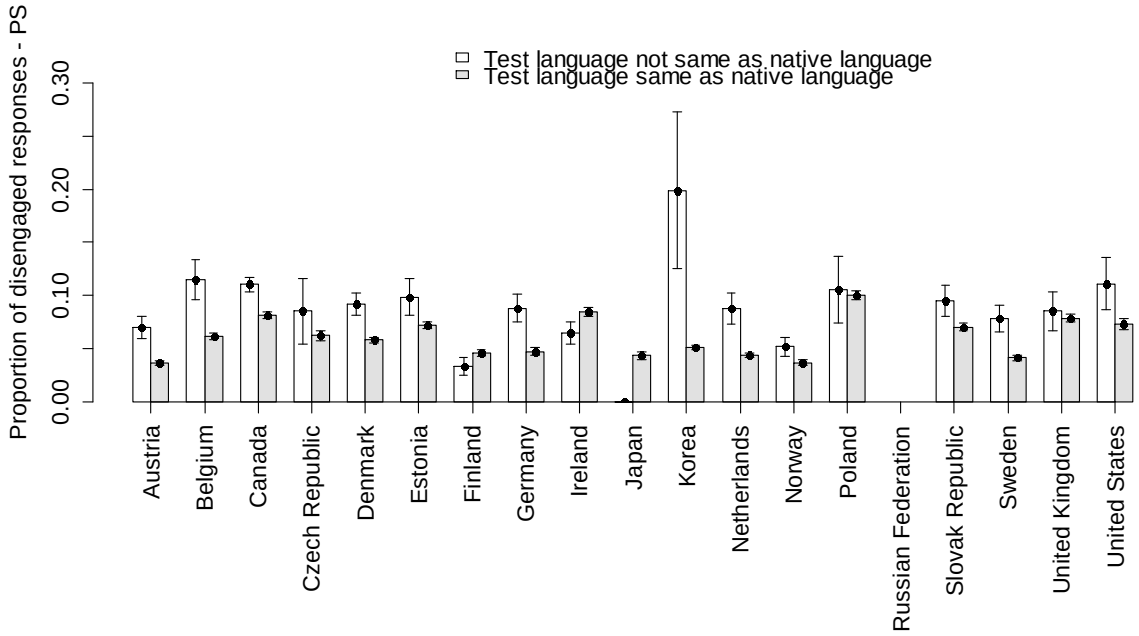
In a few countries (i.e. Belgium, Denmark), testees whose native language was not the same as the test language showed a significant higher proportion of disengaged responses than testees whose native language was the same as the test language. For Estonia and Poland, this difference was reversed. Overall, the effect sizes found for significant differences were very small.

Problem Solving

Across all countries, the proportion of disengaged responses in Problem solving was a bit higher for testees whose native language was not the same as the test language, M=0.103 (SE=0.013), than for testees whose native language was the same as the test language, M=0.064 (SE=0.002). A univariate analysis of variance showed that this difference was significant ($D1$=9.81, $df1$=1, $df2$=1000, p<0.01). However, the effect size was very small, η^2=0.0057 (η=0.0753, SE_η=0.0231).

Figure 31 shows the difference between "Test language not same as native language" and "Test language same as native language" in the proportion of disengaged responses in Problem solving for each country. The corresponding statistics can be found in Table A15 (see Appendix).

Figure 31. Language differences in test-taking disengagement in Problem solving (PS) by country



Average proportion of disengaged responses in Problem solving using the P+>0% method (error bars indicate +/-1 SE). Only those testees who did not opt out of the computer-based assessment were included.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

In about half of the countries (i.e. Austria, Belgium, Canada, Denmark, Germany, Korea, the Netherlands, Sweden), testees whose native language was not the same as the test language showed a significant higher proportion of disengaged responses than testees whose native language was the same as the test language. As for the other domains, the effect sizes found for significant differences were very small.

DISCUSSION

In this study, we addressed the question of how empirical indicators of test-taking engagement can be defined and empirically validated in the context of PIAAC. Furthermore, we sought to use these findings to describe differences in test-taking engagement in PIAAC at the country and the subgroup level. The general approach for the definition of an indicator was to distinguish disengaged response behavior from engaged response behavior by means of response time thresholds. Disengaged response behavior is assumed to reflect low effort, preventing task success above the chance level, whereas engaged response behavior is characterized by effort, which enables task success above the chance level (depending on individual proficiency). Various methods for the definition of constant response time thresholds (3000 ms and 5000 ms) and item-specific response time thresholds (VI, P+>0%) were considered. Overall, validity checks showed that the P+>0% method performed slightly better than the VI method, followed by the methods assuming constant thresholds of 3000 ms and 5000 ms, respectively.

The results for Literacy and Numeracy by module revealed that the extent of disengaged response behavior depended on whether items were completed in Module 1 or Module 2. More specifically, there was an increase from Module 1 to Module 2 in the proportion of responses classified as disengaged, suggesting a drop in test-taking effort from Module 1 to 2. An increase in disengaged response behavior over time could also be observed for Problem solving; however, given the test design, this increase may be confounded with the item content presented in Modules 1 and 2.

The investigation of country differences in test-taking engagement by domain based on the P+>0% method revealed that the proportion of disengaged responses was quite low, even though PIAAC is a low-stakes assessment. This is consistent with the findings by Lee and Jia (2014) reporting proportions of less than 2% for a NAEP study, whereas Wise and DeMars (2006) showed more than 5% and Wise, Pastor, and Kong (2009) even more than 10% for other low-stakes assessments. Our findings based on the P+>0% method revealed that in Literacy the proportion of disengaged responses was well below 5% across the majority of countries, while in Numeracy the proportion was even smaller than 1% for almost all countries. However, for Problem solving the P+>0% method revealed a proportion of disengaged responses of more than 5% but usually well below 10%.

Between domains there were substantial differences in the proportion of disengagement as shown by the P+>0% method and also by the VI method. The greatest proportion was found for Problem solving and the lowest for Numeracy; simultaneously, population differences were very consistent across populations. These substantial differences between domains may reflect differences in the complexity of the item material and item difficulty (and, in turn, the items' time intensity). This might not only lengthen the period of time in which disengaged responses can be observed, but also increase the rate of disengaged responses. There were also remarkable differences in the proportion of disengagement between methods. In particular, in Problem solving the P+>0% method was associated with the greatest proportion of disengaged response behavior compared to the 3000 or the 5000 method. In Problem solving, it certainly takes much more than 5000 ms to have the chance to obtain a correct response, and obviously a great number of testees skipped items before reaching this threshold but after having spent some time (> 5000 ms) taking a look at the item. Thus, to some extent being disengaged might be the result of an informed decision. Nevertheless, this behavior should still be considered disengagement. To strengthen this interpretation, the relation between person-specific item difficulty (i.e. the difference between proficiency and item difficulty) and disengagement in an item could be investigated.

There were significant differences in test-taking engagement across countries which were most pronounced for Problem solving, followed by Literacy and Numeracy. However, the obtained effect sizes

were small (Numeracy) or medium (Problem solving and Literacy). Population differences in test-taking engagement proved to be highly correlated between the three domains. If a population shows a high average level of test-taking engagement in one measure, this could also be expected for another one (and vice versa). Thus, test-taking engagement can be conceived of as a characteristic that is highly consistent across domains. Most interestingly, there was a clear negative association between test-taking disengagement and proficiency in Literacy, Numeracy and Problem solving, respectively. Thus, higher test-taking engagement was related to higher proficiency.

The analysis of subgroup differences revealed non-significant differences or only very small effects. Regarding gender, there was the tendency for males to show higher disengagement than females. Interestingly, in some countries this was reversed for Problem solving; that is, females were more disengaged than males. This might reflect gender differences in ICT engagement, with males showing higher levels of ICT interest and self-concept (e.g. Christoph, Goldhammer, Zylka and Hartig, 2015). There was no clear association between age and disengagement for Literacy and Numeracy. However, for Problem solving disengagement tended to increase with age, particularly for the age groups "Aged 35-54", and "Aged 55 or more". As regards educational attainment, there was some evidence that disengagement decreases with increasing educational attainment. In particular, this pattern could be observed for Literacy and Problem solving. Finally, disengagement seemed to be associated with language. Testees whose native language was the same as the test language showed lower disengagement than those whose native language was not the same as the test language. This difference was more pronounced for Literacy and Problem solving than for Numeracy.

Limitations

A potential limitation of using response time thresholds for detecting disengaged responses is that there may be disengaged responses with long response times, for instance, if in the interview situation the testee pretends to be engaged with task completion without making an effort. Such disengaged responses cannot be discovered using the current approach.

The design of the PIAAC study does not allow for causal inferences. That is, the direction of associations between disengagement and other variables such as proficiency and background variables cannot be clarified by means of the presented data and analysis. For instance, with regard to the substantial relation with proficiency, it may be that higher levels of disengagement give rise to lower levels of proficiency or that lower proficiency causes higher disengagement, since testees predict that enduring effort will probably not pay out.

For further investigations of test-taking engagement, the P+>0% method seems to be the most promising. However, the P+>0% method might be further improved because the chance level for success may be greater than zero for a few items. Among the Numeracy items, there are a few which include MC or MC-like response formats (e.g. selection from a drop-down menu). For instance, if one of four graphical elements needs to be clicked, the chance level can be assumed to be 0.25. Then, the response time threshold would be defined as the lowest response time which is associated with the proportion correct being greater than this chance level (cf. Lee and Jia, 2014).

The measure of individual differences in test-taking disengagement was obtained by averaging across (different sets of) items. Strictly speaking, this kind of aggregation assumes that a measurement model holds, justifying computing the average. Thus, future research should additionally use a model-based approach to individual differences in test-taking disengagement (as it is typically done for item response data).

Implications

In principle, there are two ways dealing with disengaged responses: i) avoiding them by choosing appropriate testing conditions, or ii) identifying and taking them into account when deriving test scores in such a way so as to make valid inferences possible. The first way could imply the use of incentives, reducing testing time or adapting item difficulty to the testee's proficiency level. Current research suggests that there is no best practice on how to shape testing conditions (for an overview see e.g. Bridgin, 2015). For instance, the effect of incentives on enhancing test-taking engagement seems to depend on various factors such as grade level and age. Thus, in any case, it is important to consider disengaged responses in the phase of scoring and data analysis. Such a method does not reduce disengagement, but rather the impact of individual differences in disengagement on derived test scores.

In the first round of PIAAC, omitted responses (without an interaction) were scored as wrong if the testee spent more than five seconds on an item. If the testee spent less than five seconds on an item, it was considered not attempted and treated as an ignorable missing value (OECD, 2013a). In the future, this procedure could be adapted by applying item-specific response time thresholds as they have been empirically defined in the present study. They allow for the separation of non-responses without engagement (i.e. missing response) from those with engagement (i.e. wrong response). As a further next step, the scoring of any observation could take into account item-specific thresholds regardless of whether there was an interaction or not. Even if an interaction with the stimulus occurred, as long as it is associated with a short response time below the threshold, the item may be scored as not attempted in an engaged way (i.e. missing response).

REFERENCES

Braun, H., Kirsch, I., Yamamoto, K., Park, J., and Eagan, M. K. (2011), An experimental study of the effects of monetary incentives on performance on the 12th-grade NAEP reading assessment. *Teachers College Record,* Vol. 113, pp. 2309-2344.

Brennan, R. L. (Ed.) (2006), *Educational measurement* (4th ed.). Westport, CT. American Council on Education/Praeger.

Bridgin, F. (2015), *Measuring Motivation in Low-Stakes Assessments*. ETS Research Report No. RR-15-19. Princeton, NJ: Educational Testing Service (ETS).

Christoph, G., Goldhammer, F., Zylka, J. and Hartig, J. (2015), Adolescents' computer performance: The role of self-concept and motivational aspects. *Computers & Education*, Vol. 81, pp. 1-12.

Cohen. J. (1988), *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

DeMars, C. E. (2007), Changes in rapid-guessing behavior over a series of assessments. *Educational Assessment*, Vol. 12, pp. 23-45.

Enders, C. K. (2010), *Applied missing data analysis*. Guilford Press.

Hauser, C., and Kingsbury, G.G. (2009), *Individual score validity in a modest-stakes adaptive educational testing setting*. San Diego, CA: Paper presented at the annual meeting of the National Council on Measurement in Education.

Kong, X., Wise, S.L., and Bhola, D.S. (2007), Setting the response time threshold parameter to differentiate solution behavior from rapid-guessing behavior. *Educational and Psychological Measurement*, Vol. 67, pp. 606–619.

Lee, Y.-H., and Jia, Y. (2014), Using response time to investigate students' test-taking behaviors in a NAEP computer-based study. *Large-scale Assessments in Education,* Vol. 2, pp. 1-24.

Ma, L., Wise, S.L., Thum, Y.M., and Kingsbury, G. (2011), *Detecting response time threshold under the computer adaptive testing environment*. New Orleans, LA: Paper presented at the annual meeting of the National Council of Measurement in Education.

Organisation for Economic Co-Operation and Development (OECD) (2013a), *Technical Report of the Survey of Adult Skills (PIAAC)*. Paris: OECD: http://www.oecd.org/site/piaac/_Technical%20Report_17OCT13.pdf

Organisation for Economic Co-Operation and Development (OECD) (2013b), *OECD Skills Outlook 2013: First Results from the Survey of Adult Skills*, OECD Publishing, Paris. http://dx.doi.org/10.1787/9789264204256-en

R Core Team (2014), R: *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. URL: www.R-project.org/.

Revelle, W. (2015), *psych: Procedures for Personality and Psychological Research*, Northwestern University, Evanston, Illinois, USA, http://CRAN.R-project.org/package=psych Version = 1.5.1.

Robitzsch, A., and Oberwimmer, K. (2015), BIFIEsurvey: Tools for Survey Statistics in Educational Assessment. R package version 1.2-6. URL http://CRAN.R-project.org/package=BIFIEsurvey

Setzer, J. C., Wise, S. L., van den Heuvel, J. R., and Ling, G. (2013), An investigation of examinee test-taking effort on a large-scale assessment. *Applied Measurement in Education*, Vol. 26, pp. 34-49.

Shrout, P. E. and Fleiss, J. L. (1979), Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, Vol. 86, pp. 420-3428.

Wise, S. (2006), An investigation of the differential effort received by items on a low-stakes computer-based test. *Applied Measurement in Education*, Vol. 19, pp. 95–114.

Wise, S., and DeMars, C. (2006), An application of item response time: the effort-moderated IRT model. *Journal of Educational Measurement*, Vol. 43, pp. 19–38.

Wise, S, and Kong, X. (2005), Response time effort: a new measure of examinee motivation in computer-based tests. *Applied Measurement in Education*, Vol. 18, pp. 163–183.

Wise, SL, and Ma, L. (2012), *Setting response time thresholds for a CAT item pool: the normative threshold method*. Vancouver, Canada: Paper presented at the annual meeting of the National Council on Measurement in Education.

Wise, S., Pastor, D.A., and Kong, X. (2009), Correlates of rapid-guessing behavior in low- stakes testing: Implications for test development and measurement practice. *Applied Measurement in Education*, Vol. 22, pp. 185–205.

APPENDIX

Country differences in test-taking engagement

Table A1. Average percentage of disengaged response behavior in Literacy by country

Countries	Mean proficiency Score (*SE*)[1]		Mean proportion of disengagement by method (*SE*)								Opted out of CBA % (*SE*)[2]	
			5000		3000		VI		P+>0%			
Australia	280.4	(0.9)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	13.7	(0.6)
Austria	269.5	(0.7)	0.008	(0.001)	0.004	(0.001)	0.014	(0.002)	0.011	(0.001)	11.3	(0.5)
Belgium	275.5	(0.8)	0.015	(0.002)	0.006	(0.001)	0.026	(0.002)	0.021	(0.002)	4.7	(0.3)
Canada	273.5	(0.6)	0.034	(0.002)	0.016	(0.001)	0.052	(0.003)	0.042	(0.002)	6.3	(0.3)
Cyprus[3]	268.8	(0.8)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	18.0	(0.5)
Czech Republic	274.0	(1.0)	0.008	(0.001)	0.002	(0.000)	0.017	(0.002)	0.012	(0.002))	12.1	(0.8)
Denmark	270.8	(0.6)	0.021	(0.002)	0.008	(0.001)	0.032	(0.002)	0.024	(0.002)	6.4	(0.3)
Estonia	275.9	(0.7)	0.017	(0.001)	0.006	(0.001)	0.030	(0.002)	0.021	(0.001)	15.8	(0.4)
Finland	287.5	(0.7)	0.012	(0.002)	0.005	(0.001)	0.019	(0.002)	0.014	(0.001)	9.7	(0.4)
France	262.1	(0.6)	0.022	(0.001)	0.006	(0.000)	0.040	(0.002)	0.026	(0.001)	11.6	(0.4)
Germany	269.8	(0.9)	0.011	(0.001)	0.004	(0.001)	0.020	(0.002)	0.016	(0.001)	6.1	(0.5)
Ireland	266.5	(0.9)	0.023	(0.002)	0.010	(0.001)	0.036	(0.003)	0.027	(0.002)	17.4	(0.7)
Italy	250.5	(1.1)	0.040	(0.003)	0.014	(0.001)	0.061	(0.004)	0.042	(0.003)	14.6	(0.9)
Japan	296.2	(0.7)	0.015	(0.002)	0.007	(0.001)	0.020	(0.002)	0.018	(0.002)	15.9	(0.9)
Korea	272.6	(0.6)	0.006	(0.001)	0.002	(0.000)	0.011	(0.001)	0.009	(0.001)	5.4	(0.3)
Netherlands	284.0	(0.7)	0.007	(0.001)	0.002	(0.000)	0.014	(0.001)	0.01	(0.001)	4.5	(0.3)
Norway	278.4	(0.6)	0.009	(0.001)	0.003	(0.000)	0.016	(0.001)	0.014	(0.001)	6.7	(0.4)
Poland	266.9	(0.6)	0.035	(0.003)	0.019	(0.002)	0.045	(0.003)	0.039	(0.003)	23.8	(0.7)
Russian Federation	275.2	(2.7)	0.073	(0.014)	0.029	(0.008)	0.106	(0.017)	0.082	(0.013)	12.8	(1.6)
Slovak Republic	273.8	(0.6)	0.014	(0.002)	0.005	(0.001)	0.026	(0.002)	0.019	(0.002)	12.2	(0.4)
Spain	251.8	(0.7)	0.034	(0.002)	0.010	(0.001)	0.053	(0.003)	0.036	(0.002)	10.7	(0.5)
Sweden	279.2	(0.7)	0.016	(0.002)	0.006	(0.002)	0.021	(0.002)	0.018	(0.002)	5.7	(0.3)
United Kingdom	272.5	(1.0)	0.018	(0.002)	0.007	(0.001)	0.030	(0.003)	0.022	(0.002)	4.5	(0.4)
United States	269.8	(1.0)	0.014	(0.002)	0.006	(0.001)	0.024	(0.003)	0.018	(0.002)	6.3	(0.6)

Notes: [1]From international report (OECD, 2013b, Table A2.2a). [2]From international report (OECD, 2013b, Table A2.10a). *SE*=Standard Error, CBA=Computer-Based Assessment, n/a= this country did not participate in the assessment or did not publish the data. [3]See note 1 on page 10.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A2. Average percentage of disengaged response behavior in Numeracy by country

Countries	Mean proficiency Score (*SE*)[1]		Mean proportion of disengagement by method (*SE*)								Opted out of CBA % (*SE*)[2]	
			5000		3000		VI		P+>0%			
Australia	267.6	(1.0)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	13.7	(0.6)
Austria	275.0	(0.9)	0.007	(0.001)	0.003	(0.001)	0.008	(0.001)	0.002	(0.000)	11.3	(0.5)
Belgium	280.4	(0.8)	0.016	(0.002)	0.008	(0.001)	0.018	(0.002)	0.004	(0.001)	4.7	(0.3)
Canada	265.5	(0.7)	0.036	(0.002)	0.020	(0.001)	0.038	(0.002)	0.008	(0.001)	6.3	(0.3)
Cyprus[3]	264.6	(0.8)	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	18.0	(0.5)
Czech Republic	275.7	(0.9)	0.016	(0.002)	0.006	(0.001)	0.020	(0.002)	0.002	(0.000)	12.1	(0.8)
Denmark	278.3	(0.7)	0.017	(0.002)	0.008	(0.001)	0.018	(0.002)	0.004	(0.000)	6.4	(0.3)
Estonia	273.1	(0.5)	0.018	(0.001)	0.009	(0.001)	0.022	(0.001)	0.004	(0.000)	15.8	(0.4)
Finland	282.2	(0.7)	0.007	(0.001)	0.004	(0.001)	0.008	(0.001)	0.002	(0.000)	9.7	(0.4)
France	254.2	(0.6)	0.025	(0.002)	0.013	(0.001)	0.027	(0.002)	0.005	(0.000)	11.6	(0.4)
Germany	271.7	(1.0)	0.006	(0.001)	0.003	(0.001)	0.008	(0.001)	0.001	(0.000)	6.1	(0.5)
Ireland	255.6	(1.0)	0.021	(0.002)	0.010	(0.001)	0.023	(0.002)	0.005	(0.001)	17.4	(0.7)
Italy	247.1	(1.1)	0.035	(0.003)	0.013	(0.002)	0.041	(0.003)	0.007	(0.001)	14.6	(0.9)
Japan	288.2	(0.7)	0.008	(0.001)	0.004	(0.001)	0.009	(0.001)	0.002	(0.000)	15.9	(0.9)
Korea	263.4	(0.7)	0.004	(0.001)	0.002	(0.000)	0.007	(0.001)	0.001	(0.000)	5.4	(0.3)
Netherlands	280.3	(0.7)	0.009	(0.001)	0.005	(0.001)	0.011	(0.001)	0.002	(0.000)	4.5	(0.3)
Norway	278.3	(0.8)	0.008	(0.001)	0.003	(0.001)	0.009	(0.001)	0.002	(0.000)	6.7	(0.4)
Poland	259.8	(0.8)	0.023	(0.003)	0.012	(0.002)	0.028	(0.003)	0.006	(0.001)	23.8	(0.7)
Russian Federation	269.9	(2.7)	0.095	(0.015)	0.030	(0.006)	0.103	(0.014)	0.012	(0.003)	12.8	(1.6)
Slovak Republic	275.8	(0.8)	0.014	(0.002)	0.004	(0.001)	0.018	(0.002)	0.002	(0.000)	12.2	(0.4)
Spain	245.8	(0.6)	0.047	(0.003)	0.019	(0.002)	0.050	(0.003)	0.008	(0.001)	10.7	(0.5)
Sweden	279.1	(0.8)	0.012	(0.001)	0.006	(0.001)	0.014	(0.001)	0.002	(0.000)	5.7	(0.3)
United Kingdom	261.7	(1.1)	0.018	(0.002)	0.008	(0.001)	0.020	(0.002)	0.003	(0.000)	4.5	(0.4)
United States	252.8	(1.2)	0.016	(0.002)	0.008	(0.001)	0.019	(0.002)	0.003	(0.001)	6.3	(0.6)

Notes: [1]From international report (OECD, 2013b, Table A2.6a). [2]From international report (OECD, 2013b, Table A2.10a). SE=Standard Error, CBA=Computer-Based Assessment, n/a= this country did not participate in the assessment or did not publish the data. [3]See note 1 on page 10.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A3. Average percentage of disengaged response behavior in Problem solving by country

Countries	% at level 2 or 3[1]	Mean proportion of disengagement by method (*SE*)								Opted out of CBA % (*SE*)[2]	
		5000		3000		VI		P+>0%			
Australia	38.0	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	13.7	(0.6)
Austria	32.5	0.003	(0.001)	0.000	(0.000)	0.021	(0.002)	0.041	(0.002)	11.3	(0.5)
Belgium	34.5	0.009	(0.002)	0.000	(0.000)	0.047	(0.003)	0.065	(0.003)	4.7	(0.3)
Canada	36.6	0.025	(0.002)	0.001	(0.000)	0.077	(0.003)	0.087	(0.003)	6.3	(0.3)
Cyprus[3]	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	18.0	(0.5)
Czech Republic	33.1	0.002	(0.000)	0.000	(0.000)	0.043	(0.006)	0.063	(0.005)	12.1	(0.8)
Denmark	38.7	0.008	(0.001)	0.000	(0.000)	0.045	(0.003)	0.061	(0.003)	6.4	(0.3)
Estonia	27.6	0.007	(0.001)	0.000	(0.000)	0.057	(0.003)	0.073	(0.003)	15.8	(0.4)
Finland	41.6	0.005	(0.001)	0.000	(0.000)	0.032	(0.003)	0.046	(0.003)	9.7	(0.4)
France	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	11.6	(0.4)
Germany	36.0	0.006	(0.001)	0.000	(0.000)	0.034	(0.004)	0.052	(0.003)	6.1	(0.5)
Ireland	25.3	0.010	(0.001)	0.000	(0.000)	0.059	(0.004)	0.083	(0.004)	17.4	(0.7)
Italy	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	14.6	(0.9)
Japan	34.6	0.005	(0.001)	0.000	(0.000)	0.037	(0.004)	0.044	(0.004)	15.9	(0.9)
Korea	30.4	0.005	(0.001)	0.000	(0.000)	0.033	(0.002)	0.053	(0.002)	5.4	(0.3)
Netherlands	41.5	0.004	(0.001)	0.000	(0.000)	0.032	(0.003)	0.047	(0.002)	4.5	(0.3)
Norway	41.0	0.005	(0.001)	0.000	(0.000)	0.026	(0.003)	0.039	(0.002)	6.7	(0.4)
Poland	19.2	0.012	(0.002)	0.000	(0.000)	0.081	(0.005)	0.100	(0.004)	23.8	(0.7)
Russian Federation	25.9	0.042	(0.008)	0.000	(0.000)	0.165	(0.022)	0.167	(0.019)	12.8	(1.6)
Slovak Republic	25.6	0.005	(0.001)	0.000	(0.000)	0.049	(0.004)	0.072	(0.003)	12.2	(0.4)
Spain	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	n/a	10.7	(0.5)
Sweden	44.0	0.008	(0.001)	0.000	(0.000)	0.032	(0.003)	0.047	(0.003)	5.7	(0.3)
United Kingdom	34.8	0.009	(0.001)	0.000	(0.000)	0.056	(0.004)	0.079	(0.004)	4.5	(0.4)
United States	31.1	0.011	(0.002)	0.000	(0.000)	0.058	(0.007)	0.077	(0.006)	6.3	(0.6)

Notes: [1]From international report (OECD, 2013b, Table A2.10a). [2]From international report (OECD, 2013b, Table A2.10a). *SE*=Standard Error, CBA=Computer-Based Assessment, n/a= this country did not participate in the assessment or did not publish the data. [3]See note 1 on page 10.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Subgroup differences in test-taking engagement

Gender

Table A4. Gender differences in test-taking disengagement in Literacy by country

Country	M_{male}	SE_{male}	M_{female}	SE_{female}	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0125	0.0020	0.0104	0.0018	0.5296	0.4669	0.0004	0.0192	0.0268
Belgium	0.0206	0.0027	0.0214	0.0023	0.0486	0.8256	0.0000	0.0046	0.0200
Canada	0.0418	0.0026	0.0415	0.0030	0.0092	0.9237	0.0000	0.0014	0.0114
Czech Republic	0.0145	0.0030	0.0092	0.0018	2.0351	0.1540	0.0028	0.0528	0.0345
Denmark	0.0263	0.0027	0.0220	0.0024	1.5567	0.2124	0.0006	0.0236	0.0191
Estonia	0.0240	0.0023	0.0189	0.0012	3.9430	0.0473	0.0012	0.0339	0.0164
Finland	0.0155	0.0021	0.0133	0.0020	0.5956	0.4405	0.0002	0.0153	0.0197
France	0.0291	0.0016	0.0230	0.0014	10.8421	0.0010	0.0014	0.0372	0.0110
Germany	0.0178	0.0021	0.0145	0.0016	1.5005	0.2209	0.0007	0.0258	0.0206
Ireland	0.0278	0.0035	0.0266	0.0035	0.0612	0.8047	0.0000	0.0064	0.0231
Italy	0.0452	0.0036	0.0389	0.0035	1.8672	0.1721	0.0009	0.0301	0.0219
Japan	0.0181	0.0023	0.0173	0.0025	0.0530	0.8179	0.0000	0.0047	0.0201
Korea	0.0083	0.0010	0.0098	0.0010	1.2242	0.2688	0.0003	0.0180	0.0164
Netherlands	0.0091	0.0013	0.0106	0.0015	0.5532	0.4572	0.0003	0.0161	0.0214
Norway	0.0144	0.0016	0.0129	0.0017	0.4390	0.5078	0.0002	0.0128	0.0194
Poland	0.0378	0.0035	0.0398	0.0046	0.1174	0.7319	0.0001	0.0078	0.0212
Russian Federation	0.0932	0.0164	0.0711	0.0125	4.0578	0.0442	0.0055	0.0745	0.0330
Slovak Republic	0.0211	0.0020	0.0175	0.0022	1.6530	0.1988	0.0007	0.0272	0.0215
Spain	0.0364	0.0027	0.0365	0.0026	0.0000	0.9947	0.0000	0.0001	0.0181
Sweden	0.0211	0.0033	0.0157	0.0020	2.3833	0.1230	0.0013	0.0355	0.0214
United Kingdom	0.0256	0.0034	0.0175	0.0018	6.2359	0.0127	0.0028	0.0526	0.0185
United States	0.0212	0.0029	0.0148	0.0023	4.5097	0.0339	0.0019	0.0436	0.0201

Note: Degrees of freedom of *D1*: *df1*=1, *df2*=1000; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A5. Gender differences in test-taking disengagement in Numeracy by country

Country	M_{male}	SE_{male}	M_{female}	SE_{female}	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0020	0.0006	0.0012	0.0004	1.2301	0.2677	0.0005	0.0231	0.0200
Belgium	0.0038	0.0009	0.0032	0.0005	0.3509	0.5537	0.0001	0.0117	0.0190
Canada	0.0084	0.0007	0.0080	0.0007	0.1427	0.7057	0.0000	0.0047	0.0125
Czech Republic	0.0024	0.0005	0.0025	0.0006	0.0094	0.9230	0.0000	0.0020	0.0148
Denmark	0.0035	0.0007	0.0039	0.0007	0.2643	0.6073	0.0001	0.0095	0.0191
Estonia	0.0034	0.0006	0.0039	0.0005	0.3216	0.5708	0.0001	0.0102	0.0181
Finland	0.0016	0.0005	0.0019	0.0004	0.3123	0.5764	0.0001	0.0106	0.0215
France	0.0050	0.0005	0.0056	0.0006	0.7272	0.3940	0.0001	0.0111	0.0127
Germany	0.0011	0.0003	0.0012	0.0003	0.0053	0.9418	0.0000	0.0013	0.0153
Ireland	0.0049	0.0011	0.0052	0.0012	0.0371	0.8472	0.0000	0.0052	0.0215
Italy	0.0075	0.0009	0.0061	0.0010	1.6605	0.1978	0.0005	0.0228	0.0176
Japan	0.0017	0.0005	0.0020	0.0006	0.1418	0.7066	0.0001	0.0088	0.0232
Korea	0.0007	0.0002	0.0008	0.0002	0.1091	0.7413	0.0000	0.0059	0.0145
Netherlands	0.0017	0.0005	0.0032	0.0006	3.1528	0.0761	0.0013	0.0362	0.0194
Norway	0.0016	0.0005	0.0018	0.0004	0.1438	0.7047	0.0000	0.0060	0.0163
Poland	0.0067	0.0011	0.0062	0.0012	0.0819	0.7748	0.0000	0.0065	0.0216
Russian Federation	0.0141	0.0037	0.0101	0.0023	3.0585	0.0806	0.0031	0.0557	0.0277
Slovak Republic	0.0016	0.0004	0.0023	0.0004	1.7683	0.1839	0.0007	0.0271	0.0192
Spain	0.0075	0.0008	0.0084	0.0008	0.6968	0.4041	0.0002	0.0123	0.0147
Sweden	0.0027	0.0007	0.0022	0.0005	0.3118	0.5767	0.0002	0.0129	0.0222
United Kingdom	0.0039	0.0007	0.0030	0.0006	1.0105	0.3150	0.0004	0.0204	0.0197
United States	0.0043	0.0010	0.0027	0.0006	1.6996	0.1926	0.0010	0.0313	0.0236

Note: Degrees of freedom of *D1*: *df1*=1, *df2*=1000; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A6. Gender differences in test-taking disengagement in Problem solving by country

Country	M_{male}	SE_{male}	M_{female}	SE_{female}	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0373	0.0029	0.0443	0.0027	3.2043	0.0737	0.0014	0.0376	0.0212
Belgium	0.0576	0.0037	0.0727	0.0047	5.8918	0.0154	0.0032	0.0563	0.0230
Canada	0.0902	0.0037	0.0847	0.0037	1.0866	0.2975	0.0003	0.0158	0.0152
Czech Republic	0.0619	0.0063	0.0636	0.0068	0.0409	0.8398	0.0001	0.0074	0.0352
Denmark	0.0625	0.0037	0.0599	0.0038	0.2245	0.6357	0.0001	0.0096	0.0204
Estonia	0.0711	0.0043	0.0753	0.0036	0.5722	0.4496	0.0002	0.0145	0.0193
Finland	0.0489	0.0045	0.0436	0.0033	0.8882	0.3462	0.0005	0.0231	0.0242
Germany	0.0518	0.0046	0.0523	0.0052	0.0067	0.9347	0.0000	0.0024	0.0229
Ireland	0.0810	0.0071	0.0855	0.0050	0.2903	0.5902	0.0002	0.0153	0.0287
Japan	0.0384	0.0046	0.0497	0.0049	3.5583	0.0595	0.0022	0.0470	0.0257
Korea	0.0545	0.0038	0.0514	0.0026	0.4231	0.5155	0.0002	0.0144	0.0220
Netherlands	0.0421	0.0031	0.0527	0.0035	5.1797	0.0231	0.0022	0.0470	0.0205
Norway	0.0391	0.0032	0.0388	0.0035	0.0050	0.9434	0.0000	0.0016	0.0165
Poland	0.0992	0.0060	0.1012	0.0060	0.0532	0.8176	0.0000	0.0059	0.0248
Russian Federation	0.1747	0.0214	0.1590	0.0196	1.1469	0.2845	0.0015	0.0387	0.0356
Slovak Republic	0.0715	0.0048	0.0719	0.0041	0.0043	0.9478	0.0000	0.0015	0.0180
Sweden	0.0474	0.0038	0.0472	0.0039	0.0014	0.9697	0.0000	0.0009	0.0205
United Kingdom	0.0720	0.0050	0.0865	0.0044	5.4126	0.0202	0.0027	0.0517	0.0223
United States	0.0836	0.0076	0.0711	0.0056	4.4173	0.0358	0.0018	0.0425	0.0197

Note: Degrees of freedom of D1: df1=1, df2=1000; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Age

Table A7. Age differences in test-taking disengagement in Literacy by country

Country	$M_{<25}$	$SE_{<25}$	M_{25-34}	SE_{25-34}	M_{35-54}	SE_{35-54}	$M_{>54}$	$SE_{>54}$	$D1$	p	η^2	η	SE_η
Austria	0.0148	0.0033	0.0091	0.0024	0.0105	0.0016	0.0135	0.0046	2.6133	0.0500	0.0015	0.0390	0.0236
Belgium	0.0111	0.0019	0.0179	0.0036	0.0207	0.0026	0.0350	0.0057	22.4183	0.0000	0.0076	0.0873	0.0215
Canada	0.0394	0.0038	0.0395	0.0047	0.0421	0.0027	0.0460	0.0044	1.6201	0.1830	0.0003	0.0173	0.0148
Czech Republic	0.0096	0.0027	0.0055	0.0012	0.0165	0.0039	0.0149	0.0042	14.1173	0.0000	0.0085	0.0923	0.0203
Denmark	0.0174	0.0034	0.0189	0.0039	0.0239	0.0026	0.0361	0.0033	20.3962	0.0000	0.0050	0.0709	0.0165
Estonia	0.0105	0.0016	0.0244	0.0029	0.0244	0.0019	0.0266	0.0035	49.2635	0.0000	0.0066	0.0809	0.0111
Finland	0.0082	0.0016	0.0150	0.0033	0.0130	0.0022	0.0232	0.0041	18.2342	0.0000	0.0045	0.0671	0.0198
France	0.0203	0.0020	0.0286	0.0027	0.0257	0.0017	0.0319	0.0030	13.7583	0.0000	0.0020	0.0453	0.0118
Germany	0.0125	0.0021	0.0142	0.0022	0.0180	0.0023	0.0184	0.0033	3.9630	0.0080	0.0014	0.0373	0.0182
Ireland	0.0202	0.0037	0.0195	0.0030	0.0292	0.0032	0.0619	0.0148	13.4966	0.0000	0.0149	0.1222	0.0343
Italy	0.0320	0.0045	0.0436	0.0047	0.0439	0.0041	0.0553	0.0087	7.0471	0.0001	0.0038	0.0615	0.0235
Japan	0.0103	0.0022	0.0136	0.0039	0.0133	0.0023	0.0483	0.0078	21.1568	0.0000	0.0212	0.1457	0.0288
Korea	0.0054	0.0011	0.0060	0.0014	0.0111	0.0013	0.0190	0.0045	19.6541	0.0000	0.0067	0.0820	0.0199
Netherlands	0.0050	0.0009	0.0053	0.0011	0.0124	0.0019	0.0137	0.0022	30.1216	0.0000	0.0060	0.0774	0.0112
Norway	0.0140	0.0026	0.0125	0.0025	0.0104	0.0015	0.0238	0.0037	11.5237	0.0000	0.0062	0.0785	0.0229
Poland	0.0298	0.0027	0.0353	0.0051	0.0447	0.0063	0.0692	0.0264	9.0352	0.0000	0.0062	0.0790	0.0452
Russian Federation	0.0729	0.0118	0.1075	0.0281	0.0645	0.0100	0.1002	0.0211	4.4987	0.0038	0.0162	0.1272	0.0649
Slovak Republic	0.0119	0.0015	0.0201	0.0028	0.0234	0.0027	0.0183	0.0039	20.3100	0.0000	0.0046	0.0677	0.0148
Spain	0.0283	0.0031	0.0342	0.0034	0.0355	0.0023	0.0605	0.0070	18.0175	0.0000	0.0072	0.0848	0.0202
Sweden	0.0142	0.0026	0.0169	0.0044	0.0211	0.0040	0.0189	0.0032	2.6740	0.0461	0.0012	0.0347	0.0221
United Kingdom	0.0245	0.0055	0.0226	0.0035	0.0197	0.0023	0.0222	0.0073	1.1246	0.3380	0.0006	0.0241	0.0209
United States	0.0159	0.0033	0.0187	0.0034	0.0194	0.0026	0.0154	0.0032	1.9365	0.1220	0.0005	0.0234	0.0167

Note: Degrees of freedom of D1: df1=3, df2=1000; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A8. Age differences in test-taking disengagement in Numeracy by country

Country	$M_{<25}$	$SE_{<25}$	M_{25-34}	SE_{25-34}	M_{35-54}	SE_{35-54}	$M_{>54}$	$SE_{>54}$	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0010	0.0005	0.0014	0.0006	0.0019	0.0006	0.0017	0.0012	1.4996	0.2131	0.0004	0.0211	0.0198
Belgium	0.0012	0.0005	0.0017	0.0005	0.0047	0.0009	0.0049	0.0014	17.3734	0.0000	0.0041	0.0639	0.0107
Canada	0.0070	0.0010	0.0080	0.0012	0.0081	0.0008	0.0103	0.0017	2.7737	0.0404	0.0007	0.0259	0.0163
Czech Republic	0.0030	0.0010	0.0019	0.0005	0.0020	0.0005	0.0040	0.0015	2.3080	0.0750	0.0018	0.0427	0.0242
Denmark	0.0025	0.0006	0.0037	0.0012	0.0033	0.0006	0.0058	0.0009	9.0849	0.0000	0.0021	0.0461	0.0153
Estonia	0.0017	0.0004	0.0042	0.0007	0.0048	0.0007	0.0025	0.0009	16.6181	0.0000	0.0029	0.0541	0.0120
Finland	0.0015	0.0006	0.0022	0.0010	0.0015	0.0005	0.0020	0.0007	0.7696	0.5111	0.0003	0.0175	0.0170
France	0.0034	0.0006	0.0067	0.0012	0.0048	0.0005	0.0073	0.0012	15.3778	0.0000	0.0022	0.0473	0.0124
Germany	0.0009	0.0004	0.0007	0.0004	0.0014	0.0004	0.0014	0.0007	2.3698	0.0692	0.0008	0.0277	0.0173
Ireland	0.0037	0.0014	0.0046	0.0010	0.0045	0.0011	0.0121	0.0048	2.6859	0.0454	0.0051	0.0714	0.0419
Italy	0.0050	0.0011	0.0078	0.0016	0.0072	0.0010	0.0071	0.0014	4.3502	0.0047	0.0011	0.0335	0.0148
Japan	0.0005	0.0003	0.0013	0.0006	0.0019	0.0006	0.0043	0.0017	9.9193	0.0000	0.0040	0.0632	0.0271
Korea	0.0006	0.0002	0.0010	0.0005	0.0004	0.0001	0.0024	0.0010	6.5616	0.0002	0.0030	0.0547	0.0215
Netherlands	0.0006	0.0002	0.0016	0.0007	0.0032	0.0007	0.0035	0.0011	19.4202	0.0000	0.0028	0.0531	0.0119
Norway	0.0028	0.0007	0.0018	0.0008	0.0012	0.0005	0.0016	0.0007	4.3457	0.0047	0.0014	0.0371	0.0204
Poland	0.0055	0.0008	0.0064	0.0015	0.0072	0.0016	0.0067	0.0026	1.2886	0.2770	0.0004	0.0192	0.0177
Russian Federation	0.0111	0.0023	0.0185	0.0063	0.0090	0.0023	0.0089	0.0026	3.0520	0.0278	0.0127	0.1127	0.0543
Slovak Republic	0.0017	0.0006	0.0025	0.0008	0.0014	0.0003	0.0031	0.0009	4.6099	0.0033	0.0019	0.0440	0.0181
Spain	0.0061	0.0013	0.0083	0.0013	0.0078	0.0008	0.0104	0.0022	3.0232	0.0289	0.0009	0.0300	0.0181
Sweden	0.0043	0.0016	0.0021	0.0008	0.0026	0.0006	0.0005	0.0003	20.6899	0.0000	0.0036	0.0599	0.0212
United Kingdom	0.0039	0.0013	0.0026	0.0008	0.0039	0.0008	0.0029	0.0009	1.9654	0.1175	0.0006	0.0245	0.0185
United States	0.0024	0.0006	0.0024	0.0007	0.0043	0.0010	0.0040	0.0014	3.9860	0.0078	0.0013	0.0362	0.0145

Note: Degrees of freedom of D1: df1=3, df2=1000; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A9. Age differences in test-taking disengagement in Problem solving by country

Country	$M_{<25}$	$SE_{<25}$	M_{25-34}	SE_{25-34}	M_{35-54}	SE_{35-54}	$M_{>54}$	$SE_{>54}$	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0302	0.0039	0.0299	0.0045	0.0463	0.0036	0.0584	0.0079	20.5645	0.0000	0.0113	0.1061	0.0233
Belgium	0.0417	0.0045	0.0378	0.0048	0.0713	0.0047	0.1065	0.0088	78.0885	0.0000	0.0314	0.1773	0.0210
Canada	0.0627	0.0053	0.0849	0.0070	0.0890	0.0047	0.1150	0.0070	33.0733	0.0000	0.0078	0.0882	0.0154
Czech Republic	0.0385	0.0075	0.0393	0.0053	0.0790	0.0088	0.0935	0.0114	24.6913	0.0000	0.0355	0.1885	0.0323
Denmark	0.0472	0.0058	0.0426	0.0068	0.0644	0.0043	0.0850	0.0044	31.3321	0.0000	0.0118	0.1084	0.0218
Estonia	0.0457	0.0048	0.0520	0.0056	0.0884	0.0045	0.1274	0.0083	96.3077	0.0000	0.0345	0.1857	0.0196
Finland	0.0292	0.0044	0.0286	0.0046	0.0495	0.0044	0.0847	0.0081	42.7771	0.0000	0.0274	0.1656	0.0253
Germany	0.0325	0.0046	0.0434	0.0067	0.0539	0.0047	0.0847	0.0115	22.0207	0.0000	0.0181	0.1344	0.0286
Ireland	0.0629	0.0061	0.0620	0.0067	0.1010	0.0071	0.1206	0.0121	37.3428	0.0000	0.0221	0.1487	0.0227
Japan	0.0333	0.0081	0.0198	0.0042	0.0482	0.0047	0.0832	0.0133	26.8154	0.0000	0.0263	0.1621	0.0339
Korea	0.0228	0.0035	0.0338	0.0036	0.0719	0.0044	0.1055	0.0118	109.9216	0.0000	0.0545	0.2335	0.0209
Netherlands	0.0317	0.0039	0.0304	0.0047	0.0451	0.0040	0.0829	0.0073	44.7760	0.0000	0.0274	0.1654	0.0233
Norway	0.0361	0.0050	0.0187	0.0036	0.0420	0.0037	0.0588	0.0073	39.0263	0.0000	0.0138	0.1174	0.0196
Poland	0.0717	0.0047	0.1036	0.0074	0.1071	0.0092	0.1606	0.0204	42.2021	0.0000	0.0199	0.1412	0.0259
Russian Federation	0.1623	0.0282	0.1607	0.0275	0.1709	0.0171	0.1798	0.0233	0.2940	0.8298	0.0009	0.0306	0.0395
Slovak Republic	0.0543	0.0050	0.0632	0.0072	0.0874	0.0064	0.0851	0.0108	21.6031	0.0000	0.0121	0.1098	0.0258
Sweden	0.0322	0.0050	0.0288	0.0048	0.0532	0.0053	0.0716	0.0073	32.9986	0.0000	0.0181	0.1345	0.0220
United Kingdom	0.0510	0.0053	0.0756	0.0082	0.0867	0.0051	0.1020	0.0087	37.3548	0.0000	0.0137	0.1169	0.0200
United States	0.0719	0.0095	0.0899	0.0115	0.0746	0.0069	0.0737	0.0113	3.1740	0.0235	0.0021	0.0453	0.0251

Note: Degrees of freedom of D1: df1=3, df2=1000; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Educational attainment

Table A10. Educational attainment differences in test-taking disengagement in Literacy by country

Country	$M_{<highschool}$	$SE_{<highschool}$	$M_{highschool}$	$SE_{highschool}$	$M_{>highschool}$	$SE_{>highschool}$	*D1*	*p*	η^2	η	SE_η
Austria	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Belgium	0.0311	0.0043	0.0231	0.0027	0.0148	0.0025	14.0232	0.0000	0.0048	0.0695	0.0196
Canada	0.0502	0.0052	0.0459	0.0038	0.0336	0.0025	14.5909	0.0000	0.0028	0.0529	0.0139
Czech Republic	0.0126	0.0043	0.0129	0.0023	0.0095	0.0035	0.6938	0.4999	0.0008	0.0287	0.0341
Denmark	0.0314	0.0043	0.0255	0.0029	0.0187	0.0023	7.3741	0.0007	0.0027	0.0519	0.0188
Estonia	0.0237	0.0034	0.0230	0.0021	0.0195	0.0018	2.2967	0.1011	0.0006	0.0247	0.0164
Finland	0.0186	0.0041	0.0158	0.0024	0.0105	0.0018	5.1903	0.0057	0.0021	0.0462	0.0203
France	0.0336	0.0027	0.0288	0.0019	0.0174	0.0015	32.1417	0.0000	0.0059	0.0768	0.0129
Germany	0.0239	0.0051	0.0180	0.0020	0.0102	0.0015	16.2832	0.0000	0.0067	0.0819	0.0226
Ireland	0.0399	0.0058	0.0230	0.0044	0.0253	0.0030	5.9951	0.0026	0.0038	0.0620	0.0268
Italy	0.0499	0.0052	0.0386	0.0031	0.0354	0.0045	4.6771	0.0095	0.0034	0.0582	0.0263
Japan	0.0200	0.0085	0.0252	0.0032	0.0119	0.0018	16.8542	0.0000	0.0056	0.0748	0.0167
Korea	0.0106	0.0024	0.0111	0.0014	0.0065	0.0008	9.3280	0.0001	0.0026	0.0506	0.0155
Netherlands	0.0137	0.0019	0.0084	0.0015	0.0072	0.0014	7.6932	0.0005	0.0033	0.0574	0.0213
Norway	0.0202	0.0037	0.0166	0.0019	0.0091	0.0014	18.3441	0.0000	0.0056	0.0750	0.0181
Poland	0.0347	0.0064	0.0483	0.0048	0.0309	0.0044	7.9031	0.0004	0.0045	0.0673	0.0248
Russian Federation	0.0898	0.0309	0.0876	0.0224	0.0800	0.0129	0.9776	0.3766	0.0005	0.0230	0.0293
Slovak Republic	0.0147	0.0035	0.0221	0.0021	0.0147	0.0026	7.1552	0.0008	0.0030	0.0546	0.0206
Spain	0.0417	0.0033	0.0394	0.0040	0.0302	0.0026	8.9387	0.0001	0.0028	0.0533	0.0176
Sweden	0.0344	0.0090	0.0190	0.0025	0.0110	0.0019	12.8924	0.0000	0.0112	0.1060	0.0296
United Kingdom	0.0548	0.0167	0.0200	0.0023	0.0170	0.0025	6.3119	0.0019	0.0141	0.1186	0.0515
United States	0.0215	0.0049	0.0204	0.0035	0.0146	0.0023	3.7880	0.0230	0.0018	0.0427	0.0213

Note: Degrees of freedom of D1: df1=2, df2=1000; NA=Statistic could not be derived; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A11. Educational attainment differences in test-taking disengagement in Numeracy by country

Country	$M_{<highschool}$	$SE_{<highschool}$	$M_{highschool}$	$SE_{highschool}$	$M_{>highschool}$	$SE_{>highschool}$	*D1*	*p*	η^2	η	SE_η
Austria	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Belgium	0.0029	0.0008	0.0033	0.0007	0.0039	0.0008	0.7883	0.4549	0.0002	0.0155	0.0169
Canada	0.0097	0.0018	0.0096	0.0012	0.0068	0.0006	5.2175	0.0056	0.0012	0.0340	0.0145
Czech Republic	0.0038	0.0017	0.0028	0.0005	0.0011	0.0005	9.1892	0.0001	0.0024	0.0490	0.0166
Denmark	0.0025	0.0006	0.0033	0.0006	0.0047	0.0008	5.0051	0.0069	0.0012	0.0351	0.0148
Estonia	0.0035	0.0009	0.0036	0.0006	0.0038	0.0006	0.1039	0.9013	0.0000	0.0052	0.0166
Finland	0.0008	0.0003	0.0027	0.0007	0.0012	0.0003	7.1078	0.0009	0.0021	0.0462	0.0165
France	0.0066	0.0010	0.0061	0.0006	0.0034	0.0005	13.1880	0.0000	0.0020	0.0452	0.0122
Germany	0.0015	0.0006	0.0012	0.0003	0.0010	0.0003	0.4833	0.6169	0.0002	0.0143	0.0208
Ireland	0.0040	0.0011	0.0067	0.0016	0.0047	0.0009	2.0598	0.1280	0.0010	0.0309	0.0218
Italy	0.0063	0.0011	0.0072	0.0013	0.0071	0.0014	0.2183	0.8039	0.0002	0.0126	0.0253
Japan	0.0006	0.0004	0.0031	0.0008	0.0011	0.0003	5.6957	0.0035	0.0035	0.0595	0.0211
Korea	0.0014	0.0005	0.0007	0.0002	0.0006	0.0002	1.7901	0.1675	0.0007	0.0256	0.0245
Netherlands	0.0022	0.0006	0.0019	0.0005	0.0033	0.0010	1.4586	0.2330	0.0008	0.0285	0.0222
Norway	0.0023	0.0011	0.0022	0.0006	0.0010	0.0004	3.6163	0.0272	0.0015	0.0389	0.0223
Poland	0.0056	0.0017	0.0075	0.0011	0.0055	0.0013	1.7734	0.1703	0.0008	0.0279	0.0227
Russian Federation	0.0128	0.0036	0.0172	0.0047	0.0108	0.0026	5.6265	0.0037	0.0045	0.0672	0.0232
Slovak Republic	0.0010	0.0005	0.0021	0.0004	0.0020	0.0006	3.0684	0.0469	0.0005	0.0229	0.0128
Spain	0.0077	0.0009	0.0086	0.0015	0.0077	0.0009	0.3265	0.7215	0.0001	0.0110	0.0173
Sweden	0.0022	0.0007	0.0031	0.0008	0.0017	0.0005	1.6984	0.1835	0.0011	0.0325	0.0231
United Kingdom	0.0053	0.0015	0.0029	0.0005	0.0038	0.0010	2.8966	0.0557	0.0008	0.0277	0.0160
United States	0.0057	0.0022	0.0043	0.0011	0.0022	0.0005	4.1852	0.0155	0.0025	0.0503	0.0217

Note: Degrees of freedom of D1: df1=2, df2=1000; NA=Statistic could not be derived; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A12. Educational attainment differences in test-taking disengagement in Problem solving by country

Country	$M_{<highschool}$	$SE_{<highschool}$	$M_{highschool}$	$SE_{highschool}$	$M_{>highschool}$	$SE_{>highschool}$	*D1*	*p*	η^2	η	SE_η
Austria	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Belgium	0.0926	0.0082	0.0779	0.0047	0.0418	0.0046	39.4366	0.0000	0.0239	0.1545	0.0258
Canada	0.1287	0.0085	0.0857	0.0057	0.0694	0.0031	45.7627	0.0000	0.0130	0.1138	0.0164
Czech Republic	0.0623	0.0084	0.0744	0.0069	0.0324	0.0052	28.6182	0.0000	0.0220	0.1482	0.0242
Denmark	0.0922	0.0094	0.0692	0.0044	0.0374	0.0031	63.8620	0.0000	0.0235	0.1533	0.0191
Estonia	0.0850	0.0090	0.0863	0.0053	0.0595	0.0033	22.6371	0.0000	0.0084	0.0915	0.0185
Finland	0.0677	0.0072	0.0536	0.0041	0.0316	0.0028	35.2511	0.0000	0.0143	0.1195	0.0194
Germany	0.0682	0.0074	0.0652	0.0054	0.0305	0.0038	46.2444	0.0000	0.0223	0.1494	0.0215
Ireland	0.1330	0.0109	0.0859	0.0081	0.0684	0.0050	32.3918	0.0000	0.0252	0.1586	0.0276
Japan	0.0518	0.0106	0.0518	0.0064	0.0365	0.0042	6.1228	0.0023	0.0041	0.0640	0.0258
Korea	0.0609	0.0105	0.0620	0.0037	0.0426	0.0030	14.9109	0.0000	0.0078	0.0883	0.0220
Netherlands	0.0797	0.0056	0.0438	0.0033	0.0267	0.0040	62.9762	0.0000	0.0330	0.1818	0.0234
Norway	0.0786	0.0105	0.0455	0.0041	0.0215	0.0023	53.7306	0.0000	0.0361	0.1899	0.0254
Poland	0.0889	0.0111	0.1241	0.0069	0.0817	0.0061	20.7136	0.0000	0.0145	0.1205	0.0269
Russian Federation	0.1944	0.0275	0.1709	0.0334	0.1630	0.0178	1.7574	0.1730	0.0018	0.0421	0.0387
Slovak Republic	0.0613	0.0077	0.0830	0.0045	0.0470	0.0058	22.6572	0.0000	0.0145	0.1204	0.0247
Sweden	0.0910	0.0093	0.0496	0.0040	0.0250	0.0028	62.8134	0.0000	0.0418	0.2045	0.0284
United Kingdom	0.1452	0.0146	0.0877	0.0046	0.0584	0.0055	32.7789	0.0000	0.0239	0.1547	0.0272
United States	0.1045	0.0109	0.0935	0.0095	0.0577	0.0056	28.0112	0.0000	0.0171	0.1308	0.0242

Note: Degrees of freedom of D1: df1=2, df2=1000; NA=Statistic could not be derived; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Language

Table A13. Language differences in test-taking disengagement in Literacy by country

Country	$M_{not\ same}$	$SE_{not\ same}$	M_{same}	SE_{same}	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0281	0.0073	0.0095	0.0011	6.1783	0.0131	0.0107	0.1036	0.0389
Belgium	0.0501	0.0113	0.0190	0.0018	7.1665	0.0075	0.0071	0.0843	0.0308
Canada	0.0467	0.0046	0.0404	0.0023	1.5205	0.2178	0.0004	0.0191	0.0155
Czech Republic	0.0058	0.0030	0.0122	0.0017	3.0866	0.0792	0.0004	0.0195	0.0106
Denmark	0.0502	0.0054	0.0219	0.0019	28.7017	0.0000	0.0072	0.0851	0.0161
Estonia	0.0178	0.0063	0.0214	0.0013	0.2968	0.5860	0.0001	0.0084	0.0155
Finland	0.0434	0.0191	0.0129	0.0013	2.5563	0.1102	0.0046	0.0675	0.0413
France	0.0448	0.0061	0.0247	0.0012	10.9312	0.0010	0.0036	0.0601	0.0182
Germany	0.0351	0.0075	0.0141	0.0012	7.4987	0.0063	0.0101	0.1005	0.0347
Ireland	0.0314	0.0114	0.0267	0.0022	0.1766	0.6744	0.0002	0.0138	0.0221
Italy	0.0748	0.0116	0.0399	0.0028	8.3772	0.0039	0.0072	0.0847	0.0297
Japan	0.0313	0.0338	0.0177	0.0017	0.1591	0.6900	0.0001	0.0089	0.0068
Korea	0.0152	0.0089	0.0090	0.0007	0.4812	0.4880	0.0001	0.0109	0.0145
Netherlands	0.0240	0.0054	0.0087	0.0009	7.6875	0.0057	0.0072	0.0850	0.0282
Norway	0.0296	0.0055	0.0118	0.0012	9.4927	0.0021	0.0089	0.0943	0.0315
Poland	0.0158	0.0077	0.0391	0.0030	7.7337	0.0055	0.0004	0.0205	0.0075
Russian Federation	NA	NA	NA	NA	NA	NA	NA	NA	NA
Slovak Republic	0.0173	0.0061	0.0194	0.0017	0.1018	0.7498	0.0000	0.0065	0.0130
Spain	0.0420	0.0083	0.0361	0.0020	0.4656	0.4952	0.0002	0.0143	0.0210
Sweden	0.0345	0.0117	0.0160	0.0018	2.4360	0.1189	0.0070	0.0834	0.0500
United Kingdom	0.0467	0.0115	0.0190	0.0021	5.6499	0.0176	0.0110	0.1050	0.0427
United States	0.0358	0.0098	0.0155	0.0021	4.0923	0.0433	0.0080	0.0893	0.0431

Note: Degrees of freedom of D1: df1=1, df2=1000; NA=Statistic could not be derived; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A14. Language differences in test-taking disengagement in Numeracy by country

Country	$M_{not\ same}$	$SE_{not\ same}$	M_{same}	SE_{same}	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0021	0.0014	0.0015	0.0004	0.1580	0.6911	0.0001	0.0110	0.0159
Belgium	0.0089	0.0029	0.0032	0.0005	3.9924	0.0460	0.0029	0.0539	0.0266
Canada	0.0082	0.0010	0.0082	0.0006	0.0014	0.9702	0.0000	0.0005	0.0101
Czech Republic	0.0107	0.0045	0.0022	0.0004	3.5578	0.0596	0.0060	0.0775	0.0394
Denmark	0.0067	0.0014	0.0034	0.0005	4.9740	0.0260	0.0013	0.0363	0.0172
Estonia	0.0011	0.0011	0.0037	0.0003	6.0479	0.0141	0.0004	0.0191	0.0078
Finland	0.0000	0.0000	0.0018	0.0003	NA	NA	NA	NA	NA
France	0.0086	0.0023	0.0051	0.0004	2.3415	0.1263	0.0009	0.0294	0.0181
Germany	0.0023	0.0009	0.0010	0.0002	2.0549	0.1520	0.0011	0.0339	0.0232
Ireland	0.0050	0.0024	0.0050	0.0008	0.0001	0.9916	0.0000	0.0002	0.0211
Italy	0.0117	0.0036	0.0065	0.0009	1.7469	0.1866	0.0018	0.0421	0.0325
Japan	0.0129	0.0140	0.0019	0.0004	0.6197	0.4313	0.0010	0.0318	0.0291
Korea	0.0049	0.0039	0.0007	0.0002	1.1791	0.2778	0.0016	0.0397	0.0378
Netherlands	0.0074	0.0032	0.0020	0.0003	2.8507	0.0916	0.0046	0.0679	0.0344
Norway	0.0032	0.0014	0.0015	0.0004	1.3887	0.2389	0.0010	0.0320	0.0272
Poland	0.0006	0.0006	0.0065	0.0008	38.2003	0.0000	0.0004	0.0189	0.0026
Russian Federation	NA	NA	NA	NA	NA	NA	NA	NA	NA
Slovak Republic	0.0030	0.0015	0.0019	0.0003	0.5068	0.4767	0.0003	0.0174	0.0245
Spain	0.0091	0.0027	0.0078	0.0007	0.2046	0.6511	0.0001	0.0083	0.0184
Sweden	0.0051	0.0016	0.0020	0.0004	3.1633	0.0756	0.0029	0.0535	0.0304
United Kingdom	0.0045	0.0015	0.0034	0.0005	0.4845	0.4866	0.0002	0.0132	0.0195
United States	0.0070	0.0024	0.0030	0.0005	2.5497	0.1106	0.0025	0.0505	0.0309

Note: Degrees of freedom of D1: df1=1, df2=1000; NA=Statistic could not be derived; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.

Table A15. Language differences in test-taking disengagement in Problem solving by country

Country	$M_{not\ same}$	$SE_{not\ same}$	M_{same}	SE_{same}	*D1*	*p*	η^2	η	SE_{η}
Austria	0.0701	0.0102	0.0371	0.0021	9.3650	0.0023	0.0123	0.1108	0.0361
Belgium	0.1151	0.0185	0.0616	0.0030	7.8531	0.0052	0.0085	0.0921	0.0324
Canada	0.1106	0.0066	0.0814	0.0029	15.0906	0.0001	0.0045	0.0669	0.0171
Czech Republic	0.0855	0.0305	0.0625	0.0049	0.5571	0.4556	0.0005	0.0229	0.0306
Denmark	0.0922	0.0106	0.0587	0.0026	10.2530	0.0014	0.0043	0.0659	0.0206
Estonia	0.0988	0.0168	0.0725	0.0029	2.3556	0.1251	0.0010	0.0309	0.0205
Finland	0.0336	0.0086	0.0462	0.0028	2.0680	0.1507	0.0003	0.0178	0.0122
Germany	0.0884	0.0134	0.0477	0.0033	9.5402	0.0021	0.0115	0.1071	0.0345
Ireland	0.0651	0.0103	0.0848	0.0046	3.6618	0.0560	0.0013	0.0359	0.0186
Japan	0.0000	NA	0.0437	0.0037	NA	NA	0.0001	0.0104	NA
Korea	0.1988	0.0732	0.0518	0.0020	4.0790	0.0437	0.0146	0.1206	0.0563
Netherlands	0.0875	0.0146	0.0440	0.0024	8.4624	0.0037	0.0106	0.1032	0.0355
Norway	0.0520	0.0085	0.0373	0.0025	2.6261	0.1054	0.0018	0.0428	0.0259
Poland	0.1056	0.0316	0.1002	0.0042	0.0301	0.8624	0.0000	0.0032	0.0147
Russian Federation	NA	NA	NA	NA	NA	NA	NA	NA	NA
Slovak Republic	0.0955	0.0146	0.0706	0.0035	2.6052	0.1068	0.0016	0.0397	0.0245
Sweden	0.0783	0.0124	0.0419	0.0026	7.6212	0.0059	0.0124	0.1115	0.0384
United Kingdom	0.0854	0.0179	0.0790	0.0037	0.1181	0.7311	0.0002	0.0126	0.0246
United States	0.1112	0.0245	0.0734	0.0052	2.5585	0.1100	0.0059	0.0771	0.0466

Note: Degrees of freedom of D1: df1=1, df2=1000; NA=Statistic could not be derived; disengagement was defined via the P+>0% method.

Source: Survey of Adult Skills (PIAAC) (2012), www.oecd.org/site/piaac/publicdataandanalysis.htm.