7

Using Log-File Data to Understand What Drives Performance in PISA (Case Study)

In computer-based tests, machines keep track (in log files) of – and, if so instructed, could analyse – all the steps and actions students take in finding a solution to a given problem. This chapter uses three tasks from the PISA 2012 computer-based reading assessment to illustrate how process data recorded by the assessment can enhance educators' ability to monitor students' test-taking behaviour and measure their skills.

Information and communication technology (ICT) tools have the potential to improve education and teaching in several ways. In the domain of assessment, digital tools can improve precision and efficiency of measurements, and expand the kinds of knowledge and skills that can be assessed (e.g. problem solving [OECD, 2014]). Perhaps more important, ICT tools make it easier for students to identify their learning needs as they participate in the assessment. Indeed, computers' interactive nature, rapid feedback loops and powerful analytical possibilities can be harnessed in the interest of universal learning principles. When digital tools support students' engagement with challenging material, thus extending learning time and practice, or help students to assume control over the learning situation, by individualising the pace with which new material is introduced or by providing immediate feedback, students probably learn more.

Bunderson, Inouye and Olsen (1988) were among the first to describe the potential of "intelligent measurement". In their vision, computers could provide the kind of advice that, in the past, only experts could give to learners, and only if they closely monitored learners' progress (an extremely time-consuming activity). In computer-administered tasks, machines keep track of all the steps and actions taken towards a solution (in log files). If correctly instructed, computers could also analyse those actions along with students' performance on the tasks. Thus, computers could eventually produce not only static test scores, but also an interpretation of scores (a profile) and personalised feedback for learners and their instructors.

Yet more than 25 years later, this vision of intelligent measurement is still far from being implemented on a large scale. One reason for the slow progress is the scarcity of studies investigating the use of process data (log-file data) for analysing students' learning. As Csapó et al. (2012) note, it is very difficult to "make sense of the hundreds of pieces of information students may produce when engaging in a complex assessment" (p. 216).

What the data tell us

- In computer-based tests, log files record information about the timing and type of actions students perform while trying to solve tasks. Analyses of log files allow for investigating how fluently students read, how persistent they are in trying to solve challenging problems and, more generally, analysing differences in how students handle tasks.
- Students who need a long time to read and understand a simple test question are likely to lack fluency in reading, as well as other basic reading skills. Data culled from one task in the digital reading assessment show that students in Brazil, Chile, Colombia, Hungary and the United Arab Emirates are significantly more likely to read slowly than students in other countries. In contrast, in Japan, Korea, and other countries with small shares of low performers in reading, few students read slowly.
- The largest proportions of students who could solve a complex digital reading task in less than the seven minutes usually required to succeed were observed in Australia, Canada, France and the United States. But when also considering students who persisted beyond seven minutes to solve the problem correctly, other countries and economies – notably Hong Kong-China, Ireland, Japan, Macao-China, Shanghai-China and Singapore – performed on par with or sometimes better than the former group of countries.

This chapter uses three assessment items, or tasks, to illustrate how process data recorded by computer-based assessments enhance the ability to measure students' behaviour and skills. All three items analysed here belong to a same test unit (*SERAING*), which means that they share a common text as stimulus material.

The case study is organised as follows. First, the general features of the unit *SERAING* are introduced. In addition to providing the context for the later analyses, this unit illustrates how students' skills in digital reading were assessed in PISA 2012. Next, students' reading fluency, their persistence, and their navigation behaviour are analysed, drawing on comparisons of students across countries and across performance levels.

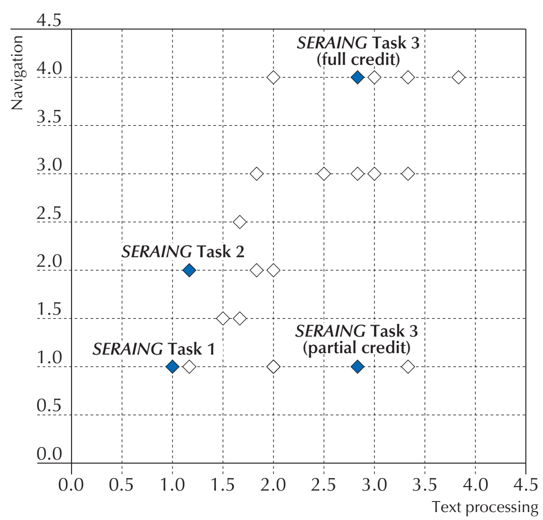
All findings in this section are based on a limited number of tasks, sometimes on a single task. For this reason, they should be interpreted with caution. Further research is necessary to generalise and corroborate results.

DESCRIPTION OF THE UNIT *SERAING*

The three items analysed in greater detail here were chosen to illustrate the range of navigation demands in the PISA assessment. Figure 7.1 shows that the tasks in the unit *SERAING* span the full range of navigation difficulties.

■ Figure 7.1 ■

How text processing and navigation demands vary across tasks in the unit *SERAING*



Notes: The horizontal axis shows experts' ratings of text-processing demands; the vertical axis shows experts' ratings of navigation demands (both ratings are expressed on a 1-4 scale, with 4 corresponding to the most difficult).
Each diamond represents one or more tasks in the PISA 2012 assessment of digital reading.
Source: OECD (2011), *PISA 2009 Results: Students on Line: Digital Technologies and Performance (Volume VI)*, p. 43, http://dx.doi.org/10.1787/9789264112995-en.
StatLink http://dx.doi.org/10.1787/888933253361

■ Figure 7.2 ■

Screenshots of webpages used in the unit *SERAING*

Source: Australian Council for Educational Research (2015), "PISA examples of computer-based items: Digital Reading 2012: *SERAING*", http://cbasq.acer.edu.au/index.php?cmd=toEra2012 (accessed 30 April 2015).

The items are briefly described below and are available for viewing on the website of the Australian Council for Educational Research (http://cbasq.acer.edu.au/index.php?cmd=toEra2012 [accessed 30 April 2015]). The extent and nature of text-processing skills and navigation required in each of the tasks can be best appreciated by trying to solve the items.

The unit *SERAING* is built around the fictional website of the Belgian city of Seraing (Figure 7.2). The unit comprises three tasks. Box 7.1 summarises the main characteristics of each task. The first two tasks measure students' proficiency in "accessing and retrieving" information. In an online environment, this involves searching for information by interacting with the available navigation tools, using knowledge of the typical structures encountered on line and feedback received along the way as a guide. To limit the search difficulty, both tasks provide directions to guide and orient students ("Look at the Seraing home page", "Find the page for…"). The third task is classified as "complex" because several cognitive processes are involved at the same time to solve this task. As in previous tasks, students need to search and locate information, and are provided with explicit directions to assist them. They also need to "integrate and interpret" the information found, contrasting two concert descriptions, and to reflect on these descriptions in light of their personal preferences, in order to justify a choice of one concert over the other. All aspects of the reading framework ("access and retrieve", "integrate and interpret", "reflect and evaluate": see OECD, 2013a) are important, making this a "complex" item. Answers given to this last item were scored by trained coders.

Box 7.1. General description of tasks in the unit *SERAING*

TASK 1

Question stem: Look at the Seraing home page. What are the dates of the Heritage Days?

Answer format: Simple multiple choice (4 options)

Framework aspect: Access and retrieve

Difficulty on the digital reading scale: 248 points (below Level 1b)

Number of pages available for navigation: 21

Minimum number of navigation steps required for locating the necessary information: 0 (but scrolling is required)

TASK 2

Question stem: Find the page for the Seraing Community Cultural Centre. Which film will be shown in the first week of November?

Answer format: Simple multiple choice (4 options)

Framework aspect: Access and retrieve

Difficulty on the digital reading scale: 382 points (Level 1a)

Number of pages available for navigation: 21

Minimum number of navigation steps required for locating the necessary information: 3

…

TASK 3

Question stem: Open the e-mail inbox and read John's e-mail message to Max. Click on "Reply" and write a reply to John from Max. In the reply, recommend which concert to buy tickets for (5 December or 12 December). Explain why this concert would be more enjoyable, referring to the concert descriptions.

Click on "Send your message" to send the reply.

Answer format: Constructed response, expert coded

Framework aspect: Complex

Difficulty on the digital reading scale: 570 points (Level 4) for full credit, 547 points (Level 3) for partial credit. Partial credit is given to students who could indicate their preference and give an explanation, but their explanation is not related to concert descriptions. No credit is given to students who could indicate the preference, but did not give any explanation.

Number of pages available for navigation: 25

Minimum number of navigation steps required for locating the necessary information: 11

Over 38 000 students from 32 countries participating in the assessment of digital reading were given these tasks. Process data are available for 38 506 students for the first task, 38 370 for the second, and 37 474 for the third task. The unit appeared at the end of one of the two test clusters, which means that half of the students were administered this unit before reaching the middle of the 40-minute test session, and half of the students at the very end.

On average across OECD countries, 38% of students received full credit for all three tasks in this unit, 43% solved two out of three tasks correctly, 16% solved only one task correctly, and 4% of students did not solve any task correctly, among students with available process data and excluding students who did not reach this unit in the test (Table 7.6). In general, the proportion of correct answers across countries/economies is in line with the ranking of countries and economies by their mean performance on the digital reading scale.[1]

Throughout the case study, the behaviour of strong and weak students will be contrasted by comparing students who received full credit for all three tasks and students who solved at most one of the three tasks correctly. As defined here, the groups of strong and weak students account for 38% and 20% of the student population, respectively.

HOW FLUENTLY CAN STUDENTS READ?

Good readers are able to process the words and sentences they read fast and accurately (Catts et al., 2002), as if they were listening to a natural voice rather than reading a text. In simple tasks where the stimulus material contains a short explicit direction, *initial reaction time* can be used to measure reading fluency. Initial reaction time refers to the amount of time, measured in seconds, from the moment the student sees the item to the first action a student takes.[2] This measure can be extracted from process data recorded automatically by the test-administration platform.

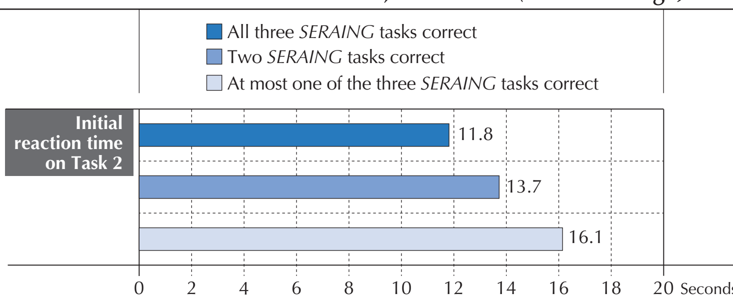
Although it was not designed for this purpose, Task 2 in the unit *SERAING* offers a good setting to measure and compare reading fluency across students. When students reach Task 2 in the unit, they have already had the opportunity to familiarise themselves with the homepage of Seraing – the stimulus material – in Task 1. The only material that is new to them is the question in the bottom part of the screen, which states: "Find the page for the Seraing Community Cultural Centre. Which film will be shown in the first week of November?" The phrase "Community Cultural Centre" prominently appears at the centre of the homepage. Initial reaction time on this task typically corresponds to the time it takes students to click on this link; it is therefore most likely related to the time it takes to read and understand the question, and only to a limited extent to other processes (such as locating the information, developing a plan, etc.).

Figure 7.3 shows how the initial reaction time varies across students of different proficiency in (digital) reading. The weakest students – those who are able to solve at most one of the three *SERAING* tasks correctly – have the longest reaction times, on average, indicating that many of these students may lack fluency in decoding words and sentences.

■ Figure 7.3 ■

Initial reaction time on Task 2 in the unit *SERAING*, by performance categories

Median time between the start of the task and the first mouse click, in seconds (OECD average)



Note: Because time indicators typically have a skewed distribution and some very large outliers, the median time (50th percentile) is shown in this figure.
Source: OECD, PISA 2012 Database, Table 7.1.
StatLink http://dx.doi.org/10.1787/888933253377

On average across OECD countries, 10% of students have reaction times shorter than 6 seconds, while another 10% of students take 28 seconds or more to react (Table 7.1). Further analyses show that reaction time is negatively related to success in this simple task. Some 83% of the fastest students (those who reacted in under 6 seconds) went on to answer the question correctly, while only 46% of the slowest students (those who reacted in over 28 seconds) answered the question correctly.[3] The latter group is thus likely to have poor fluency in reading and difficulties related to basic reading skills. But how are these students distributed across countries?

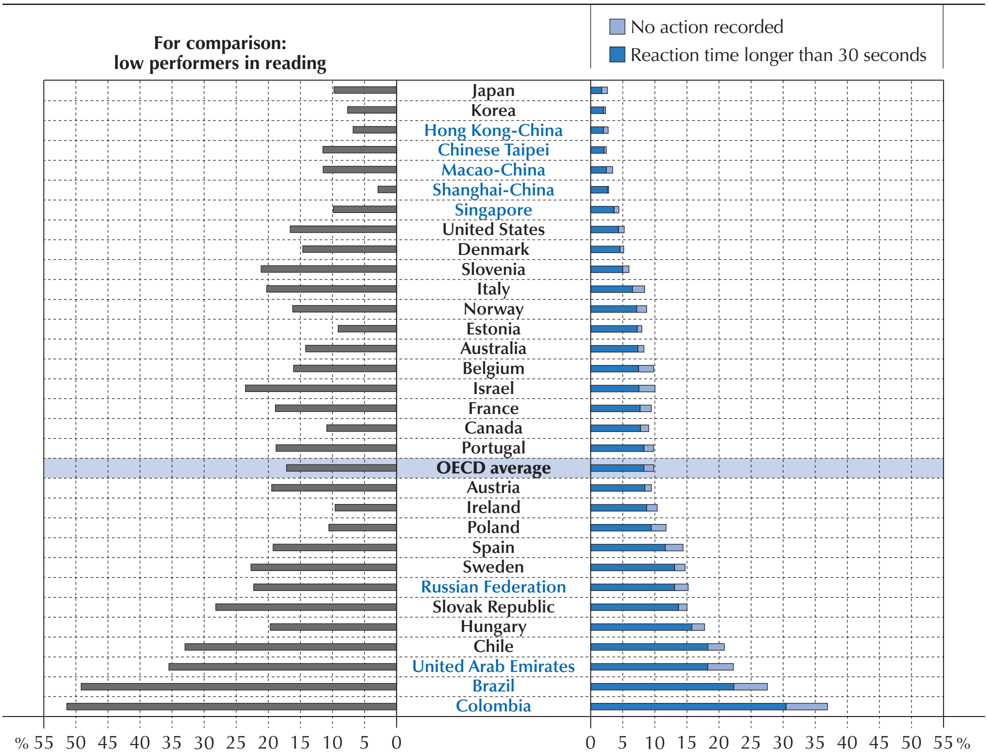
In some countries, such as Brazil, Chile, Colombia, Hungary and the United Arab Emirates, students are significantly more likely to read slowly than in Japan, Korea, and other East Asian countries and economies. The fact that this variation across countries is strongly related to the variation

in the proportion of low performers in (print) reading may indicate that one reason behind students' difficulties in reading is insufficient basic skills, such as the automatic decoding of words (Figure 7.4).

■ Figure 7.4 ■

Relationship between long reaction time on Task 2 in the unit *SERAING* and low performance in reading

Across countries and economies



Countries and economies are ranked in ascending order of the percentage of students whose reaction time is longer than 30 seconds.

Source: OECD, PISA 2012 Database, Table 7.1.

StatLink http://dx.doi.org/10.1787/888933253389

By identifying a lack of reading fluency as a possible reason behind the large proportion of low-performing students, the analysis of log files in digital reading can also indicate possible ways to improve the reading skills of low performers and adjust instruction to better serve their needs. The high correlation (0.90), across countries and economies, between the percentage of students performing below Level 2 in print reading, and the percentage of students whose initial reaction time exceeds 30 seconds shows that reaction time is a good predictor of mastery of basic reading skills (Table 7.1). Further research would be needed, however, to ensure that measures of reading fluency based on reaction time are comparable across languages.[4]

HOW DO STUDENTS ALLOCATE EFFORT AND TIME TO TASKS?

Timing measures may not only function as indicators of how fluently students process certain cognitive tasks; they may also indicate the degree of effort and motivation. For this reason, they are often difficult to interpret.

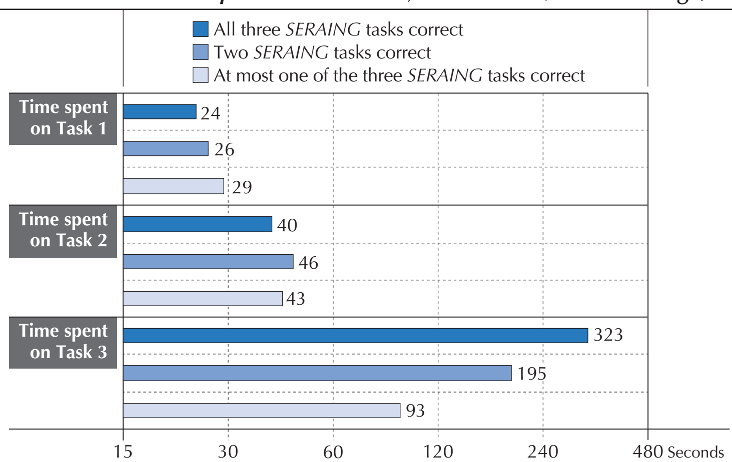
Time on task is calculated as the total time spent on a task from start to finish. How is time on task related to performance? If longer times indicate greater care in completing the task, then better-performing students may spend more time on a task. At the same time, more proficient students work more fluently and thus faster on tasks, and this could explain a negative association between success and time on task (Goldhammer et al., 2014).

Which effect dominates – greater care or greater fluency – is a matter of empirical investigation, and may vary depending on task demands, as well as on external time pressures.[5] Tasks that lend themselves to automatic processing will give rise to a negative association between time on task and performance. Tasks that instead require students to regulate their cognitive processing more, and that cannot be solved without the active engagement of students, will give rise to positive associations.

■ Figure 7.5 ■

Time on tasks in the unit *SERAING*, by performance categories

Median time spent on each task, in seconds (OECD average)



Note: The horizontal axis is in logarithmic scale: each tick corresponds to a doubling in the value.
Source: OECD, PISA 2012 Database, Table 7.2.
StatLink http://dx.doi.org/10.1787/888933253392

The third and most difficult task in the unit *SERAING* illustrates the complex relationship between time on task and performance (Figure 7.5). In Tasks 1 and 2 of this unit, the strongest students, i.e. those who are able to solve all three *SERAING* tasks correctly, work faster than weaker students, on average. This is consistent with the observation that these tasks are relatively straightforward. Even Task 2 lends itself to automatic processing. Although it requires several steps to find the solution, students who follow the expected path do not run into unexpected

impasses, and the type of navigation they are asked to perform – narrowing down the search while moving towards a solution – is relatively linear, and corresponds to a familiar structure that is often encountered on line.

In contrast, several features make Task 3 more demanding. To start, students need to work on two websites at the same time (a webmail site and the Seraing website). They must use several navigation tools (tabs and links), and need to navigate back and forth across pages. In addition, students need to use their short-term memory to contrast two descriptions and then encapsulate one of the differences found in these descriptions in an argument stating their preference. Not surprisingly, the strongest students are those who take the longest time to reach a solution. Indeed, this task requires high levels of metacognitive regulation in order to select a strategy, apply adequate cognitive processes at each step, and sustain effort throughout the task.

Success rates on Task 3 vary significantly across countries. For the most part, these differences are in line with overall performance differences across countries/economies. The main exception is Korea, where fewer students solved the task correctly than across OECD countries, on average (Tables 3.1 and 7.3).

Analyses of time spent on this task reveal that students' willingness and ability to sustain and regulate effort in demanding tasks may play a major role in between-country differences in performance. On average across OECD countries, 41% of students received full credit for Task 3; 32% of students were able to solve the task correctly in less than seven minutes; and a further 9% took seven minutes or longer to solve this task correctly (Table 7.3). The latter group of students shows remarkable persistence in a task that may be at the upper limit of their ability.[6]

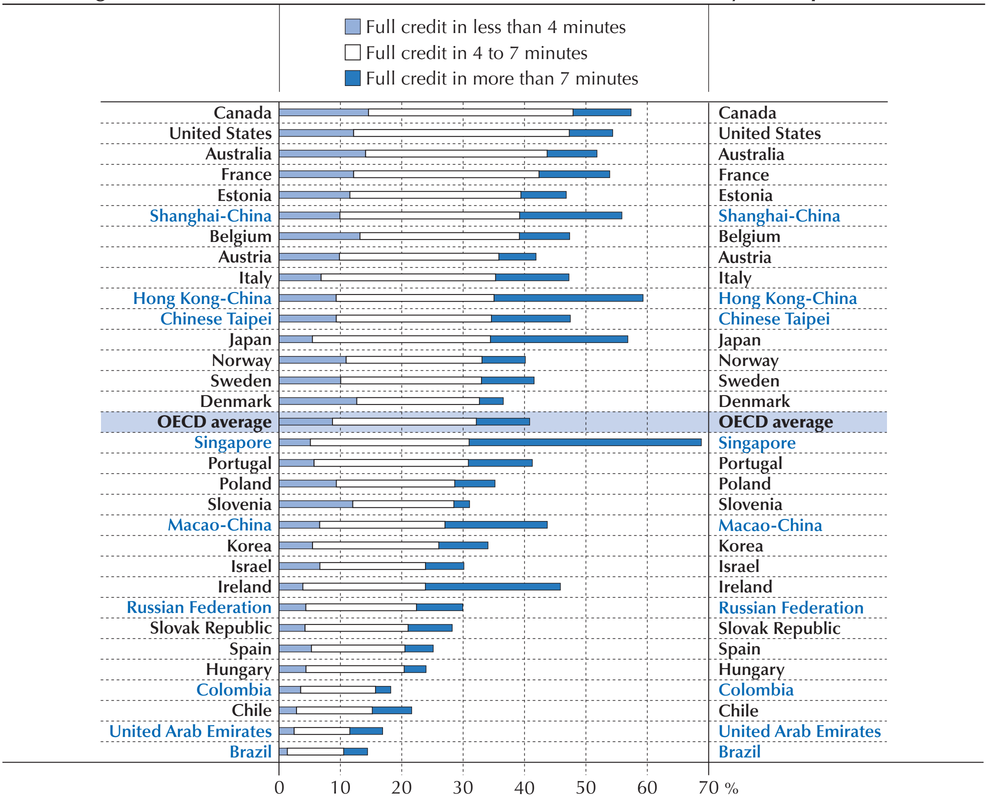
Figure 7.6 shows that in Australia, Canada, France and the United States, more than four in ten students were able to solve Task 3 correctly in less than seven minutes, a higher proportion than on average across OECD countries. However, after including slower but persistent students who were able to solve Task 3 accurately, but needed more than seven minutes to do so, other countries performed on par with or sometimes better than this group of countries. In Singapore, 38% of students succeeded on this task after spending more than seven minutes working on it – making it the most successful country in completing this task. In Hong Kong-China, Ireland, Japan, Macao-China and Shanghai-China, more than one in six students belong to this group of relatively slow, but highly persistent students.[7]

These first results based on log files show how measures of students' use of time during the PISA test can be related to cognitive and non-cognitive aspects of students' performance at both ends of the performance distribution. The findings relate the variation in students' reaction time and time on task to their ease with and motivation to complete different tasks. Further work is required to investigate the robustness of findings based on case studies. Such work may also extend the analyses of timing data recorded in log files to other aspects, such as task dependencies and order effects. Does the time required to solve the previous task influence students' willingness and ability to solve the next task? Do students allocate time and effort differently at the beginning and end of the test? Do students strategically allocate time to tasks, skipping questions when they recognise them as "too hard to crack"?

■ Figure 7.6 ■

Students' perseverance and success

Percentage of students who succeed on Task 3 in the unit SERAING*, by time spent on the task*



Countries and economies are ranked in descending order of the percentage of students who solved Task 3 in less than seven minutes and obtained full credit for their solution.

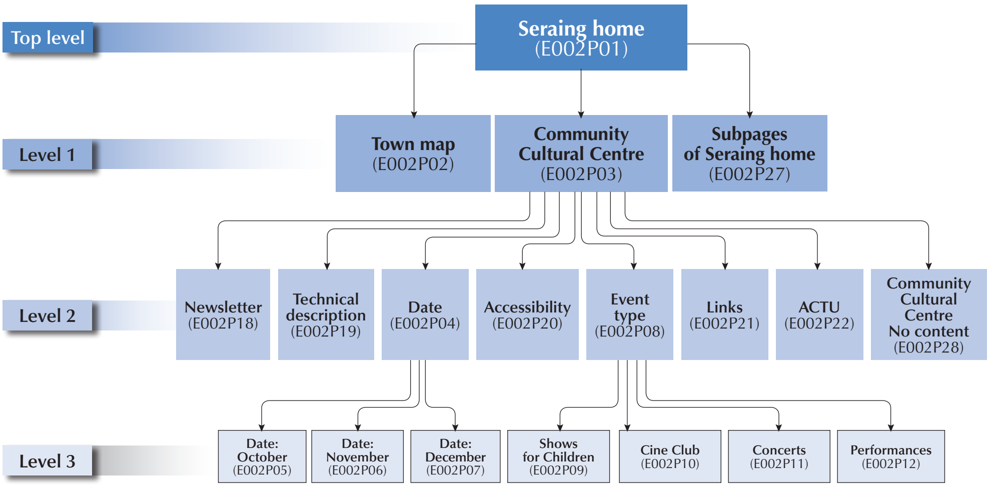
Source: OECD, PISA 2012 Database, Table 7.3.

StatLink http://dx.doi.org/10.1787/888933253404

HOW DO STUDENTS NAVIGATE A SIMPLE WEBSITE?

The unit *SERAING* is built around the fictional website of a city (Seraing). The structure of Seraing's website corresponds to a hierarchical hypertext, a relatively common structure encountered on line (Figure 7.7). Hierarchical hypertexts have a branching structure that resembles a tree and have been found to facilitate information retrieval, compared to linear or networked structures (Mohageg, 1992). The typical navigation sequences on these websites have one beginning (the "home page") but many possible endings, depending on decisions made by the reader. Readers who search for information in hierarchical sites typically move to the next level at each step, thus narrowing down their search. In addition to movements across levels along "branches", in Seraing's website students could also move across pages belonging to the same hierarchical level by means of a navigation menu, a feature that is often present in real websites as well.

■ Figure 7.7 ■
Hierarchical structure of Seraing's fictitious website



Note: "Top Level", "Level 1", "Level 2" and "Level 3" indicate hierarchical levels: Level 3 pages can only be accessed from the Top Level by going through a Level 1 and a Level 2 page first.
Source: Australian Council for Educational Research (2015), "PISA examples of computer-based items: Digital Reading 2012: *SERAING*", http://cbasq.acer.edu.au/index.php?cmd=toEra2012 (accessed 30 April 2015).

Figure 7.8 shows which pages are available for navigation and were coded as relevant in each task. Pages are identified by a title that describes the content, and by a unique code (not visible to the student) used for the purpose of analysis. The first and second tasks in the unit *SERAING* start on the home page, which is the website's top-level page. Several thematic portals can be found at the first level, and detailed information is provided at the second and third levels. The third task starts on a mailing website; students can reach Seraing's website by clicking on a link embedded in an e-mail. As shown in the figure, there is only one relevant page in Task 1; there are six relevant pages in Task 2, which form two efficient navigation paths; and there are 12 task-relevant pages in Task 3, seven of which appear on Seraing's website, and five appear on the mailing website.

Several links on the website were intentionally "broken", and students who clicked on them landed on pages with "no content available", and received immediate feedback inviting them to go back to the previous page. There are two such pages, one at Level 1 of the hierarchy, one at Level 2. These pages may not be perceived as the same page by students but are counted as such throughout this report.

Successful and unsuccessful navigation in Task 2 in the unit *SERAING*

Task 2 in the unit *SERAING* corresponds to a relatively common, if simple, online search task. The task requires students to identify the movie to be shown in the Community Cultural Centre during the first week of November, information they can find on the third level of the hierarchy.

■ Figure 7.8 ■

List of page codes and relevant pages in the unit *SERAING*

PAGE TYPE

- S Starting page
- R Relevant page
- N Non-relevant page
- (blank) Not available for this task

PAGE CODE	PAGE NAME	Task 1 (CR002Q01)	Task 2 (CR002Q03)	Task 3 (CR002Q05)
E002P01	Seraing: home	S	S	N
E002P02	Town map	N	N	N
E002P03	Community Cultural Centre	N	R	R
E002P04	Community Cultural Centre: Date	N	R	R
E002P05	Community Cultural Centre: Date – October	N	N	N
E002P06	Community Cultural Centre: Date – November	N	R	N
E002P07	Community Cultural Centre: Date – December	N	N	R
E002P08	Community Cultural Centre: Event type	N	R	R
E002P09	Community Cultural Centre: Event type – Shows for children	N	N	N
E002P10	Community Cultural Centre: Event type – Cine club	N	R	
E002P11	Community Cultural Centre: Event type – Concerts	N	N	R
E002P12	Community Cultural Centre: Event type – Performances	N	N	N
E002P13	Email reading			R
E002P14	Community Cultural Centre: Event type – Madredeus 1	N	N	R
E002P15	Community Cultural Centre: Concerts – Madredeus 2	N	N	R
E002P16	Send email			R
E002P17	Email frame			R
E002P18	Community Cultural Centre: Newsletter	N	N	N
E002P19	Community Cultural Centre: Technical description	N	N	N
E002P20	Community Cultural Centre: Accessibility	N	N	N
E002P21	Community Cultural Centre: Links	N	N	N
E002P22	Community Cultural Centre: ACTU	N	N	N
E002P27	Subpages of Seraing home page without content	N	N	N
E002P28	Subpages of Community Cultural Centre without content	N	N	N
E002P29	Confirmation for sending the reply email			R
E002P30	Email home			S
	Number of relevant pages in each task	**1**	**6**	**12**

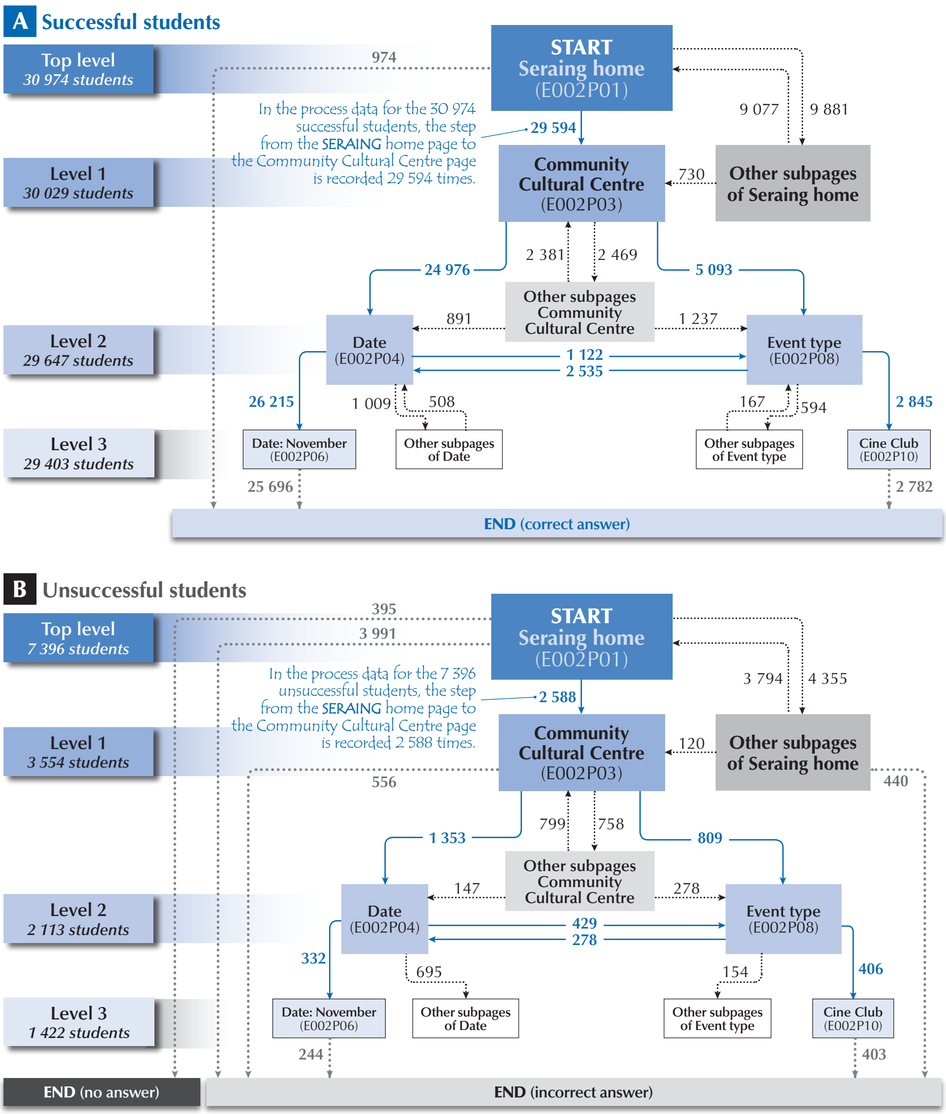
Note: The PISA digital reading unit *SERAING* can be seen at the website http://cbasq.acer.edu.au/index.php?cmd=toEra2012 (accessed 30 April 2015).

On average, four out of five of students across OECD countries solved this task correctly (Table 7.5). The most efficient navigation sequence involves visiting three pages beyond the starting page (three steps) (Figure 7.9). Students may fail this unit either because they make wrong decisions at one or more points in their navigation, or because they fail to continue navigating for as long as required.

Figure 7.9 contrasts the behaviour of successful and unsuccessful students. The figure shows the main navigation paths in this task (not all paths are shown). Arrows indicate possible steps along these paths: blue arrows indicate task-relevant steps (from relevant to relevant pages), whereas black dotted arrows indicate other kinds of steps (missteps, corrections, and task-irrelevant steps). For each step shown in these figures, a number counts the frequency with which it is recorded in students' log files. Students could exit the navigation to give an answer or move to the next task at any point; such decisions are shown by grey arrows.

■ Figure 7.9 ■

Navigation behaviour in Task 2 in the unit *SERAING*



Notes: The figure shows the main pages available for viewing and the main navigation paths in this task (not all paths are shown). Blue and black dotted arrows indicate possible steps between pages: blue arrows indicate task-relevant steps (from relevant to relevant pages), whereas black dotted arrows indicate other kinds of steps (missteps, corrections, and task-irrelevant steps). Grey dotted arrows indicate the last page visited before ending navigation. For each step shown in these figures, the number of cases recorded in log files is indicated next to the arrows (numbers are unweighted).
Pages are placed based on their position at different levels in the hierarchical hypertext. Relevant pages are highlighted in blue. Non-relevant pages having the same position in the hierarchy were grouped to simplify exposition.
Source: OECD, PISA 2012 Database, Table 7.4.

Figure 7.9 clearly shows the two efficient navigation sequences. The first goes through the "Date" page, the second through the "Event type" page. Panel A in this figure shows that successful students were more likely to visit the "Date" page than the "Event type" page, after reaching the portal of the Community Cultural Centre. This may be related to the fact that the former link appears above the latter in the navigation menu, and is therefore more prominent. It is also likely that students who solved this task perceived a closer match between the question and the stimulus text for the former link ("November" – "Date") than for the latter ("film" – "Event type"). Indeed, some students who initially followed the "Event type" path reverted to the "Date" path in their next step, when continuing on the "Event type" branch would have required them to match the question stem ("film") with "Cine-club" among a list of 12 possibilities (including some potentially overlapping categories, such as "Various" or "Shows for children").

The top-left part of Figure 7.9 also shows that some 1 571 students (30 974 minus 29 403), representing 5% of all students who solved this item correctly, did not, in fact, reach the third level of Seraing's website during their navigation, but nevertheless gave the correct answer. Of the 30 974 students who gave the correct answer, only 29 403 reached Level 3. If their log files are complete, this means that they gave a guess answer. Given that this is a multiple-choice item where students had to select among four options, it can be estimated that for each correct guess, there should be about three times as many incorrect guesses. Indeed, further analyses show that 5 251 students, representing 71% of the unsuccessful students, similarly tried to give an answer, despite the fact that they had not reached the third level of Seraing's website.

Why did students fail to answer this question correctly? Panel B in Figure 7.9 shows that a majority of the unsuccessful students (3 991 students) did not perform any action other than attempt a guess answer. It also shows that, among the remaining students, those who did navigate the website had relatively unfocused browsing behaviour, on average. Visits to non-relevant pages are much more likely to be observed among students who failed the task.

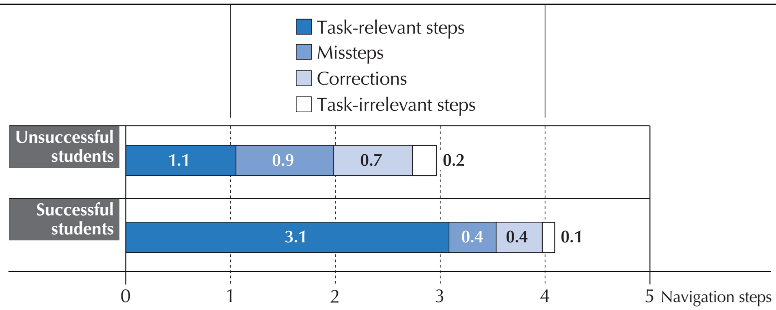
Differences between successful and unsuccessful students in the quantity and quality of navigation are confirmed by the indicators of navigation used in Chapter 4. Figure 7.10 shows that successful students had longer navigation sequences (four navigation steps) than unsuccessful students (three navigation steps), on average. It also shows that unsuccessful students had a larger number of non-task-relevant steps (missteps, corrections, task-irrelevant steps) than successful students, whose typical navigation sequence included the three required task-relevant steps. In short, the navigation of successful students is characterised by longer sequences and a more careful selection of pages to visit.

Overall, it is possible to distinguish four reasons for failing to solve this task, three of which are related to navigation. First, students may not have navigated the website and simply remained on the Seraing home page. Second, students may have ended their navigation at some point along the task-relevant path before reaching the page containing the crucial information to solve the task. Third, students may have deviated from the task-relevant navigation path, ending their navigation on a non-relevant page – perhaps they were lost after a misstep. Finally, some students may have completed their navigation as expected, by visiting either the "November" or the "Cine-club" page, but nevertheless failed to give the correct answer.

■ Figure 7.10 ■

Quality and quantity of navigation steps in Task 2 in the unit *SERAING*, by performance on the task

OECD average values

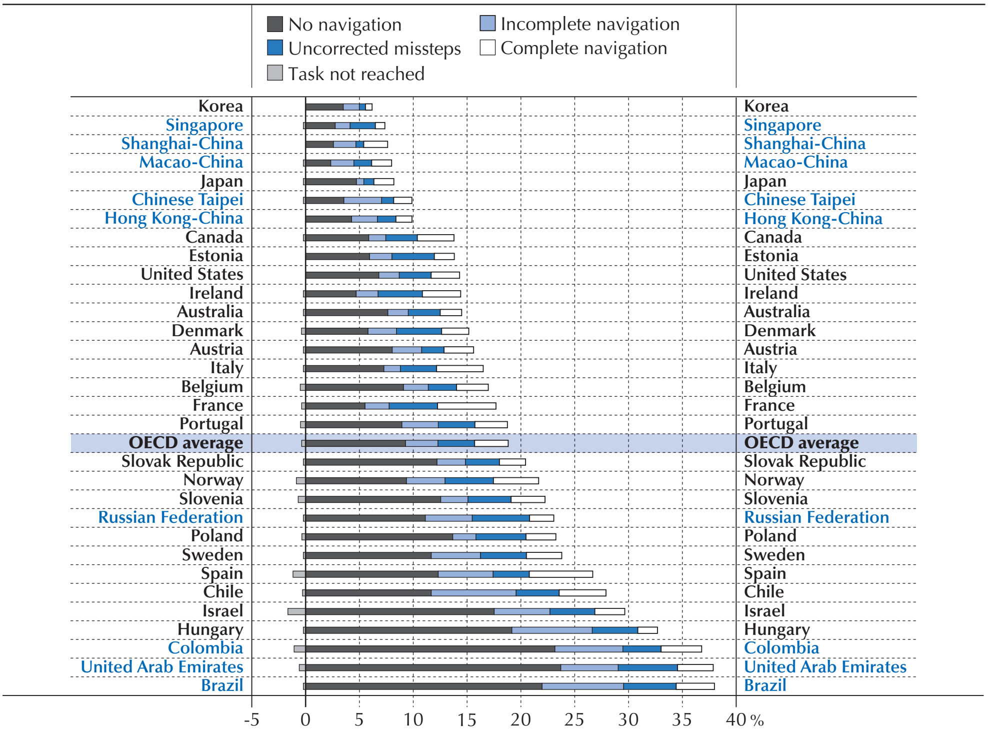


Source: OECD, PISA 2012 Database, Table 7.5.

StatLink http://dx.doi.org/10.1787/888933253410

■ Figure 7.11 ■

Navigation behaviour of students unsuccessful on Task 2 in the unit *SERAING*



Countries and economies are ranked in ascending order of the share of students with no credit for Task 2 in the unit SERAING.

Source: OECD, PISA 2012 Database, Table 7.5.

StatLink http://dx.doi.org/10.1787/888933253420

Across all countries/economies, most students who were unsuccessful in this item have navigation-related mistakes recorded in their log files. In detail, on average across OECD countries, 9% of students did not attempt any navigation step in this item. However, in Korea, Macao-China, Shanghai-China, Singapore and Chinese Taipei, less than 4% of students did not try to navigate in this item. By contrast, an average of 3% of students completed the navigation but still failed to select the correct answer; in Spain and France, this was the case for more than 5% of students (Figure 7.11).

IMPLICATIONS OF THE CASE STUDY FOR COMPUTER-BASED ASSESSMENTS

The analyses of reaction time and of time on task, as well as the detailed description of students' navigation behaviour, based on data recorded in log files for a single unit in digital reading, illustrate three advantages of computer-based assessments.

First, the detailed information on the interactions of unsuccessful students with assessment tasks may be used to improve the ability to measure proficiency at lower ends of the performance distribution. In particular, the scoring rubrics could be expanded to give partial credit for certain elementary processes observed in isolation, in addition to giving credit for the joint product of these processes (task success). For example, partial credit could be awarded to students who understand the simple instruction "Find the page for the Community Cultural Centre" and click on the corresponding link.

Second, log files often reveal to a greater extent than students' answers alone what the most frequent mistakes are, and allow for investigating the reasons behind them. This information, in turn, can be used to identify learners' profiles and improve teaching practices. In mathematics, there is a long tradition of identifying common mistakes on assessment tasks in order to diagnose misconceptions and weaknesses among students. Teachers use this to inform instruction and design learning experiences (OECD, 2005; Wiliam, 2010). Furthermore, existing studies at the national level have similarly analysed the traces left by students on PISA paper booklets (OECD, 2013b; DEPP, 2007). However, such analyses have been limited by the fact that few students (about 10% in the case of France) actually leave traces of their reasoning on paper booklets. The data captured by a computer system can expand the possibilities for such analyses – at lower cost.

In addition, the above analysis shows that several invalid test-taking behaviours, such as guessing, can be detected in on-screen tests. This, in turn, may lead to significant gains in the precision of performance measures derived from test answers. In the PISA computer-based assessment of problem solving, for instance, several multiple-choice items were designed to require certain interactions to arrive at the correct solution and were scored as correct only if these interactions had been recorded in log files (OECD, 2014).

Notes

1. The correlation of the sum of scores on *SERAING* tasks with the latent performance measure (PV1CREA) is 0.68.

2. "Action" refers here to mouse clicks only (on a link, a tab, on an answer option, or elsewhere on the page). Mouse scrolls or keyboard strokes are not recorded in the data used for this analysis.

3. Percentages refer to the pooled sample with students from all countries and economies.

4. There are only few studies in which comparable, timed fluency tasks were administered across languages and orthographies. Most of these are not based on representative samples, focus on younger students, and are confined to European languages (e.g. Frith, Wimmer and Landerl, 1998; Mann and Wimmer, 2002; Seymour et al., 2003). It is not clear how well their results can be generalised to the population level and to later stages in the development of reading proficiency, and how other languages would compare (also see Abadzi, 2011).

5. The PISA computer-based test in 2012 was a timed test: students had 40 minutes to complete all questions in their forms. Questions had been used in a Field Trial, and the length of test forms was determined, after the Field Trial, to ensure that approximately 80% of all students (across countries) would complete the test without running into the time limit.

6. Only full-credit answers are considered here, for two reasons. First, there were few students who received partial credit. Second, students could have received partial credit without navigating the Seraing website at all, at least in theory. Only students receiving full credit demonstrate the kind of skills that justify the classification of this item as "complex" within the reading framework.

7. Between-country differences do not change much by whether students see this task towards the middle of the test, or at the end of it. See Table 7.3 for full results.

Chapter 7 tables are available on line at http://dx.doi.org/10.1787/edu-data-en.

Note regarding Israel

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

References

Abadzi, H. (2011), Reading fluency measurements in EFA FTI partner countries: Outcomes and improvement prospects, *Partnership for Education (GPE) Working Paper Series on Learning,* No. 1, World Bank, http://documents.worldbank.org/curated/en/2011/09/18042914/reading-fluency-measurements-efa-fti-partner-countries-outcomes-improvement-prospects.

Australian Council for Educational Research (2015), "PISA examples of computer-based items: Digital Reading 2012: *SERAING*", http://cbasq.acer.edu.au/index.php?cmd=toEra2012 (accessed 30 April 2015).

Bunderson, C.V., D.K. Inouye and **J.B. Olsen** (1988), "The four generations of computerized educational measurement", *ETS Research Report Series,* Vol. 1988/1, pp. i-148.

Catts, H.W., M. Gillispie, L.B. Leonard, R.V. Kail and **C.A. Miller** (2002), "The role of speed of processing, rapid naming, and phonological awareness in reading achievement", *Journal of Learning Disabilities*, Vol. 35/6, pp. 510-525.

Csapó, B., J. Ainley, R.E. Bennett, T. Latour and **N. Law** (2012), "Technological issues for computer-based assessment", in P. Griffin, B. McGaw and E. Care (eds.), *Assessment and Teaching of 21st Century Skills*, Springer Netherlands, pp. 143-230.

DEPP (2007), "Résultats des français par champs mathématiques (Chapter 5)", in *L'évaluation internationale PISA 2003 : Compétences des élèves français en mathématiques, compréhension de l'écrit et sciences*, Les Dossiers, No. 180, Direction de l'évaluation, de la prospective et de la performance (DEPP), Ministère de l'éducation nationale, de l'enseignement supérieur et de la recherche, http://media.education.gouv.fr/file/83/1/4831.pdf.

Frith, U., H. Wimmer and **K. Landerl** (1998), "Differences in phonological recoding in German- and English-speaking children", *Scientific Studies of Reading,* Vol. 2/1, pp. 31-54.

Goldhammer, F., J. Naumann, A. Stelter, K. Tóth, H. Rölke and **E. Klieme** (2014), "The time on task effect in reading and problem solving is moderated by task difficulty and skill: Insights from a computer-based large-scale assessment", *Journal of Educational Psychology,* Vol. 106/3, pp. 608-626.

Mann, V. and **H. Wimmer** (2002), "Phoneme awareness and pathways into literacy: A comparison of German and American children", *Reading and Writing,* Vol. 15/7-8, pp. 653-682.

Mohageg, M.F. (1992), "The influence of hypertext linking structures on the efficiency of information retrieval", *Human Factors,* Vol. 34/3, pp. 351-367.

OECD (2014), *PISA 2012 Results: Creative Problem Solving (Volume V): Students' Skills in Tackling Real-Life Problems*, PISA, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264208070-en.

OECD (2013a), *PISA 2012 Assessment and Analytical Framework: Mathematics, Reading, Science, Problem Solving and Financial Literacy*, PISA, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264190511-en.

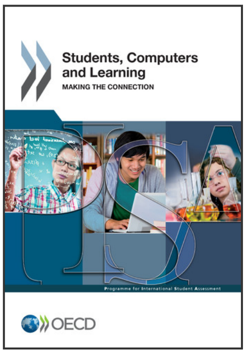
OECD (2013b), "Strengths and Weaknesses of American Students in Mathematics", in OECD, *Lessons from PISA 2012 for the United States*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264207585-5-en.

OECD (2011), *PISA 2009 Results: Students On Line: Digital Technologies and Performance (Volume VI)*, PISA, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264112995-en.

OECD (2005), *Formative Assessment: Improving Learning in Secondary Classrooms*, OECD Publishing, Paris, http://dx.doi.org/10.1787/9789264007413-en.

Seymour, P.H.K., M. Aro and **J. Erskine** and collaboration with **COST Action A8 network** (2003), "Foundation literacy acquisition in European orthographies", *British Journal of Psychology*, Vol. 94, pp. 143-174.

Wiliam, D. (2010), "The role of formative assessment in effective learning environments", in *The Nature of Learning: Using Research to Inspire Practice*, OECD Publishing, Paris, http://dx.doi.org/10.1787/ 9789264086487-8-en.



From:
Students, Computers and Learning
Making the Connection

Access the complete publication at:
https://doi.org/10.1787/9789264239555-en

Please cite this chapter as:

OECD (2015), “Using Log-File Data to Understand What Drives Performance in PISA (Case Study)”, in *Students, Computers and Learning: Making the Connection*, OECD Publishing, Paris.

DOI: https://doi.org/10.1787/9789264239555-10-en

OECD