Hahnel *et al. Large-scale Assessments in Education* (2023) 11:16
https://doi.org/10.1186/s40536-023-00169-5

Large-scale Assessments in Education

RESEARCH

Open Access

Rule-based process indicators of information processing explain performance differences in PIAAC web search tasks

Carolin Hahnel[1,2*], Ulf Kroehne[1] and Frank Goldhammer[1,2]

*Correspondence:
Carolin Hahnel
c.hahnel@dipf.de
[1]TBA – Centre for Technology-Based Assessment, DIPF | Leibniz Institute for Research and Information in Education, Rostocker Straße 6, 60323 Frankfurt, Germany
[2]Centre for International Student Assessment (ZIB), Frankfurt, Germany

Abstract

Background A priori assumptions about specific behavior in test items can be used to process log data in a rule-based fashion to identify the behavior of interest. In this study, we demonstrate such a top-down approach and created a process indicator to represent what type of information processing (flimsy, breadth-first, satisficing, sampling, laborious) adults exhibit when searching online for information. We examined how often the predefined patterns occurred for a particular task, how consistently they occurred within individuals, and whether they explained task success beyond individual background variables (age, educational attainment, gender) and information processing skills (reading and evaluation skills).

Methods We analyzed the result and log file data of ten countries that participated in the *Programme for the International Assessment of Adult Competencies* (PIAAC). The information processing behaviors were derived for two items that simulated a web search environment. Their explanatory value for task success was investigated with generalized linear mixed models.

Results The results showed item-specific differences in how frequently specific information processing patterns occurred, with a tendency of individuals not to settle on a single behavior across items. The patterns explained task success beyond reading and evaluation skills, with differences across items as to which patterns were most effective for solving a task correctly. The patterns even partially explained age-related differences.

Conclusions Rule-based process indicators have their strengths and weaknesses. Although dependent on the clarity and precision of a predefined rule, they allow for a targeted examination of behaviors of interest and can potentially support educational intervention during a test session. Concerning adults' digital competencies, our study suggests that the effective use of online information is not inherently based on demographic factors but mediated by central skills of lifelong learning and information processing strategies.

Keywords Web search, Online information, PIAAC, Log data, Reading, Evaluation, Age, Educational attainment

Introduction

The analysis of log data from computer-based assessments holds great potential for measuring competencies. Among others, it provides information beyond item scores and allows inferences about the cognitive processes that took place during the completion of a task (e.g., Goldhammer et al., 2021; Keller et al., 2016). In the context of large-scale assessment, data-driven approaches of log data analysis helped to uncover specific solution strategies (e.g., Eichmann et al., 2020; Hahnel et al., 2022; He et al., 2019; He & von Davier, 2015; Tóth et al., 2017; Ulitzsch et al., 2021). However, such bottom-up methods require that the data contains more information that contributes to pattern detection than irrelevant information blurring it. Alternatively, process indicators and solution strategies may be identified a priori based on theoretical assumptions and previous research (e.g., Hahnel et al., 2019; Salles et al., 2020), allowing for a targeted examination of behaviors of interest without regard to their probability of occurrence or the application of computationally intensive methods.

In the present study, we demonstrate a top-down approach that leverages previous research findings on information processing behaviors in web search tasks. Such research identified behavioral patterns by exploring log data (Jenkins et al., 2003; Juvina & Oostendorp, 2006; Reader & Payne, 2007). Following their descriptions, we constructed coding rules to derive a process indicator indicating specific information processing patterns. We applied the rules to the log data of ten countries participating in the *Programme for the International Assessment of Adult Competencies* (PIAAC; OECD, 2013; also Maehler & Rammstedt, 2020) and examined the resulting process indicator for differences between two tasks and for the extent to which it explains item success beyond variables of individual background (age, educational attainment, gender) and information processing skills (reading and evaluation skills).

Web search as solving information problems

Nowadays, people often use Internet search engines to find specific information. After entering a search query, search engines provide a result list of sources with potentially relevant information (e.g., Bendersky et al., 2012). Using short descriptions of each result, web users generate predictive judgments about what they will encounter on the upcoming website when clicking a link (Hilligoss & Rieh, 2008). However, the information that has to be processed — on the result list and the following websites — is often diverse in quantity and quality (e.g., Flanagin & Metzger, 2007). Accordingly, web users need to filter, evaluate, and select information to fulfill their information needs while avoiding misinformation (e.g., Braasch & Graesser, 2020; Forzani et al., 2022; Hahnel et al., 2020; Leu et al., 2014; Molerov et al., 2020; Vakkari, 2016).

Finding information online that is suitable for one's information needs is considered an information problem (Brand-Gruwel et al., 2005). Based on their current state of knowledge, individuals acquire, evaluate, and integrate information in a goal-directed way to create a solution. Brand-Gruwel and colleagues (2009) summarize the required steps in their IPS-I (information problem solving with the Internet) model. Accordingly, web users (a) define the information problem, (b) select a search strategy and websites, (c) scan the encountered information to generate an initial evaluation of a website's usefulness and, eventually, (d) thoroughly process the website for comprehension while integrating the gathered information with their knowledge. Finally, they will (e) synthesize a solution for their information problem. During this process, web users will try to find the desired information as quickly as possible with minimal effort (e.g., Blackmon, 2012; Pirolli & Card, 1999). Theories of dual-processing suggest that individuals do not systematically process all information in situations where a variety of information is available but apply heuristic strategies to deal with the amount and complexity of information and reduce cognitive efforts (Evans, 2008; Wirth et al., 2007). However, oversimplified strategies can result in inadequate outcomes (e.g., relying on self-confirmation heuristics may lead to dismissing credible sources as less trustworthy; Metzger & Flanagin, 2013).

Web search tasks in PIAAC

Given the importance of the Internet in information societies, it is not surprising that the appropriate use of online information is considered key to participate in societal activities and, therefore, is subject to international large-scale assessments (ILSAs; see Singer et al., 2018). The PIAAC study 2011–2012 (OECD, 2013; see also Maehler & Rammstedt, 2020) was one of the first ILSAs that assessed not only the literacy and numeracy of adults but their skills in dealing with digital information as part of the domain *problem solving in technology-rich environments* (PSTRE; OECD, 2012). PSTRE problems portray non-routine tasks involving problems of organizing digital folder structures, e-mail requests, calendar appointments, and tasks that require the evaluation of information from search engines. More specifically, there are three PSTRE items that simulate a web search environment[1]: u06a, u06b, and u07.

Item u06a (*Sprained Ankle - Site Evaluation Table*) is an evaluation task that requested test-takers to rate each entry of a search result page. The item displayed one page with a list of five search results about a medical treatment and asked test-takers to assign an attribute (useful, not authoritative, biased, not current) to each link that describes it best. Although the item looked like a search result page, the presented links were just decorative and did not allow navigation. Nevertheless, test-takers had to interact with the page by scrolling since it was partly covered by the area where they entered their responses via radio button clicks. Scroll events, though, were not logged.

Item u06b (*Sprained Ankle - Reliable/Trustworthy Site*) is a selection task that provided a list of clickable search result links and requested test-takers to select the search result with the most reliable and trustworthy information. The item presented a small hierarchical hypertext, including a parent node with five links (the search result page) and a total of eight child nodes, where five child nodes (main pages) could be directly accessed from the search result page and three child nodes (subpages) from a main page. Scrolling was not possible. A response was given by selecting an entry from a dropdown combo box referring to one of the five search results. Note that item u06b is not an extension of item u06a. They share a common narrative (seeking information about a medical treatment) and display a prototypical list of search results, but they provide independent content. One item could be solved without knowledge of the other.

Like u06b, item u07 (*Digital Photography Book Purchase*) is a selection task. It provided a list of six clickable search results and a total of eleven websites (6 main pages and 5 subpages) without scrolling requirements. In contrast to u06b, adults were requested to find and order a product that matches a list of specific criteria (price, shipping dates,

[1] The items are not publicly released. However, an interactive example of a comparable design is available here: https://piaac-logdata.tba-hosting.de/public/problemsolving/JobSearchPart1/pages/jsp1-home.html.

book for beginners). A response was given by visiting one of the main pages and clicking the *Buy* button displayed on that page.

The two selection tasks, u06b and u07, were of particular interest for the present study. They provided individuals with navigation options and invited them to explore available alternatives autonomously. Therefore, the log data from these items could be used to derive a process indicator that reflects different information processing behaviors when evaluating online information. Possible person-task interactions included item-specific response events (u06b: *COMBOBOX* and *GLOBAL_VAR*, u07: *BUYBOOK*) and navigation-related events that were triggered by clicking on text links (event type *TEXTLINK*) and back, forward, home, and website-embedded buttons (corresponding event types *HISTORY_BACK*, *HISTORY_NEXT*, *HISTORY_ADD*, and *BUTTON*). Based on the structure of the PIAAC items, the raw log events (e.g., an event of the type *TEXTLINK* with the attribute *unit06bpage1* at time *t*) can be interpreted as meaningful process components, so-called low-level features (e.g., visiting the first website of item u06b at time *t*). A comprehensive description of the structure of the PIAAC log data and their transformation into low-level features can be found in Goldhammer et al. (2020, 2021).

Information processing behaviors in the PIAAC web search tasks

A recent study by Gao et al. (2022) identified several patterns of web navigation for the items u06b and u07 using the log data of the US-American and British/Northern-Irish PIAAC samples. The authors performed a series of latent class analyses on variables indicating how often and when a particular page was visited. They found five patterns and described them as *Flimsy*, *Breadth-first*, *Satisficing*, *Sampling*, and *Laborious* by matching representative navigation sequences of the identified classes with descriptions from previous studies (e.g., Jenkins et al., 2003; Juvina & Oostendorp, 2006; Reader & Payne, 2007). While Gao and colleagues detected all five patterns in the item u06b, only four (US sample) and three (UK sample) emerged in item u07.

With the present study, we pick up on the patterns recovered by Gao et al. (2022). We determined coding rules for them based on their description and supposed function (Table 1). Although these patterns primarily refer to web navigation, they nevertheless provide insight to adults' approach of information processing. For instance, the flimsy pattern summarizes a parsimonious navigation style in which most time is spent dwelling on the search results or a single website rather than exploring the hypertext structure (Juvina & Oostendorp, 2006). Such a pattern may indicate that web users prefer to finish a search task quickly and invest as little effort as possible (see Pirolli & Card, 1999). Given the constrained information space in the PIAAC web search tasks (i.e., total of 8 or 11 websites) and the fact that PIAAC items were not time restricted (see Sect. 2.1), we assigned the label *Flimsy* when individuals visited no more than a single source. Other patterns express an information processing behavior that is focused on browsing the surface of the available options (breadth-first), finishing the search task as soon as one's subjective information need has been satisfied (satisficing), inspecting all information options to identify the best option (sampling), or making an effort to build up a good representation of the information and hypertext structure (laborious). However, our derived coding rules of *Satisficing* and *Breadth-first*, as well as *Satisficing* and *Sampling*, could not be formulated to be mutually exclusive since the underlying descriptions do

Table 1 Overview of the investigated behaviors

Pattern		Description	Operationalization
(1)	Flimsy	Individuals primarily focus on the information provided on the search results page and visit only a few websites, if any (see Juvina & Oostendorp, 2006).	Number of main pages visited <= 1 (for item u07, this number is always 1 when a response is given)
(2)	Breadth-first	Individuals browse the surface of the available options. They visit only the main pages from the search results page and do not go deeper to explore their content (see Jenkins et al., 2003).	Number of main pages visited > 1[1], number of nested subpages visited = 0, and the conditions (5) Laborious[2] and (6) Multiple are not met
(3)	Satisficing	Individuals browse the available options until they find the website that matches their subjective level of aspiration for solving the search problem (see Reader & Payne, 2007).	Number of main pages visited > 1[1], number of page revisits = 0, and the condition (6) Multiple is not met. Item u06b: Last visited website was chosen as response, participants did not change their response, and they did not engage in navigation activities afterward
(4)	Sampling	Individuals go through the available options and explore and compare them until a website has been inspected to their satisfaction. They then decide on the subjectively best solution for the search problem (see Reader & Payne, 2007).	Item u06b: Number of main pages visited = 5, number of nested subpages visited > 0, and the conditions (5) Laborious[2] and (6) Multiple are not met Item u07: Number of main pages visited = 6, number of nested subpages visited > 0, and the conditions (5) Laborious[2] and (6) Multiple are not met
(5)	Laborious	Individuals visit a large number of websites and make extensive use of the infrastructure ensuring to create a good representation of the provided information and its structure (see Juvina & Oostendorp, 2006).	Number of website visits reaches or exceeds the sum of the number of unique (back and forth) navigation steps and the number of navigation steps required to revisit the main pages. Item u06b: Number of website visits ≥ 26; Item u07: Number of website visits ≥ 34
(6)	Multiple	As the above descriptions of the behaviors (2) and (3) as well as (3) and (4) are not mutually exclusive, individuals can demonstrate behavior that is in line with both descriptions.	Participants' behavior fits multiple patterns (e.g., satisficing and sampling: participants visit all main pages and at least some subpages before selecting the last visited page as their response) and the condition (5) Laborious is not met
(7)	Unassigned	None of the above described behaviors were observed.	Participants gave a response but their behavior could not be assigned to the categories above

Note. [1]Adults might follow a breadth-first or satisficing strategy while leaving little traces of navigation. This is the case when the first available option is already considered sufficient. Therefore, to distinguish flimsy behavior from breadth-first and satisficing behavior, we made it a condition that at least two main pages were visited to assign the category breadth-first or satisficing. [2]Eight cases across all three subsamples showed a laborious pattern without also demonstrating a breadth-first or sampling pattern.

not provide sufficient evidence for further distinction. In this case, we assigned the label *Multiple*. If none of these patterns applied, we labelled the behavior as *Unassigned*.

Three points should be noted. First, the exact coding rules slightly differ between items, as they are adjusted to the item-specific structures (number of websites) and response formats (drop-down combo box selection without explicit navigation requirements vs. clicking a button on a main page). Second, the *Laborious* category did not exclude breadth-first or sampling behavior. We still distinguished this category because it highlights the behavior of extensive hypertext use. Third, the behaviors should not be ranked in terms of effectiveness, as their effectiveness is likely to depend on multiple factors with respect to the task (e.g., personal value of a task; Rouet et al., 2017), the text (e.g., availability of relevance and credibility cues; Hahnel et al., 2020), and the reader (e.g., individual standards of coherence; van den Broek et al., 2011).

Interindividual differences in web search

Process indicators are useful when they explain variation in performance and shed light on the role of individual differences. Therefore, the second objective of our study was to investigate whether the derived process indicator explains item success beyond variables of individual background (age, educational attainment, gender) and information processing skills (reading and evaluation skills). The PIAAC study 2011–2012 provides a unique database for this objective, as participants come from different backgrounds and cover a wide range of ages (16–65 years).

Although there are many factors associated with adults' web search performance (e.g., prior knowledge: Lucassen et al., 2013; Monchaux et al., 2015; Rouet, 2003; web experiences: Hölscher & Strube, 2000; epistemic beliefs: Kammerer et al., 2013, 2015), age stands out as a background factor. Information processing skills tend to decline throughout life (e.g., Bryan et al., 1997; Warrington et al., 2018). This trend is reflected in the PIAAC data, which shows a curvilinear decline in problem solving performance with increasing age (OECD, 2013, p. 191). More specifically, empirical research indicates that, compared to young adults, older adults have difficulties in finding targets in navigation-oriented tasks (Etcheverry et al., 2012), frequently fail to recognize the primary interests of websites (Morrison, 2015), are more susceptible to online fraud (Grimes et al., 2010), and judge web information less accurately (Chevalier et al., 2015; Kubeck et al., 1999). However, such comparisons often neglect middle-aged groups and progression across adulthood, and do not consider the contribution of acquired information processing skills, such as reading skills.

Moreover, age-related effects do not necessarily occur from aging but can arise from being born in a particular cohort with its respective learning opportunities (e.g., higher education access in the 60s vs. the 90s; van der Kamp & Boudard, 2003). Such cohort effects, as well as adults without a university degree as a target population (Kammerer et al., 2015), are often neglected in small-scale studies on web search. Studies associating web search skills with different educational backgrounds are rare, although competencies and the acquisition of educational qualifications are highly related (Massing & Schneider, 2017; OECD, 2013). A higher level of education was positively associated with a higher awareness of security hazards (Grimes et al., 2010) and with self-reported knowledge of computer and Internet-related terms (e.g., "PDF", "newsgroup"; Hargittai & Hinnant, 2008). Other studies use the level of education as an indicator of expected skill level (e.g., information skills in Lucassen et al., 2013).

Considering gender as a third background variable, the evidence of differences in web search activities seems inconclusive. Although studies, such as ICILS 2018 (Fraillon et al., 2020), indicate an overall gender effect in computer and information literacy (CIL) across countries in favor of girls, gender effects in studies of digital reading were mixed, varying in favor of girls (Naumann & Sälzer, 2017) or boys (Rasmusson & Åberg-Bengtsson, 2015). Other studies found no gender effects in the credibility evaluation of online information (Hämäläinen et al., 2020). Still, the search efficiency seems to differ between genders (Zhou, 2014), with males being more active in formulating and updating search queries than females.

Regarding cognitive skills, the IPS-I model identifies reading, evaluation, and computer skills as conditional skills that create the basis for solving information problems (Brand-Gruwel et al., 2009). Since web users actively "create" their own text base by selecting information pieces, reading skills are indispensable to make sense of the information encountered and to form an integrated mental representation of what was read (for an overview, see Salmerón et al., 2018). Evaluation skills support this process, as they enable web users to make predictive judgments about the content and trustworthiness of available websites (Hahnel et al., 2020; Hilligoss & Rieh, 2008; Kiili et al., 2018). Accordingly, reading and evaluation skills can also affect navigation decisions (Goldman et al., 2012; Hahnel et al., 2018). As for computer skills, it is reasonable to assume that they are required to access and manage digital information within a hypertext. This assumption is reflected in the design of the PIAAC study, as participants were required to demonstrate a minimum level of computer proficiency before participating in the computer-based assessment (Kirsch et al., 2016).

The present study

The present study demonstrates a top-down approach to defining process indicators based on theoretical considerations. We leveraged evidence from previous research (e.g., Gao et al., 2022) to recover different types of information processing in adults searching the web. In this regard, our approach agrees with the classic paradigm of evidence-centered design (Mislevy et al., 2003; adapted to reason from log data, see Goldhammer et al., 2021): We apply evidence identification rules to summarize adults' behavior within a task to an observable variable (evidence model) representing a process-related construct of interest (student model). Based on the PIAAC log file data, we derived an indicator representing different behaviors expected to emerge when adults work on web search tasks (Table 1). We examined how often and consistently these patterns occurred across the two tasks and explored the success of adults who showed a specific pattern. Specifically, we investigated the unique contribution of these behaviors to predicting task success beyond variables of individual background (age, educational attainment, gender) and information processing skills (reading and evaluation skills). Since time allocation affects item success (e.g., Goldhammer et al., 2014) and substantial time differences can be expected between patterns (e.g., flimsy vs. laborious), we also included time on task as a control variable in our analysis of task success.

Materials and methods

Design of the PIAAC study

Mainly administered as a computer-based assessment (CBA), the PIAAC study 2011–2012 included several components (Kirsch et al., 2016). As shown in Fig. 1, background information about the participants was first collected through a computer-assisted personal interview. Supervised by an interviewer, the participants worked afterwards directly and autonomously on test items ("direct assessment") to assess their competencies in literacy, numeracy and, eventually, problem solving in technology-rich environments (PSTRE). The administration of the direct assessment was highly standardized. Interviewers were supposed to read standardized instructions and interact with participants as little as possible during the cognitive assessment. The PIAAC Consortium provided detailed training materials for interviewers and encouraged the participating countries to implement multiple measures of quality assurance and control (e.g., mentoring of data collection, feedback). Still, countries were allowed to adjust the measures to the needs of their interviewers (Montalvan et al., 2016). There was no global time

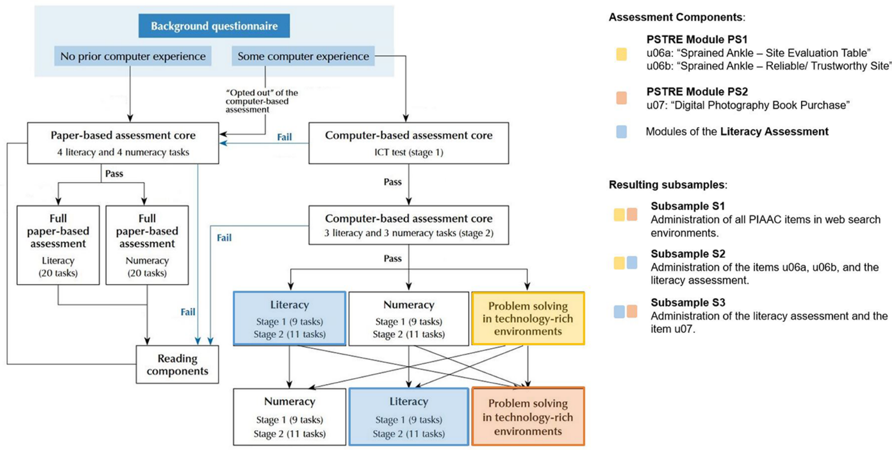

Fig. 1 Illustration of the PIAAC study design (adapted from OECD, 2013, p.62) and the composition of the subsamples S1, S2, and S3

limit, but the competence assessment was designed to take about an hour. If participants reported no computer experience, failed the basic computer tasks of the CBA Core Assessment, or refused to take the CBA option, participants were routed to a paper-based assessment of literacy and numeracy. However, most participants took the CBA, in which their click behavior was recorded with time stamps in log files (OECD, 2019). The entries in those log files (events) reflect that items have been started or completed and that participants have accessed different stimulus pages (navigation events, such as clicks on links, buttons, and menu entries) or processed the items (response events, such as clicks on radio buttons, checkboxes, or combo boxes; for detailed information, see Goldhammer et al., 2020).

Aiming at an efficient estimation of skills, PIAAC employed an adaptive design in which each participant received two so-called modules in the direct assessment. Figure 1 shows the possible combinations of modules. For reading literacy (and numeracy), a module consisted of two adaptively assigned testlets (stage 1 and stage 2) with a total of 20 items. The PSTRE assessment was not adaptive and included two modules (PS1 and PS2) consisting of seven items each.[2] The order of the items and modules was fixed (i.e., PS1 was always administered in the first module slot). It was possible to receive both PSTRE modules (without assessing reading literacy), whereas only one of the PSTRE modules could be assigned with a reading literacy module.

The PSTRE items u06a and u06b were included in the first module PS1, whereas the PSTRE item u07 was part of the second module PS2. An indicator of adults' information processing (Table 1) could only be derived for the items u06b and u07. However, we also derived indicators based on participants' performance in item u06a that served as measures of their evaluation skill (see Sect. 2.3.3).

Samples

The study of Gao et al. (2022) provided insights into different information processing behaviors based on the US and UK data of PIAAC 2011–2012. We applied this knowledge to an independent sample and investigated the PIAAC result and log file data of

[2] Further details on the items, modules, and testlets can be retrieved from the Online PIAAC Item and Log Data Documentation. https://piaac-logdata.tba-hosting.de.

ten other countries that released their data for public use (OECD, 2017a, 2017b, 2017c, 2017d, 2017e, 2017f, 2017g, 2017h, 2017i, 2017j). These countries participated in the optional PSTRE assessment and released their participants' exact age (variable AGE_R in the PIAAC Public Use Files), education level (variable EDLEVEL3: low=adults who have not attained upper secondary education; medium=adults who have attained upper secondary education; high=adults who have attained tertiary education), and gender (variable GENDER_R: 1=male; 2=female).

To pursue our research objectives, we drew three independent subsamples from the total data set (Fig. 1). For examining intraindividual differences in adults' behavior and performance on both items (u06b and u07), we selected participants who completed both PSTRE modules as subsample S1 (n=5921). We needed another sample of adults who had participated in the PIAAC literacy assessment to explain task success beyond individual background and information processing skills. Accordingly, we constructed the S2 subsample (n=3103), in which participants received the PS1 and a literacy module, and the S3 subsample (n=3278), in which participants received a literacy module and the PS2 module. Cases with missing log data or missing responses to the web search tasks were not considered.

Table 2 provides details on the subsamples' composition compared to the total sample. It is noteworthy that the participants in the subsamples were on average three to five years younger than the participants in the overall sample. They were also less likely to have a low education level and more likely to have a high education level. These findings correspond with previous results showing that older participants and participants with a low education level were less likely to participate in the PSTRE assessment (OECD, 2013, p.92; see also Goldhammer et al., 2016). Further country-specific information about the samples can be found in the supplementary material *Description by Country*.

Measures

Product and process measures of the items u06b and u07

As described in Sect. 1.2, the interactive PSTRE items u06b and u07 requested the participants to choose an option from several search results. During the tasks, the participants could click on the links and navigate the corresponding websites and their subpages. Participants' item responses were scored dichotomously according to whether they made the objectively best choice or not (0=incorrect, 1=correct). The item scores

Table 2 Sample description of the total PIAAC sample and the subsamples S1, S2, and S3

		Total	S1	S2	S3
n		63,924	5,921	3,103	3,278
Gender	% female	51.47	51.11	51.14	51.04
Age (in years)	*M*	39.68	35.31	35.91	36.72
	SD	14.61	13.57	13.90	13.79
Education level	% low	21.53	15.54	16.27	14.55
	% medium	46.82	44.18	44.89	45.18
	% high	31.66	40.28	38.83	40.27
Reading skill (literacy)	*M*	0.00	-	0.13	0.16
Evaluation skill (item u06a)	$M_{\text{Determining usefulness}}$	3.92	4.10	4.05	-
	$M_{\text{Identifying deficiencies}}$	1.65	1.76	1.74	-
Item success rate	u06b	0.55	0.57	0.54	-
	u07	0.52	0.58	-	0.50

showed that both items were of moderate difficulty, with item success rates between 50 and 58% in the subsamples (Table 2).

Using the R package *LogFSM* (Kroehne, 2019; Kroehne & Goldhammer, 2018), we translated the log events in the PIAAC logfile data into item-specific low-level features (Goldhammer et al., 2021; e.g., actions of accessing main and subpages) and summarized them into several process indicators (i.e., number of total page visits; number of main pages visited; number of subpages visited; number of revisits; last website visited; number of changes in the item response; occurrence of navigation activities after a response was given). We further contextualized these process indicators according to our coding rules (Table 1, column Operationalization). As a result, we created a process indicator representing adults' information processing pattern (flimsy, breadth-first, satisficing, sampling, laborious). Note that we also derived a time on task indicator for each item as a control variable for predicting item success. Details about the specific raw log events and transformations can be found in the supplementary material.

Reading skill

The PIAAC literacy items assessed reading skills with a variety of text formats (e.g., continuous, non-continuous texts), text types (e.g., narrations, argumentations, instructions), cognitive operations (access and identify, integrate and interpret, evaluate and reflect), and social contexts in which reading takes place (e.g., work and occupation, personal uses; OECD, 2012). The items also differentiated the medium in which reading is situated (i.e., digital vs. print context). Although the "digital" literacy items looked similar to the PSTRE items u06b and u07 (e.g., displaying a browser and navigation buttons), they focused on assessing how adults construct meaning from text (e.g., locating explicitly stated information, identifying evidence supporting or disproving a given claim, reflecting on meta-information). Accordingly, they only provided a few pages, with one to three content pages and eventually some non-content pages that were accessed when participants clicked on decorative links. Some items allowed for basic interactions, where participants could scroll or click on links, buttons, and menu entries. However, their instructions were strongly guided, specifying where to find requested information (e.g., "Look at page x"). The item response formats of the literacy items included mainly highlighting text passages but also clicking on links, radio buttons, or checkboxes.

Based on the dichotomous scores of 49 items (OECD, 2016), we determined factor scores to represent reading skill. Scores of zero indicate average reading skills; higher values represent higher skill levels (min = -3.04; max=2.26). For the score estimation, we considered the country, the final sampling weights, the hierarchical data structure, and the complex sampling design (type=complex in Mplus 7.4; Muthén & Muthén, 1998–2012). Item parameters were estimated using the maximum likelihood procedure with robust standard errors (MLR) and a logit link (n_{scaling} = 44,076). Since the likelihood of receiving a particular testlet was affected by some variables (i.e., level of education, test language, score in the CBA Core Assessment, score in the stage-1 testlet; see Kirsch et al., 2016), correlations of the latent reading factor with these variables were included in the measurement model to avoid a violation of the missing at random (MAR) assumption.

Evaluation skills

The PSTRE item u06a was used to represent evaluation skills. Based on the PIAAC log file data, we rescored participants' performance to derive indicators of how well adults determine whether an information source provides useful information ("determining usefulness") and how accurately they identify why an information source is flawed ("identifying deficiencies"). Like in other PIAAC web search items, participants received five search results. In this task, they could not interact with the links and were asked to assign an attribute to each link describing the link best. The possible attributes and short descriptions were displayed while participants worked on the task. The links could be assigned to be useful (i.e., the website appears to have reliable and trustworthy information), not authoritative (the website author appears not to be an expert), biased (the website author appears to seek influence over the audience), or not current (the website appears to be out of date). The measure for evaluating usefulness was created as the sum of links correctly assigned to be useful (ranging from 0 to 5). The indicator for identifying deficiencies was created as the sum of links correctly assigned as not authoritative, biased, or not current (ranging from 0 to 3). Both indicators were highly but not perfectly correlated ($r = .68$, $p < .001$).

Data analysis

For the examination of item success, we conducted a series of generalized linear mixed models (GLMM; De Boeck et al., 2011) to predict the probability of correctly solving a web search item for each subsample using the R package *lme4* (Bates et al., 2015). For this purpose, we first specified a baseline model with an intercept and random intercepts for the countries (models A0-C0). In the case of subsample S1, a fixed effect for items and random person intercepts were also added (A0). We then successively extended the baseline models with additional predictors. First, fixed effects were added for the variables age, education level, and gender (A1-C1). Then, depending on availability, fixed effects were added for reading skills (B2-C2), evaluation skills (A2-B2), and, as a control variable, time on task (A2-C2). Finally, fixed effects were added for the behavioral patterns of information processing (A3-C3) to investigate their unique contribution to predicting task success. We choose the flimsy pattern as reference category, as this behavior was expected to be the least successful one of all investigated behaviors. The reference category for level of education was set to 'medium' (upper secondary education) and for gender to 'male'. All continuous predictor variables were z-standardized. The time on task indicator was additionally log-transformed before standardization.

Results

Occurrence, description, and consistency across tasks

We examined how often the distinguished information processing patterns occurred per item and subsample (Table 3). Information about the occurrence of the information processing patterns by country can be found in the supplementary material *Description by Country*. We also compared our results with those of Gao et al. (2022), although the pattern frequencies are not absolutely comparable due to different operationalizations (rule-based coding vs. assignment of the most likely latent class). Nevertheless, with some exceptions, the frequencies appeared to be highly similar. The most noticeable exceptions are that the US sample showed the sampling pattern more often than the

Table 3 Occurrence of the information processing patterns in the subsamples and in the study of Gao et al. (2022; Appendix D)

Item	Pattern	Frequencies in Present Study		Class Frequencies in Gao et al. (2022)	
		S1	S2/S3	US	UK
u06b (Sprained Ankle)	Flimsy	23.1	24.3	22.7	18.7
	Breadth-first	55.2	55.3	10.7	68.9
	Satisficing	0.7	0.5	-	-
	Sampling	14.1	12.6	65.6	12.4
	Laborious	1.2	1.1	2.0	-
	Multiple	4.3	4.5	-	-
	(Multiple: Satisficing & Breadth-first / Satisficing & Sampling)	(4.2 / 0.1)	(4.5 / 0.0)	-	-
	Unassigned	1.4	1.6	-	-
u07 (Book purchase)	Flimsy	17.5	30.8	23.7	23.5
	Breadth-first	0.4	1.2	7.5	10.5
	Satisficing	15.7	22.8	15.4	10.2
	Sampling	45.4	24.3	51.6	53.3
	Laborious	7.5	4.1	1.8	2.5
	Multiple	5.9	6.6	-	-
	(Multiple: Satisficing & Breadth-first / Satisficing & Sampling)	(1.8 / 4.2)	(3.7 / 2.9)	-	-
	Unassigned	7.6	10.2	-	-

breadth-first pattern in item u06b compared to our and the UK-specific results, and that participants of S3 showed very different frequencies for some patterns compared to the other samples. This last finding is particularly interesting: While the frequencies for S1 and S2 in item u06b did not differ greatly, there were clear differences in the occurrence of the flimsy, satisficing, and sampling patterns between S1 and S3 for item u07. These differences may suggest context effects that could result from S1 participants having previously completed other PSTRE tasks (S1) instead of a different module (S3). Accordingly, S1 participants might have had a better idea of what the PSTRE tasks required from them. Like Gao and colleagues, we also observed item-specific differences in the frequency of each pattern (e.g., the satisficing pattern occurred more often in u07 than in u06b). Finally, our results showed that the behavior of a few participants could not be assigned to only one pattern (*Multiple* category) or classified according to our coding approach (*Unassigned* category). This was especially evident for item u07. Although the results suggest good coverage of the specific behaviors (i.e., 1–5 in Table 1), they also indicate that some participants process the information in the web search items in ways not clearly captured by the investigated patterns or their current operationalization.

The descriptive statistics presented in Table 4 provide further insight into adults who exhibited specific information processing patterns. Given its operationalization, it is unsurprising that participants showing a flimsy pattern finished each item the fastest (on average, it took them about a minute) and hardly ever visited a website. They also had the lowest average scores for reading and evaluation skills compared to other participants. In contrast, adults with the sampling and laborious patterns had the highest average scores for reading and evaluation skills. Both groups showed high website coverage (i.e., they tended to visit all the main pages and most subpages). Also, unsurprising, the adults with the laborious pattern were the ones who spent the longest time with the items (about a minute longer on average than the adults with the sample pattern).

Table 4 Means and standard deviations in parenthesis of skill and process indicators describing the distinguished web search behaviors

Behavior	Item	Sample	Age	Gender (% female)	Reading	Evaluation (Usefulness)	Evaluation (Deficiencies)	Time on Task (in sec)	Number of Page Visits	Main Page Visits	Subpage Visits
Flimsy	u06b	S1	35.8 (14.0)	53.3	-	3.6 (1.3)	1.3 (1.0)	65.5 (61.1)	0.5 (0.9)	0.3 (0.5)	0.0 (0.1)
		S2	36.1 (14.4)	52.2	-0.3 (0.7)	3.5 (1.4)	1.3 (1.1)	66.7 (44.3)	0.6 (1.0)	0.4 (0.5)	0.0 (0.2)
	u07	S1	37.3 (14.2)	51.2	-	3.4 (1.4)	1.1 (1.0)	59.2 (37.0)	1.5 (0.7)	1.0 (0.0)	0.4 (0.5)
		S3	39.4 (14.3)	51.7	-0.3 (0.7)	-	-	69.2 (34.3)	1.3 (0.6)	1.0 (0.0)	0.3 (0.4)
Breadth-first	u06b	S1	35.8 (13.6)	51.9	-	4.2 (1.0)	1.8 (1.0)	163.1 (86.1)	11.9 (3.6)	4.7 (0.7)	0.0 (0.0)
		S2	36.2 (13.8)	52.1	0.3 (0.6)	4.2 (1.0)	1.9 (1.0)	166.8 (83.7)	12 (3.6)	4.7 (0.7)	0.0 (0.0)
	u07	S1	37.6 (12.7)	52.2	-	4.5 (0.7)	2.1 (1.0)	117.1 (47.6)	8.1 (2.9)	3.2 (1.1)	0.0 (0.0)
		S3	40.0 (14.6)	44.7	0.3 (0.6)	-	-	119.5 (37.8)	7.6 (3.0)	2.8 (0.9)	0.0 (0.0)
Satisficing	u06b	S1	30.8 (11.3)	29.3	-	4.1 (1.0)	1.8 (1.1)	98.8 (46.5)	7.0 (1.7)	2.4 (0.8)	1.2 (0.4)
		S2	31.9 (13.8)	23.5	0.4 (0.5)	4.1 (1.0)	1.6 (1.1)	110.6 (47.5)	7.5 (1.6)	2.4 (0.8)	1.5 (0.5)
	u07	S1	34.8 (14.2)	51.2	-	3.8 (1.2)	1.4 (1.0)	102.0 (51.1)	9.8 (3.3)	3.6 (0.9)	1.9 (0.8)
		S3	35.7 (14.0)	48.1	0.2 (0.6)	-	-	116.8 (53.1)	9.6 (3.4)	3.5 (1.0)	1.9 (0.8)
Sampling	u06b	S1	33.4 (12.4)	48.4	-	4.5 (0.8)	2.2 (0.9)	187.1 (90.9)	16.3 (3.2)	5.0 (0.0)	1.7 (0.8)
		S2	34.0 (13.0)	49.5	0.5 (0.6)	4.5 (0.9)	2.2 (1.0)	191.2 (104.0)	16.1 (3.1)	5.0 (0.0)	1.6 (0.8)
	u07	S1	34.6 (12.9)	51.2	-	4.4 (0.9)	2.1 (1.0)	168.9 (61.8)	22.9 (4.4)	6.0 (0.0)	3.4 (1.2)
		S3	34.6 (12.6)	52.4	0.5 (0.6)	-	-	181.8 (72.5)	22.6 (4.5)	6.0 (0.0)	3.3 (1.1)
Laborious	u06b	S1	37.9 (14.0)	49.3	-	4.4 (0.8)	2.0 (1.0)	296.6 (139.3)	30.0 (5.1)	5.0 (0.1)	1.1 (1.3)
		S2	38.9 (12.5)	45.7	0.6 (0.6)	4.5 (0.9)	2.2 (1.0)	295.8 (110.0)	31.1 (6.6)	5.0 (0.2)	1.1 (1.1)
	u07	S1	35.5 (13.2)	44.7	-	4.5 (0.8)	2.2 (0.9)	262.3 (86.2)	42.5 (8.7)	6.0 (0.1)	4.1 (0.8)
		S3	34.8 (13.1)	50.7	0.7 (0.5)	-	-	280.8 (122.8)	41.1 (8.3)	6.0 (0.2)	3.9 (0.7)
Multiple	u06b	S1	32.8 (13.4)	45.9	-	3.9 (1.2)	1.5 (1.0)	109.8 (80.7)	6.1 (2.7)	3.1 (1.3)	0.0 (0.2)
		S2	35.1 (14.2)	45.4	-0.1 (0.7)	3.7 (1.2)	1.5 (1.1)	107.8 (73.1)	5.6 (2.5)	2.9 (1.2)	0.0 (0.1)
	u07	S1	34.1 (13.5)	58.5	-	3.9 (1.2)	1.5 (1.0)	117.1 (76.7)	10.5 (4.9)	4.9 (1.7)	1.1 (1.0)
		S3	35.3 (13.7)	57.2	0 (0.6)	-	-	108.8 (51.1)	8.2 (5.6)	3.9 (1.9)	0.8 (1.1)
Unassigned	u06b	S1	36.9 (15.0)	40.2	-	4.1 (1.2)	1.7 (1.1)	150.4 (68.5)	10.7 (3.7)	3.2 (0.8)	1.2 (0.4)
		S2	40.4 (15.0)	44.0	0.1 (0.7)	4.0 (1.4)	1.8 (1.1)	166.5 (140.4)	10.4 (2.6)	3.5 (0.8)	1.2 (0.4)
	u07	S1	36.8 (14.8)	50.9	-	4.2 (1.0)	1.8 (1.0)	294.8 (3206.8)	16.0 (5.1)	4.3 (0.9)	2.3 (1.0)
		S3	37.3 (13.7)	49.4	0.4 (0.6)	-	-	157.4 (54.5)	15.7 (5.0)	4.2 (0.9)	2.3 (1.0)

Notably, adults showing a mixed pattern of searching broadly but stopping as soon as something suitable has been found (*Multiple* category) tended to have moderate or low reading and evaluation scores compared to the other groups.

For subsample S1, we also inspected how consistently adults stick to certain information processing behaviors across the two tasks. Figure 2 confirms what the pattern frequencies suggest across items: Most participants (79.5% of S1) switched their information processing pattern in the second PIAAC web search task. Notably, participants who showed a similar behavior in both items were likely to display either a flimsy or a sampling pattern. When switching, participants were most likely to switch from a breadth-first to a satisficing or sampling pattern, which might be explained by the design of item u07 that specifically required to pay attention to shipping dates often displayed on subpages.

Explaining item success

Table 5 presents the results for the subsample S1 (models A0-A3). The results of model A1 showed a decrease in the probability of task success by age (b = -0.15) and by a lower than medium level of education (b = -0.26) and an increase by a higher than medium level of education (b=0.39). Apart from this, task success did not significantly differ by gender. The background variables explained about 18% of the interindividual variation observed in the baseline model. Adding the two evaluation skill variables and time on task increased the explained interindividual variance by another 58%, with the effects of age and education level remaining statistically significant. Remarkably, the two evaluation variables, determining usefulness (b=0.13) and identifying deficiencies (b=0.18), each made an independent explanatory contribution to task success, also lowering the effects of having a lower (b = -0.17) or higher education level (b=0.20). Adding the information processing patterns led to a point where specifying interindividual random intercepts overfitted the data. That means interindividual differences in task performance were attributed entirely to variables of individual background, skill, and behavior, possibly because interindividual differences are estimated based on only two items. Compared to the flimsy behavior, all behaviors were associated with a higher probability of successfully solving the web search tasks, except for the *Multiple* category, which showed a significantly lower chance (b = -0.57). Across the items, the satisficing (b=2.08), sampling (b=2.03), and laborious patterns (b=2.35) were particularly effective. The effect of age was also no longer statistically significant, suggesting that the information processing patterns could explain parts of age-related differences in the web search tasks.

Tables 6 and 7 present the results of the item-specific analyses for subsamples S2 and S3. Concerning item u06b (models B0-B3), task success was initially significantly predicted only by having a lower level of education (b = -0.30). However, this effect vanished after adding the reading and evaluation skills and time on task, with only reading (b=0.28) and the aspect of determining the usefulness of online information (b=0.17) significantly contributing to the prediction of task success. Adding the information processing patterns then revealed that adults with the satisficing pattern (b=1.26), sampling pattern (b=0.92) and, interestingly, the *Unassigned* category (b=0.87) were more successful in selecting relevant and trustworthy information than adults exhibiting any other pattern. For item u07 (models C0-C3), we observed significant effects of age (b = -0.28) and having a lower (b = -0.37) or higher level of education (b=0.68). The effects

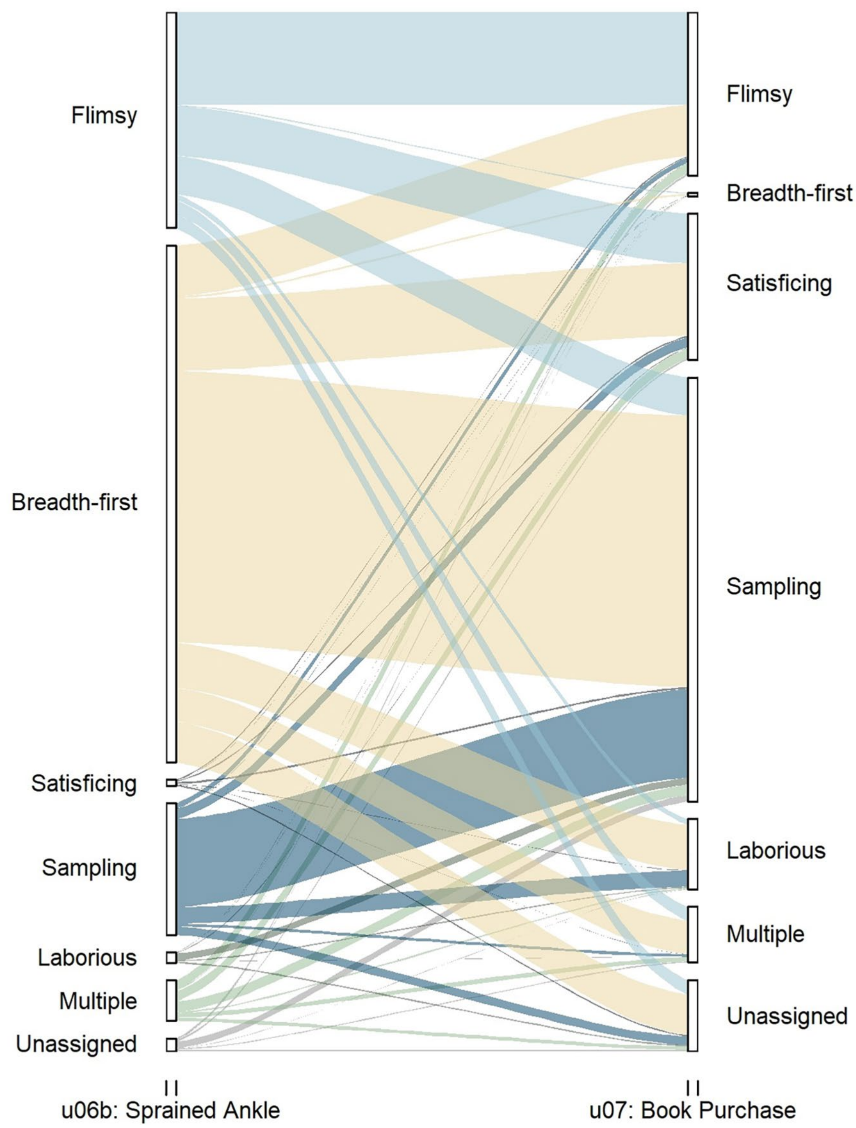

Fig. 2 Flow diagram of behavior categories across web search items (subsample S1). Box sizes are proportional to the category sizes. Line thickness represents the number of individuals within connected boxes. Colors indicate the category in the item u06b

of the education level disappeared after including reading skills (b=0.86) and time on task (b=1.23), while the effect of age remained statistically significant (b = -0.17). Like in subsample S1, the effect of age was no longer statistically significant after including information processing behavior. The most successful patterns for purchasing a particular product online were satisficing (b=2.46), sampling (b=2.41), laborious (b=2.85), and the *Unassigned* category (b=2.46).

Discussion

The present study focused on the theory-based creation and investigation of a process indicator representing different information processing patterns of adults searching for information online. We examined how often and consistently PIAAC participants exhibited these patterns and how likely they solved the corresponding web search task, controlling for age, educational background, gender, reading and evaluation skills, and

Table 5 GLMM results of predicting adults' success in web search tasks (subsample S1)

Model	A0	A1	A2	A3
Intercept	0.23 (0.09) **	0.10 (0.10)	0.09 (0.10)	0.87 (0.13) ***
Item u07	0.04 (0.04)	0.04 (0.04)	0.00 (0.04)	-1.07 (0.07) ***
Background variables				
Age (linear)	-	-0.15 (0.02) ***	-0.14 (0.02) ***	-0.02 (0.03)
Age (quadratic)	-	-0.01 (0.03)	-0.02 (0.03)	-0.04 (0.03)
Education level (low)	-	-0.26 (0.06) ***	-0.17 (0.06) **	-0.13 (0.06) *
Education level (high)	-	0.39 (0.05) ***	0.20 (0.05) ***	0.10 (0.05) *
Gender (female)		-0.04 (0.04)	-0.02 (0.04)	0.03 (0.04)
Skill variables				
Determining usefulness (u06a)	-	-	0.13 (0.03) ***	0.11 (0.03) ***
Identifying deficiencies (u06a)	-	-	0.18 (0.03) ***	0.13 (0.03) ***
Behavioral variables				
Time on task	-	-	0.48 (0.03) ***	0.02 (0.04)
Behavior: Breadth-first	-	-	-	0.33 (0.08) ***
Behavior: Satisficing	-	-	-	2.08 (0.10) ***
Behavior: Sampling	-	-	-	2.03 (0.09) ***
Behavior: Laborious	-	-	-	2.35 (0.15) ***
Behavior: Multiple	-	-	-	-0.57 (0.12) ***
Behavior: Unassigned	-	-	-	1.73 (0.12) ***
SD random country intercepts	0.20	0.22	0.20	0.19
SD random person intercepts	0.39	0.35	0.19	0.00[1]
Interindividual variance explained	-	0.18	0.76	1.00[1]

Notes. The reference categories of the categorical variables are "item u06b", "medium education level", "male", and "flimsy". [1]The model revealed a singular fit, indicating overfit. That means that the specified random effect structure (here random intercept for person) is overly complex. There is not systematic effect coming from person

Table 6 GLMM results of predicting success in selecting a reliable information source (Item u06b *Sprained ankle*; subsample S2)

Model	B0	B1	B2	B3
Intercept	0.18 (0.07) *	0.17 (0.10)	0.17 (0.10)	-0.01 (0.13)
Background variables				
Age (linear)	-	-0.03 (0.04)	0.05 (0.05)	0.09 (0.05)
Age (quadratic)	-	0.01 (0.05)	0.02 (0.05)	0.01 (0.05)
Education level (low)	-	-0.30 (0.11) **	-0.16 (0.12)	-0.18 (0.12)
Education level (high)	-	0.13 (0.08)	-0.09 (0.09)	-0.12 (0.09)
Gender (female)	-	-0.02 (0.07)	-0.01 (0.07)	0.01 (0.08)
Skill variables				
Reading skill	-	-	0.28 (0.05) ***	0.23 (0.05) ***
Determining usefulness (u06a)	-	-	0.17 (0.06) **	0.17 (0.06) **
Identifying deficiencies (u06a)	-	-	0.06 (0.05)	0.06 (0.05)
Behavioral variables				
Time on task	-	-	0.05 (0.05)	-0.07 (0.06)
Behavior: Breadth-first	-	-	-	0.18 (0.12)
Behavior: Satisficing	-	-	-	1.26 (0.59) *
Behavior: Sampling	-	-	-	0.92 (0.17) ***
Behavior: Laborious	-	-	-	0.22 (0.38)
Behavior: Multiple	-	-	-	-0.05 (0.19)
Behavior: Unassigned	-	-	-	0.87 (0.33) **
SD random country intercepts	0.19	0.19	0.19	0.21

Notes. The reference categories of the categorical variables are "medium education level", "male", and "flimsy"

Table 7 GLMM results of predicting success in buying a product that matches a list of specified criteria (Item u07 *Book Purchase*; subsample S3)

Model	C0	C1	C2	C3
Intercept	-0.00 (0.08)	-0.23 (0.12)	-0.48 (0.15) **	-1.88 (0.18) ***
Background variables				
Age (linear)	-	-0.28 (0.04) ***	-0.17 (0.05) **	0.01 (0.06)
Age (quadratic)	-	-0.06 (0.05)	-0.04 (0.06)	-0.07 (0.06)
Education level (low)	-	-0.37 (0.12) **	-0.04 (0.13)	0.07 (0.15)
Education level (high)	-	0.68 (0.08) ***	0.10 (0.10)	0.14 (0.11)
Gender (female)	-	0.00 (0.07)	0.08 (0.08)	0.11 (0.09)
Skill variables				
Reading skill	-	-	0.86 (0.06) ***	0.73 (0.07) ***
Behavioral variables				
Time on task	-	-	1.23 (0.07) ***	0.27 (0.10) **
Behavior: Breadth-first	-	-	-	-0.64 (0.55)
Behavior: Satisficing	-	-	-	2.46 (0.15) ***
Behavior: Sampling	-	-	-	2.41 (0.18) ***
Behavior: Laborious	-	-	-	2.85 (0.32) ***
Behavior: Multiple	-	-	-	0.01 (0.23)
Behavior: Unassigned	-	-	-	2.46 (0.19) ***
SD random country intercepts	0.23	0.27	0.36	0.36

Notes. The reference categories of the categorical variables are "medium education level", "male", and "flimsy"

differences in the amount of time spent on the task. In summary, we found item-specific differences in how frequently the behavioral patterns of interest occurred in the data, with a tendency of individuals not to settle on a single behavior across items. The patterns explained task success beyond reading and evaluation skills, with the satisficing and sampling patterns and the *Unassigned* category being consistently among the most effective patterns for solving a task correctly across items. In the following, we discuss implications for using and interpreting rule-based process indicators and reflect on the results in light of interindividual differences in web search.

Pros and cons of rule-based process indicators

Informed by previous research and theoretical concepts of why specific behavior occurs, assumptions about behavior in test situations can guide the creation of process indicators (see Hahnel et al., 2019; Salles et al., 2020). Our results stress this reasoning for web search tasks. Instead of using simple process indicators, such as the number of website visits, we depicted specific behavioral patterns by contextualizing multiple process indicators in a rule-based way. This approach enabled us to obtain results largely similar to those of Gao et al. (2022), who used a data-driven approach with a PIAAC dataset independent of the dataset used in our study. Our results do not imply that rule-based process indicators should be preferred over data-driven ones. They rather demonstrate how the rule-based approach can support the construction of process indicators with an interpretation (e.g., solution strategy) comparable across tasks but regarding the underlying task-specific low-level features. We discuss some of the advantages and limitations of the rule-based approach in detail below.

Manually constructed rule-based indicators are strict in terms of the behavior they capture, reducing interpretative ambiguity at the cost of flexibility. Gaps or alternative interpretations within a category are rather a consequence of incomplete and imprecise rules. For example, according to our operationalization of the sampling pattern, we know that adults who showed it have visited all main pages and at least one subpage. However, one might wonder whether the rule should include individuals who did not access a website that could be identified as inappropriate by its link information on the search result page or whether the sampling label should apply exclusively to adults who visited all available websites (including all subpages). These are questions that are ideally considered during the design of a task. Knowing up front what behavior should be considered, rule-based process indicators can then even be created and analyzed in real time, making data preparation and analysis steps more efficient. As intermediate products within an ongoing assessment, they could be used predict how a test session progresses (e.g., Ulitzsch et al., 2022) and whether or not corrective measures are required. For example, due to its low activity profile, the flimsy pattern might indicate that a person is having difficulties completing the task. Accordingly, the pattern's observation could be incorporated into feedback to teachers or trainers during a test session.

The clarity and precision of the behavior descriptions are central to deriving operationalization rules and may be the most troublesome element in creating rule-based process indicators. Deriving the rule of the laborious pattern, for example, was problematic since there was no clear idea of what an "extensive use of the infrastructure" is. We decided on a rule that considers the item's hypertext structure but, despite being our best guess, the resulting thresholds are still arbitrary. Similarly, we did not integrate time on task in our rule definition because we did not have informed assumptions to set appropriate thresholds, which ideally should consider individual processing speed. Alternatives that circumvent the problem of setting thresholds a priori could be approaches that combine rule specification with data-driven learning, such as first-order inductive learner (FOIL). FOIL is a rule-based machine learning algorithm that determines rules bottom-up while top-down constraints can be applied. Such an algorithm can also be used to determine which rule is most important (or has the greatest discriminative power) for classifying a given case with respect to a particular information processing pattern. Provided training data is available, FOIL could be a promising method for studies seeking to evaluate and improve the accuracy of pattern classifications.

With the description of expected behavior, new or unexpected patterns can also become visible. Adults in the *Unassigned* category were surprisingly among the best performers. Inspecting their descriptive statistics (Table 4), one could speculate that these adults singled out a small set of information options and systematically compared them. For both items, u06b and u07, they visited more main pages and subpages than the satisficing group but less than the sampling group and seemingly revisited some pages. Such a pattern might represent an adaptive combination of satisficing and sampling strategies (i.e., making a preselection of pages that look "good enough" and then inspecting and comparing them to find "the best one"). In this case, the order of actions would be particularly important, which could be further investigated by sequence mining techniques (e.g., He et al., 2019). Moreover, the *Unassigned* category could be a more suitable representation of actual web search behavior than, for example, sampling or laborious patterns that make sense in closed information spaces with preselected content. In this respect, research aiming at cross-validation and investigating differences in information processing behaviors for tasks in closed (e.g., PIAAC web search tasks) and open information spaces (tasks involving the Internet) would be of great value.

Interpreting process indicators as strategy use

Interpreting the occurrence of behavioral patterns in terms of strategy use is a strong inference. Following a strategy is a conscious and deliberate act of systematically controlling and modifying one's efforts to work toward a self-set goal (see Afflerbach et al., 2008). Accordingly, strategies are the products of intentions, but with log data, we can only observe the aftermath of how both intentions and chosen strategies played out. It becomes clear that inferring strategy use from the patterns requires validation.

For the satisficing and sampling patterns, there is a theory-based plausibility to the argument that they result from the application of satisficing and sampling strategies, motivated by the goal to perform well (in the case of satisficing with the secondary goal of saving time and effort; Reader & Payne, 2007). In contrast, the line of argument is less clear for the flimsy and the breadth-first patterns. On the one side, adults with these patterns might have also followed the goal of performing well but lacked the means to choose appropriate actions or the metacognitive skill to monitor and adjust their behavior according to the task demands. On the other side, these adults might have been instead motivated by finishing up as fast as possible, resulting in superficial test-taking. For the flimsy pattern, the result that adults tend to stick to the flimsy pattern across tasks speaks for the latter. Accordingly, it may indicate test-taking disengagement (see Goldhammer et al., 2016; OECD, 2019). Another validation strategy for such an interpretation might involve a microanalytic look at talk and gesture data (e.g., Maddox, 2017). Such fine-grained observations, obtained during test situations, could offer insights into how tests are received and administered in different socioeconomic and cultural settings (e.g., how interviewers deal with fatigued participants or participants seeking support).

Another example that highlights the necessity of validation concerns the interpretation of the laborious pattern. Juvina and Oostendorp (2006) argued that individuals exhibit this pattern to ensure a good representation of the information structure. In line with this reasoning, they showed that the pattern was correlated with high episodic memory and low spatial ability. Our results contribute to their idea by indicating that the laborious pattern was associated with item success, especially in the *Book Purchase* task (u07). Speaking against it, though, participants with the laborious pattern in the study of Gao et al. (2022) showed only moderate task success (60–78% success rates across items and samples). The authors argued that adults might get lost in the information space because they fail to interpret and integrate ideas across sources due to low reading skills. We controlled in our analyses for information processing skills, such as reading, and the adults with the laborious pattern in our samples were descriptively among the better readers. In this respect, the interpretation of Gao and colleagues might not represent a contradiction but demonstrates the need to examine the contexts in which specific patterns occur closely.

Revisiting interindividual differences in web search

There are implications that our study can contribute to the body of research on individual differences in web search. First, we showed that specific patterns associated with a systematic way of information processing explain web search performance, independent of distal variables of individual background and proximal variables of information processing skills. This is good news for educators, as this result suggests that individuals can improve their use of online information despite unfavorable preconditions. Potential strategy instructions should consider task-specific needs, as our results, like those of Gao et al. (2022), indicate differences in the appropriateness and effectiveness of certain behaviors for tasks focusing on information location versus information evaluation. Directions of future research could address changes in information processing patterns across tasks, examining individuals' repertoires of strategies and their ability to adapt to changing conditions in the information landscape.

Furthermore, our results align with findings from small-scale research on age differences (e.g., Chevalier et al., 2015; Morrison, 2015), extending them to a broad international sample of different ages in adulthood. We found age differences in two of our three subsamples: Adults were less likely to solve the PIAAC web search tasks with increasing age correctly. This was especially pronounced in the *Book Purchase* task (u07), which corresponds with the finding that older adults experience more difficulty locating specific targets on the web than younger adults (Etcheverry et al., 2012). After accounting for educational background, gender, and information processing skills, the remaining age differences could be explained in terms of the information processing behaviors, suggesting that older adults can overcome difficulties in finding information by adopting appropriate information processing strategies. Although encouraging, these results should nevertheless be viewed with caution, as older adults were also more likely to not participate in the computer-based assessment of PIAAC (OECD, 2013).

The missing age-related difference in the reliability judgment task (u06b) might result from including educational attainment in our analyses, which we partly did to better separate cohort from age effects. Cohort effects could also explain why we did not observe a quadratic effect of age at all (see OECD, 2013, p. 191). Notably, the effects of the educational level disappeared after reading skills were included as a predictor. Accordingly, good web search decisions do not depend directly on a person's formal education. Instead, adults with higher educational qualifications are more likely to possess proficient reading skills, which are a crucial resource for solving information problems and lifelong learning.

Regarding gender differences and in line with the work of Hämäläinen et al. (2020), we could not find any differences in web search performance between females and males that were not already explained by age and educational background. In terms of behavioral differences (Table 4), notable descriptive differences only occurred for exhibiting the satisficing pattern in the reliability judgment task (u06b), where females seemed less likely to show this pattern. However, the satisficing pattern was rarely observed in the reliability judgment task (Table 3).

Finally, our study underlines the role of reading and evaluation skills as essential prerequisites for solving information problems (Brand-Gruwel et al., 2009; Hahnel et al., 2018; Salmerón et al., 2018). The indicator on determining the usefulness of an information source based on link information contributed to item success beyond reading skill in sample S2 (Table 6), adding support to the assumption that individuals use link information to make predictive judgments to anticipate a website's value for a task (Hilligoss & Rieh, 2008). Moreover, the two evaluation skill indicators independently predicted item success in sample S1 (Table 5), suggesting separable aspects of link evaluation. This result is consistent with the idea that web users who are able to identify useful content may still fall for deception or misinformation if they fail to perceive and interpret link features that indicate limitations in the credibility of an information source (see

George et al., 2016). For example, suppose a web user believes they read independently researched information, not noticing features that identify this information as sponsored content. In that case, there is little incentive for that web user to investigate further the intention of the source (see Wineburg et al., 2018). Further research in this direction is needed, especially as our results rely only on two indicators from a single task not explicitly designed to measure different aspects in link evaluation.

Limitations

The PIAAC study was not designed to generate generalizable results about information processing behaviors in web search. Accordingly, we could only use two items addressing a web search problem and one item tapping on evaluation skills in the context of web search. This situation raises questions on the stability of the observed effects, especially concerning different requirements in web search tasks (locating information vs. evaluating information). Our predictor selection was also restricted to available variables, therefore failing to consider the effects of other skills and experiences critical in web search (e.g., prior knowledge, topic interest, attitudes and beliefs). Furthermore, to investigate the predictive contribution of different information processing patterns beyond reading and evaluation skills, we had to split the PIAAC data into three independent datasets, ignoring participants who worked on PSTRE and numeracy tasks and reducing the test power of our analyses. As a cross-sectional study, PIAAC provides a snapshot of the individual skills at a particular point in time. Accordingly, any directed interpretations of the observed effects are not supported by the study design or our correlation-based analytical approach but are only motivated by the theoretical background and previous research. However, a major strength of the PIAAC data is its cross-national representative sample, covering a wide range of age and educational backgrounds. As our analyses show, there is still much to learn from this unique database.

Conclusions

Our study highlights the strengths and pitfalls of creating rule-based process indicators informed by previous research and assumptions about information processing behavior. Rather than using simple process indicators of test-taking behavior (e.g., number of page visits), we reviewed strategies that are supposed to be behind the specific patterns and portrayed how applying a specific strategy would result in item interactions, leading to rules on how to transform low-level features into a meaningful process indicator. This top-down approach has strengths in reducing interpretative ambiguities and pointing out patterns that might be overlooked, but also has the downside of requiring precise a priori rule specifications. Nevertheless, it allows for a targeted examination of behaviors of interest and can potentially support educational intervention during a test session. Beyond indicator construction and validation, our study also contributes by shedding light on the role that individual background, information processing skills and behavior play when dealing with online information. A highlight is that information processing skills and behavior explained large parts of the effects of individual background, suggesting that adults can overcome shortcomings in web search performance by learnable skills and strategies. In other words, they can improve their use of online information by adopting malleable information processing strategies despite unfavorable preconditions.

Supplementary Information

The online version contains supplementary material available at https://doi.org/10.1186/s40536-023-00169-5.

Supplementary Material 1

Acknowledgements

Not applicable.

Authors' contributions

CH developed the idea and wrote the first draft of the manuscript. UK extended the R package LogFSM to process PIAAC data and reviewed and edited the manuscript. FG supported the manuscript preparation by commenting on the analyses and reviewing and editing parts of the manuscript. All authors read and approved the final manuscript.

Funding

This work was supported by the Centre for International Student Assessment (ZIB) in Germany.

Data Availability

The datasets generated and/or analysed during the current study are available in the repositories of the OECD and the GESIS Leibniz Institute for the Social Sciences, http://www.oecd.org/skills/piaac/publicdataandanalysis/ and https://search.gesis.org/research_data/ZA6712.

Declarations

Competing interests

The authors declare that they have no competing interests.

Received: 21 July 2022 / Accepted: 10 May 2023

Published online: 19 May 2023

References

Afflerbach, P., Pearson, P. D., & Paris, S. G. (2008). Clarifying differences between reading skills and reading strategies. *The Reading Teacher*, *61*(5), 364–373. https://doi.org/10.1598/RT.61.5.1.

Bendersky, M., Metzler, D., & Croft, W. B. (2012). Effective query formulation with multiple information sources. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining - WSDM '12*, 443. https://doi.org/10.1145/2124295.2124349

Blackmon, M. H. (2012). Information scent determines attention allocation and link selection among multiple information patches on a webpage. *Behaviour & Information Technology*, *31*(1), 3–15. https://doi.org/10.1080/0144929X.2011.599041.

Braasch, J. L. G., & Graesser, A. C. (2020). Avoiding and Overcoming Misinformation on the Internet. In R. J. Sternberg & D. F. Halpern (Eds.), *Critical Thinking in Psychology* (2nd ed., pp. 125–151). Cambridge University Press. https://doi.org/10.1017/9781108684354.007

Brand-Gruwel, S., Wopereis, I., & Vermetten, Y. (2005). Information problem solving by experts and novices: Analysis of a complex cognitive skill. *Computers in Human Behavior*, *21*(3), 487–508. https://doi.org/10.1016/j.chb.2004.10.005.

Brand-Gruwel, S., Wopereis, I., & Walraven, A. (2009). A descriptive model of information problem solving while using internet. *Computers & Education*, *53*(4), 1207–1217. https://doi.org/10.1016/j.compedu.2009.06.004.

Bryan, J., Luszcz, M. A., & Crawford, J. R. (1997). Verbal knowledge and speed of Information Processing as Mediators of Age differences in Verbal Fluency Performance among older adults. *Psychology and Aging*, *12*(3), 473–478.

Chevalier, A., Dommes, A., & Marquié, J. C. (2015). Strategy and accuracy during information search on the web: Effects of age and complexity of the search questions. *Computers in Human Behavior*, *53*, 305–315. https://doi.org/10.1016/j.chb.2015.07.017.

Eichmann, B., Greiff, S., Naumann, J., Brandhuber, L., & Goldhammer, F. (2020). Exploring behavioural patterns during complex problem-solving. *Journal of Computer Assisted Learning*, *36*, 933–956. https://doi.org/10.1111/jcal.12451.

Etcheverry, I., Baccino, T., Terrier, P., Marquié, J. C., & Mojahid, M. (2012). Age differences in information finding tasks: Performance and visual exploration strategy with different web page layouts. *Computers in Human Behavior*, *28*(5), 1670–1680. https://doi.org/10.1016/j.chb.2012.04.005.

Evans, J., & St., B. T. (2008). Dual-Processing Accounts of reasoning, Judgment, and Social Cognition. *Annual Review of Psychology*, *59*(1), 255–278. https://doi.org/10.1146/annurev.psych.59.103006.093629.

Flanagin, A. J., & Metzger, M. J. (2007). The role of site features, user attributes, and information verification behaviors on the perceived credibility of web-based information. *New Media & Society*, *9*(2), 319–342. https://doi.org/10.1177/1461444807075015.

Forzani, E., Corrigan, J., & Kiili, C. (2022). What does more and less effective internet evaluation entail?: Investigating readers' credibility judgments across content, source, and context. *Computers in Human Behavior*, *135*, 107359. https://doi.org/10.1016/j.chb.2022.107359.

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2020). *Preparing for life in a Digital World: IEA International computer and information literacy study 2018 International Report*. Springer International Publishing. https://doi.org/10.1007/978-3-030-38781-5.

Gao, Y., Cui, Y., Bulut, O., Zhai, X., & Chen, F. (2022). Examining adults' web navigation patterns in multi-layered hypertext environments. *Computers in Human Behavior*, *129*, 107142. https://doi.org/10.1016/j.chb.2021.107142.

George, J. F., Giordano, G., & Tilley, P. A. (2016). Website credibility and deceiver credibility: Expanding prominence-interpretation theory. *Computers in Human Behavior*, *54*, 83–93. https://doi.org/10.1016/j.chb.2015.07.065.

Goldhammer, F., Naumann, J., Stelter, A., Tóth, K., Rölke, H., & Klieme, E. (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*, *106*(3), 608–626.

Goldhammer, F., Martens, T., Christoph, G., & Lüdtke, O. (2016). *Test-taking engagement in PIAAC* (133No. vol.). OECD Publishing. https://doi.org/10.1787/5jlzfl6fhxs2-en.

Goldhammer, F., Hahnel, C., & Kroehne, U. (2020). Analysing Log File Data from PIAAC. In D. B. Maehler & B. Rammstedt (Eds.), *Large-Scale Cognitive Assessment: Analyzing PIAAC Data* (pp. 239–269). Springer. https://link.springer.com/content/pdf/10.1007%2F978-3-030-47515-4.pdf

Goldhammer, F., Hahnel, C., Kroehne, U., & Zehner, F. (2021). From byproduct to design factor: On validating the interpretation of process indicators based on log data. *Large-Scale Assessments in Education*, *9*(1), 1–25. https://doi.org/10.1186/s40536-021-00113-5.

Goldman, S. R., Braasch, J. L. G., Wiley, J., Graesser, A. C., & Brodowinska, K. (2012). Comprehending and learning from internet sources: Processing patterns of Better and Poorer Learners. *Reading Research Quarterly*, *47*(4), 356–381. https://doi.org/10.1002/rrq.027.

Grimes, G. A., Hough, M. G., Mazur, E., & Signorella, M. L. (2010). Older adults' knowledge of internet hazards. *Educational Gerontology*, *36*(3), 173–192. https://doi.org/10.1080/03601270903183065.

Hahnel, C., Goldhammer, F., Kröhne, U., & Naumann, J. (2018). The role of reading skills in the evaluation of online information gathered from search engine environments. *Computers in Human Behavior*, *78*, 223–234. https://doi.org/10.1016/j.chb.2017.10.004.

Hahnel, C., Kroehne, U., Goldhammer, F., Schoor, C., Mahlow, N., & Artelt, C. (2019). Validating process variables of sourcing in an assessment of multiple document comprehension. *British Journal of Educational Psychology*, *89*, 524–537. https://doi.org/10.1111/bjep.12278.

Hahnel, C., Eichmann, B., & Goldhammer, F. (2020). Evaluation of Online Information in University students: Development and scaling of the Screening Instrument EVON. *Frontiers in Psychology*, *11*, 1–16. https://doi.org/10.3389/fpsyg.2020.562128.

Hahnel, C., Ramalingam, D., Kroehne, U., & Goldhammer, F. (2022). Patterns of reading behaviour in digital hypertext environments. *Journal of Computer Assisted Learning*, 1–14. https://doi.org/10.1111/jcal.12709.

Hämäläinen, E. K., Kiili, C., Marttunen, M., Räikkönen, E., González-Ibáñez, R., & Leppänen, P. H. T. (2020). Promoting sixth graders' credibility evaluation of web pages: An intervention study. *Computers in Human Behavior*, *110*, 106372. https://doi.org/10.1016/j.chb.2020.106372.

Hargittai, E., & Hinnant, A. (2008). Digital Inequality: Differences in young adults' use of the internet. *Communication Research*, *35*(5), 602–621. https://doi.org/10.1177/0093650208321782.

He, Q., & von Davier, M. (2015). Identifying Feature Sequences from Process Data in Problem-Solving Items with N-Grams. In L. A. van der Ark, D. M. Bolt, W.-C. Wang, J. A. Douglas, & S.-M. Chow (Eds.), *Quantitative Psychology Research* (Vol. 140, pp. 173–190). Springer International Publishing. https://doi.org/10.1007/978-3-319-19977-1_13

He, Q., Borgonovi, F., & Paccagnella, M. (2019). *Using Process Data to Understand Adults' Problem-Solving Behaviour in the Programme for the International Assessment of Adult Competencies (PIAAC): Identifying Generalised Patterns across Multiple Tasks with Sequence Mining* (OECD Education Working Papers No. 205). OECD.

Hilligoss, B., & Rieh, S. Y. (2008). Developing a unifying framework of credibility assessment: Construct, heuristics, and interaction in context. *Information Processing & Management*, *44*(4), 1467–1484. https://doi.org/10.1016/j.ipm.2007.10.001.

Hölscher, C., & Strube, G. (2000). Web search behavior of internet experts and newbies. *Computer Networks*, *33*, 337–346. https://doi.org/10.1016/S1389-1286(00)00031-1.

Jenkins, C., Corritore, C. L., & Wiedenbeck, S. (2003). Patterns of information seeking on the web: A qualitative study of domain expertise and web expertise. *IT & Society*, *1*(3), 64–89.

Juvina, I., & van Oostendorp, H. (2006). Individual differences and behavioral metrics involved in modeling web navigation. *Universal Access in the Information Society*, *4*(3), 258–269. https://doi.org/10.1007/s10209-005-0007-7.

Kammerer, Y., Bråten, I., Gerjets, P., & Strømsø, H. I. (2013). The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior*, *29*(3), 1193–1203. https://doi.org/10.1016/j.chb.2012.10.012.

Kammerer, Y., Amann, D. G., & Gerjets, P. (2015). When adults without university education search the internet for health information: The roles of internet-specific epistemic beliefs and a source evaluation intervention. *Computers in Human Behavior*, *48*, 297–309. https://doi.org/10.1016/j.chb.2015.01.045.

Keller, L., Zenisky, A. L., & Wang, X. (2016). Analyzing process data from Technology-Rich Tasks. In Y. Rosen, S. Ferrara, & M. Mosharraf (Eds.), *Handbook of Research on Technology Tools for Real-World Skill Development* (pp. 724–748). IGI Global.

Kiili, C., Leu, D. J., Marttunen, M., Hautala, J., & Leppänen, P. H. T. (2018). Exploring early adolescents' evaluation of academic and commercial online resources related to health. *Reading and Writing*, *31*(3), 533–557. https://doi.org/10.1007/s11145-017-9797-2.

Kirsch, I., Yamamoto, K., & Garber, D. (2016). PIAAC Assessment Design. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC) (2nd Edition)* (pp. 1–21). OECD Publishing. https://www.oecd.org/skills/piaac/PIAAC_Technical_Report_2nd_Edition_Full_Report.pdf

Kroehne, U. (2019). *LogFSM: Analyzing Log Data from Educational Assessments using Finite State Machines (LogFSM)*. http://www.logfsm.com

Kroehne, U., & Goldhammer, F. (2018). How to conceptualize, represent, and analyze log data from technology-based assessments? A generic framework and an application to questionnaire items. *Behaviormetrika*, *45*(2), 527–563. https://doi.org/10.1007/s41237-018-0063-y.

Kubeck, J. E., Miller-Albrecht, S. A., & Murphy, M. D. (1999). Finding information on the world wide web: Exploring older adults' exploration. *Educational Gerontology*, *25*(2), 167–183. https://doi.org/10.1080/036012799267945.

Leu, D. J., Forzani, E., Rhoads, C., Maykel, C., Kennedy, C., & Timbrell, N. (2014). The New Literacies of Online Research and Comprehension: Rethinking the Reading achievement gap. *Reading Research Quarterly*, *50*(1), 37–59. https://doi.org/10.1002/rrq.85.

Lucassen, T., Muilwijk, R., Noordzij, M. L., & Schraagen, J. M. (2013). Topic familiarity and information skills in online credibility evaluation. *Journal of the American Society for Information Science and Technology*, *64*(2), 254–264. https://doi.org/10.1002/asi.22743.

Maddox, B. (2017). Talk and gesture as process data. *Measurement: Interdisciplinary Research and Perspectives*, *15*(3–4), 113–127. https://doi.org/10.1080/15366367.2017.1392821.
Maehler, D. B., & Rammstedt, B. (Eds.). (2020). *Large-Scale Cognitive Assessment: Analyzing PIAAC Data*. Springer International Publishing. https://doi.org/10.1007/978-3-030-47515-4.
Massing, N., & Schneider, S. L. (2017). Degrees of competency: The relationship between educational qualifications and adult skills across countries. *Large-Scale Assess Educ*, *5*(1), 6. https://doi.org/10.1186/s40536-017-0041-y.
Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, *59*, 210–220. https://doi.org/10.1016/j.pragma.2013.07.012.
Mislevy, R. J., Almond, R. G., & Lukas, J. F. (2003). A brief introduction to evidence-centered design. *ETS Research Report Series*, *1*, 1–29. https://doi.org/10.1002/j.2333-8504.2003.tb01908.x.
Molerov, D., Zlatkin-Troitschanskaia, O., Nagel, M. T., Brückner, S., Schmidt, S., & Shavelson, R. J. (2020). Assessing University students' critical online reasoning ability: A conceptual and Assessment Framework with preliminary evidence. *Frontiers in Education*, *5*, 577843. https://doi.org/10.3389/feduc.2020.577843.
Monchaux, S., Amadieu, F., Chevalier, A., & Mariné, C. (2015). Query strategies during information searching: Effects of prior domain knowledge and complexity of the information problems to be solved. *Information Processing & Management*, *51*(5), 557–569. https://doi.org/10.1016/j.ipm.2015.05.004.
Montalvan, P., Lemay, M., & Thornton, N. (2016). Field operations. In OECD (Ed.), *Technical Report of the Survey of Adult Skills (PIAAC) (2nd Edition)* (pp. 1–46). OECD Publishing.
Morrison, R. (2015). Silver surfers search for gold: A study into the online information-seeking skills of those over fifty. *Ageing Int*, *40*(3), 300–310. https://doi.org/10.1007/s12126-015-9224-4.
Muthén, L. K., & Muthén, B. O. (1998-2012). *Mplus user's guide* (7th ed.). Muthén & Muthén.
Naumann, J., & Sälzer, C. (2017). Digital reading proficiency in german 15-year olds: Evidence from PISA 2012. *Zeitschrift Für Erziehungswissenschaft*, *20*(4), 585–603.
OECD. (2012). *Literacy, numeracy and problem solving in Technology-Rich environments: Framework for the OECD Survey of adult skills*. OECD Publishing. https://doi.org/10.1787/9789264128859-en.
OECD. (2013). *OECD Skills Outlook 2013: First results from the survey of adult skills*. OECD Publishing. http://www.oecd-ilibrary.org/education/oecd-skills-outlook-2013_9789264204256-en.
OECD. (2016). *The Survey of Adult Skills: Reader's Companion*, Second Edition. OECD. https://doi.org/10.1787/9789264258075-en
OECD (2019). *Beyond Proficiency: Using Log Files to Understand Respondent Behaviour in the Survey of Adult Skills*. OECD. https://doi.org/10.1787/0b1414ed-en
Organisation for Economic Co-operation and Development (OECD). (2017a). *Programme for the International Assessment of Adult Competencies (PIAAC), Belgium log file. Data file version 2.0.0 [ZA6712_BE.data.zip]*. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017b). *Programme for the International Assessment of Adult Competencies (PIAAC), Denmark log file. Data file version 2.0.0 [ZA6712_DK.data.zip]*. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017c). *Programme for the International Assessment of Adult Competencies (PIAAC), Estonia log file. Data file version 2.0.0 [ZA6712_EE.data.zip]*. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017d). *Programme for the International Assessment of Adult Competencies (PIAAC), Finland log file. Data file version 2.0.0 [ZA6712_FI.data.zip]*. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017e). *Programme for the International Assessment of Adult Competencies (PIAAC), Ireland log file. Data file version 2.0.0 [ZA6712_IE.data.zip]*. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017f). *Programme for the International Assessment of Adult Competencies (PIAAC), Netherlands log file. Data file version 2.0.0 [ZA6712_NL.data.zip]*. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017g g). Programme for the International Assessment of Adult Competencies (PIAAC), Norway log file. Data file version 2.0.0 [ZA6712_NO.data.zip]. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017h h). Programme for the International Assessment of Adult Competencies (PIAAC), Poland log file. Data file version 2.0.0 [ZA6712_PL.data.zip]. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD). (2017i). *Programme for the International Assessment of Adult Competencies (PIAAC), Slovakia log file. Data file version 2.0.0 [ZA6712_SK.data.zip]*. Cologne: GESIS Data Archive. https://doi.org/10.4232/1.12955.
Organisation for Economic Co-operation and Development (OECD) (2017j). *Programme for the International Assessment of Adult Competencies (PIAAC), South Korea log file. Data file version 2.0.0 [ZA6712_KR.data.zip]*. GESIS Data Archive, Cologne. https://doi.org/10.4232/1.12955
Pirolli, P., & Card, S. (1999). Information foraging. *Psychological Review*, *106*(4), 643–675. https://doi.org/10.1037/0033-295X.106.4.643.
Rasmusson, M., & Åberg-Bengtsson, L. (2015). Does performance in Digital Reading relate to computer game playing? A study of factor structure and gender patterns in 15-Year-Olds' reading literacy performance. *Scandinavian Journal of Educational Research*, *59*(6), 691–709. https://doi.org/10.1080/00313831.2014.965795.
Reader, W. R., & Payne, S. J. (2007). Allocating Time Across multiple texts: Sampling and Satisficing. *Human–Computer Interaction*, *22*(3), 263–298. https://doi.org/10.1080/07370020701493376.
Rouet, J. F. (2003). What was I looking for? The influence of task specificity and prior knowledge on students' search strategies in hypertext. *Interacting with Computers*, *15*(3), 409–428.
Rouet, J. F., Britt, M. A., & Durik, A. M. (2017). RESOLV: Readers' representation of reading contexts and tasks. *Educational Psychologist*, *52*(3), 200–215. https://doi.org/10.1080/00461520.2017.1329015.

Salles, F., Dos Santos, R., & Keskpaik, S. (2020). When didactics meet data science: Process data analysis in large-scale mathematics assessment in France. *Large-Scale Assessments in Education*, *8*(1), 7. https://doi.org/10.1186/s40536-020-00085-y.

Salmerón, L., Strømsø, H. I., Kammerer, Y., Stadtler, M., & van den Broek, P. (2018). Comprehension processes in digital reading. In M. Barzillai, J. Thomson, S. Schroeder, & van den P. Broek (Eds.), *Learning to read in a digital world* (pp. 91–120). John Benjamins.

Singer, J., Braun, H., & Chudowsky, N. (2018). *International Education assessments: Cautions, Conundrums, and common sense*. National Academy of Education. https://doi.org/10.31094/2018/1.

Tóth, K., Rölke, H., Goldhammer, F., & Barkow, I. (2017). Educational process mining: New possibilities for understanding students' problem-solving skills. In B. Csapó & J. Funke (Eds.), *The Nature of Problem Solving: Using Research to Inspire 21st Century Learning* (pp. 193–209). OECD Publishing. https://doi.org/10.1787/9789264273955-14-en

Ulitzsch, E., He, Q., Ulitzsch, V., Molter, H., Nichterlein, A., Niedermeier, R., & Pohl, S. (2021). Combining Clickstream Analyses and Graph-Modeled Data Clustering for Identifying Common Response Processes. *Psychometrika*. https://doi.org/10.1007/s11336-020-09743-0

Ulitzsch, E., Ulitzsch, V., He, Q., & Lüdtke, O. (2022). A machine learning-based procedure for leveraging clickstream data to investigate early predictability of failure on interactive tasks. *Behavior Research Methods*. https://doi.org/10.3758/s13428-022-01844-1.

Vakkari, P. (2016). Searching as learning: A systematization based on literature. *Journal of Information Science*, *42*(1), 7–18. https://doi.org/10.1177/0165551515615833.

van den Broek, P., Bohn-Gettler, C., Kendeou, P., Carlson, S., & White, M. J. (2011). When a reader meets a text: The role of standards of coherence in reading comprehension. In M. McCrudden, J. Magliano, & G. Schraw (Eds.), *Text relevance and learning from text* (pp. 123–140). Information Age Publishing.

van der Kamp, M., & Boudard, E. (2003). Literacy proficiency of older adults. *International Journal of Educational Research*, *39*(3), 253–263. https://doi.org/10.1016/j.ijer.2004.04.006.

Warrington, K. L., McGowan, V. A., Paterson, K. B., & White, S. J. (2018). Effects of aging, word frequency, and text stimulus quality on reading across the adult lifespan: Evidence from eye movements. *Journal of Experimental Psychology: Learning Memory and Cognition*. https://doi.org/10.1037/xlm0000543.

Wineburg, S., Breakstone, J., McGrew, S., & Ortega, T. (2018). Why Google can't save us. In O. Zlatkin-Troitschanskaia, G. Wittum, & A. Dengel (Eds.), *Positive learning in the age of information* (pp. 221–228). Springer VS.

Wirth, W., Böcking, T., Karnowski, V., & von Pape, T. (2007). Heuristic and systematic use of search engines. *Journal of Computer-Mediated Communication*, *12*(3), 778–800. https://doi.org/10.1111/j.1083-6101.2007.00350.x.

Zhou, M. (2014). Gender difference in web search perceptions and behavior: Does it vary by task performance? *Computers & Education*, *78*, 174–184. https://doi.org/10.1016/j.compedu.2014.06.005.

Publisher's Note