Media and Communication
2025, Volume 13, Article 8988
https://doi.org/10.17645/mac.8988

Article

Developing performance tests to measure digital skills: Lessons learned from a cross-national perspective

Name Lastname [1], Name Lastname [2] and Name Lastname [3]

[1] Department, Institution, Country; Email: author1@email.com; ORCID: https://orcid.org

[2] Department, Institution, Country; Email: author2@email.com; ORCID: https://orcid.org

[3] Department, Institution, Country; Email: author3@email.com; ORCID: https://orcid.org

Correspondence: Name (author@email.com)

Submitted: 19 July 2024 | Accepted: 22 September 2024 | Published: in press

Issue: This article is part of the issue "Evaluating and Enhancing Media Literacy and Digital Skills" edited by Leen d'Haenens (KU Leuven) and Willem Joris (Vrije Universiteit Brussel), fully open access at https://doi.org/10.17645/mac.i466

Abstract

This article discusses the development of task-based performance tests designed to measure digital skills among children aged between 12 and 17 years old. The tasks reflect authentic everyday situations to evaluate skill levels. The primary objective is to design performance tests that provide a comprehensive understanding of children's digital skills. The tests cover three distinct skill dimensions: (1) information navigation and processing; (2) communication and interaction; and (3) content creation and production. These include several subdimensions, offering a detailed perspective on children's digital skills. The development process itself revealed several methodological challenges that needed to be addressed, yielding valuable lessons for future applications. Key lessons from our cross-national experiences include the importance of involving children early in the design process, using a combination of open-ended and closed tasks, and allocating ample time to walk through the coding scheme.

Keywords

digital skills; performance tests; cross-nationally applicable tasks; children

1. Introduction

Digital skills are indispensable for participation in an increasingly digital society. They are associated with a wide range of online opportunities, ranging from civic and social engagement to cultural, economic or health benefits (Cortesi et al., 2020; Livingstone et al., 2021; Rodríguez-de-Dios et al., 2018). Early conceptualisations focused mostly on technical operations (e.g., operating devices or using software) and information searching (e.g., defining keywords) (Bawden, 2001;

Kolle, 2017). The advent of Web 2.0 broadened this initial understanding to include skills required for online communication and interaction and the production of online content (Authors, 2021; Iordache et al., 2017; Siddiq et al., 2016; Authors, 2014). Despite these advancements in conceptualisations, many studies continue to employ limited operationalisations restricted to technical and information skills.

In addition to conceptualisation issues, recent literature reviews show that most measures use self-assessments, wherein children evaluate their proficiency across various digital skills (Haddon et al., 2020; Livingstone et al., 2021). Such self-assessments provide rough proxies for actual skill levels and require careful interpretation, as they are prone to social-desirability bias (Authors, 2021). Performance testing is considered as a more valid way to measure digital skills (Pagani et al., 2016; Authors, 2013). Such tests consist of tasks that require participants to perform an activity or construct a response (Claro et al., 2012), thereby offering closer approximations of digital skill levels (Aesaert & Van Braak, 2015). While performance testing is more common in controlled educational settings (Aesaert et al., 2014; Alkan & Meinck, 2016; Huggins et al., 2014), the number of studies that apply this method is relatively rare.

Existing performance tests have focused mainly on dimensions such as information search or evaluation (e.g., Bilal & Gwizdka, 2018; Frerejean et al., 2019; Nygren & Guath, 2019; Kaarakainen et al., 2019) and extended perspectives on assessments of digital skills as a broader concept are lacking (Siddiq et al., 2016; Authors, 2021). Additionally, studies using a task-based approach are often conducted on a small scale and cross-country comparisons are missing (Siddiq et al., 2016). Such comparisons provide a more robust basis for analysis and are essential to generalise conclusions (Gui & Argentin, 2011). To address this gap, research needs to critically reflect on performance testing as a method to measure a broad range of digital skills across various countries. This article aims to answer the following question: *What are suitable performance tests for obtaining an in-depth understanding of children's digital skills (referring to information navigation and processing, communication and interaction, and content creation and production) across different countries?*

The purpose of this study is to develop performance tests that can be implemented across European countries, facilitating cross-country comparisons. Data from these comparisons on digital skill levels are valuable to inform policymaking at both European and national levels, allowing for targeted interventions where most needed and providing indications for the effect of implemented national policies that promote digital skills. A critical first step toward expanding this type of measurement is to develop performance tests that can be applied internationally. Based on data collected from children aged 12 to 17 years in various European countries, the current contribution examines methodological issues in measuring digital skills through performance testing. The identified issues from all participating countries informed the development of the final performance tests. The lessons learned during the development process provide valuable guidance for future test application. The next section explores the conceptual framework underlying the performance tests, followed by an overview of existing digital skills measures.

2. Theoretical background

2.1 Digital skills conceptualisation

The development of performance tests was primarily guided by the youth Digital Skills Indicator (yDSI) (Authors, 2021) that proposes four digital skills dimensions: (1) technical and operational skills; (2) information navigation and processing

skills; (3) communication and interaction skills; and (4) content creation and production skills. The yDSI conceptualises both functional and critical aspects for each dimension. Functional aspects refer to the ability to use ICT functionalities, while critical aspects focus on understanding how and why content is produced in certain ways and what its impact might be. The measures for the four digital skills dimensions are grounded in a comprehensive review of both academic and grey literature that report on survey and performance test measures. The work of Haddon et al. (2020) and Cortesi et al. (2020) served as the basis for this review.

In the current contribution, the focus is on information navigation and processing, communication and interaction, and content creation and production skills. The tasks do not address technical skills directly as these are implicitly necessary to perform the other skills tasks. Information navigation and processing skills include navigation (e.g., searching information), interpretation (e.g., selecting information), and evaluation (e.g., verifying trustworthiness). Communication and interaction skills include affordances (referring to the design and features of digital technologies such as managing contacts), privacy (sharing information of self and others), and netiquette (understanding normative and non-discriminative behaviour). Content creation and production skills are conceptualised through affordances (e.g., using multimodality, which involves integrating elements like audio, images, video to enhance user engagement), content quality (e.g., attracting attention), and ownership (e.g., intellectual property).

2.2 Indirect measurements of digital skills

A considerable body of work relies on surveys to measure digital skills. One widely applied method involves asking respondents which online activities they have engaged in (Authors, 2014). While such proxies of usage are correlated with digital skills, they do not measure them directly (Authors, 2018). The limitation is that undertaking an activity (or not) does not mean that someone has (or lacks) the required skills (Haddon et al., 2020). Furthermore, accurately recalling the frequency of specific activities can be challenging. Another commonly used method is to measure respondents' self-efficacy (Aesaert & Van Braak, 2014). This gives an estimation of how proficient people think they are in various skills (Aesaert et al., 2017). Consequently, this approach measures an individual's confidence in their skills rather than actual skills.

Self-assessments in surveys are the most used method to measure digital skills (Allmann & Blank, 2021). This method is relatively straightforward and allows for the inclusion of many questions covering a wide range of skills. Combined with the ease of scoring, this approach facilitates large-scale, cross-national research. A disadvantage is that people struggle to accurately assess their own performance. Personal expectations of a satisfactory skill level and the reference group they compare themselves to influence their assessments (Talja, 2005). Consequently, such measures are sensitive to interpretation and judgment. Another disadvantage is the susceptible to social desirability bias. People tend to present themselves in a favourable manner relative to perceived social norms (King & Bruner, 2000). Specific demographic groups, such as men and younger individuals, are more likely to overestimate their skill levels compared to objective assessments (Aesaert et al., 2017; Palczyńska & Rynko, 2021; Porat et al., 2018). Consequently, conclusions drawn from self- assessments may suffer from severe validity problems.

2.3 Direct measurement of digital skills

Performance testing is a time- and labour-intensive process that relies on task completion to demonstrate skill levels. Assessments are based on directly observable performance, providing more reliable reflections of an individual's skill level (Jin et al., 2020). Scholars gather data on people's digital skills by analysing observable behaviour, such as task performance that require specific information (e.g., choosing key words) or strategies (e.g., using advanced search settings). Performance testing is, for instance, a widely used method for assessing online reading skills (see for example Castek et al., 2011; Coiro, 2011; Kiili & Leu, 2019). To some extent, approaches to test reading skills share similarities with assessments of information navigation and processing skills, as they focus on tasks aimed at measuring people's ability to locate, evaluate, and synthesize information online. However, tasks that assess skills related to social interaction and content creation and production skills, remain largely absent.

Existing studies have developed several types of performance tests. Some employ constrained response formats where participants interact with a test environment and select correct answers from provided options (e.g., Claro et al., 2012; Hatlevik & Christophersen, 2013). Others use software simulations of real-life ICT applications within a controlled environment where participants demonstrate their skills through simulation-based tasks (e.g., Fraillon & Ainley, 2010; Siddiq et al., 2017). However, biases may arise from participant's familiarity with the software (Fraillon, 2018). Additionally, designers face decisions about which aspects to simulate and which to omit (Engelhardt et al., 2021). Furthermore, these tests often involve a few relatively large tasks, where the testing situation can have a large impact on performance (Jin et al., 2020). Assessments employing interactive standardised tests offer insights into specific skill challenges contrasting with, for instance, multiple-choice tests.

Another type of performance testing involves participants engaging in real-life tasks within an open internet environment observed by researchers (e.g., Eshet-Alkali & Amichai-Hamburger, 2004; Litt, 2013). Participants apply skills to real-life situations and develop their own responses rather than selecting predetermined answers. The results provide insight into the specific skill problems experienced in authentic settings (Frerejean et al., 2019). Challenges include measuring multiple skills in a single test, devising tasks that are applicable across different countries, and developing a systematic coding scheme (Aesaert et al., 2014; Gui & Argentin, 2011). Although there is opportunity for in-depth measurement, the limited availability suggests that their full potential has yet to be realised (Siddiq et al., 2016). Details on the design, implementation, and analysis can serve as valuable guidance for future performance tests, enriching existing literature on digital skills measurements.

3. Method

3.1 Instrument design

This paper describes the development of performance tests to measure different dimensions of digital skills of children aged 12 to 17 years. Based on the detailed yDSI skill specifications, an initial version of performance tests featuring real-life tasks was developed. The choice of real-life tasks offered the advantage of allowing children to apply their digital skills in a realistic context. The task creation process was iterative, incorporating regular feedback from the research team and country partners involved in data collection. All children received the same set of tasks. Cognitive interviews and a pilot study were conducted to refine the test and make sure the tasks were age appropriate.

First, cognitive interviews were conducted with five children in the Netherlands and five children in the UK. Children were 12, 14, and 16 years old. A cognitive interview is a qualitative research method used to explore how people think and process information when answering questions or completing tasks (Willis, 2005). Children's feedback provided insights into the comprehensibility and difficulty of tasks for children across different ages and countries. Second, a pilot study involved 143 children from Estonia, Portugal, Belgium, and the Netherlands. See Table 1. For validity purposes, the selected sample was designed for diversity in gender and age groups. Estonia and Portugal held three classroom sessions within one school. Estonia sampled 6th grade children (mostly 12-year-olds), 8th grade children (mostly 14-year-olds) and 10th grade children (mostly 16-year-olds). The sample of Portugal consisted of 8th grade children (aged 12-13), 9th grade children (aged 14-15), and 12th grade children (aged 16-17). Belgium and the Netherlands together held 34 individual sessions. Upon completion of the cognitive interviews and pilot study, the instrument was evaluated carefully, leading to the final performance tests.

Table 1. Sample of the pilot study.

		Estonia		Portugal		Belgium/The Netherlands		Total	
		N	%	N	%	N	%	N	%
Gender	Boy	31	53	22	43	13	38	66	46
	Girl	25	43	29	57	21	62	75	52
	Other	2	3	0	0	0	0	2	1
Age	12-13	17	29	16	31	1	3	34	24
	14-15	23	40	17	33	10	29	50	35
	16-17	18	31	18	35	23	68	59	41
	N total	58		51		34		143	

Notes: Percentages do not add up to 100% due to rounding

3.2 Procedure

The pilot study of the performance tests was conducted in November 2020 in Estonia, Portugal, Belgium, and the Netherlands. Before starting the test, informed consent was obtained from all children and their caregivers. The test started with demographic questions followed by skill items (yDSI), which took approximately five minutes to complete in all countries. The tasks were performed on a computer or laptop with internet access and a program for creating slides (e.g., PowerPoint). The test took approximately 50 to 60 minutes.

Due to the COVID-19 pandemic, conducting performance tests in schools was not feasible in some countries. In such cases, tests were conducted individually at home, with the child monitored by a researcher via a video conferencing program that allowed screen sharing and recording. The researcher provided verbal instruction about the procedure and stayed connected with the child throughout the session, using a form to directly score several task performance indicators. In the classroom setting, children completed the test under the supervision of a teacher and trained researchers. A classroom was prepared to accommodate 15 to 20 children simultaneously, with necessary software for

screen recording and slide creation pre-installed on the computers. Scoring was performed afterwards based on video recordings. The schools were not informed about the specific content of the performance tests to prevent teachers from instructing children on specific digital skills before the testing.

3.3 The pilot performance tests

The development of the pilot performance tests was informed by the youth Digital Skills Indicator (yDSI), an extensively cross-nationally validated survey measurement. To ensure the tests' validity, we conducted consultations with experts (face validity), cognitive interviews (content validity), and pilot surveys (construct validity) with young people across various European countries. The survey items demonstrated both convergent and discriminant validity, indicating that the four skill dimensions are clearly distinct from one another and measure variety within each dimension. The content of the survey items was carefully converted into tasks to make sure the performance tests also effectively differentiate digital skills levels.

3.3.1 Information navigation and processing: Navigation, interpretation, and evaluation

The first part of the pilot tests involved four information navigation tasks focused on fact-based searches related to Netflix and dinosaurs. These tasks test the ability of children to search and select digital sources of information. Children were asked to use the internet and start their search by using a search engine of their choice. The following aspects were coded: (1) the keywords used, (2) the number of search attempts, (3) whether an evaluation of the answer occurred, and (4) whether the correct answer was found. The assessment was based on whether a correct answer was given. Additionally, children were asked to narrow their search to news articles within a designated timeframe, and the coding process verified whether this refinement was implemented.

In the second part, four social media posts in the categories of advertisement, phishing, news, and fake news were presented. This task relates to critical processing and evaluation of digital information sources, which required verifying the trustworthiness of information online. After each post, an open question was asked about its purpose. The coding scheme evaluated whether participants correctly identified the intent behind each post (commercial, scam, news, fake news).

3.3.2 Communication and interaction: Affordances, privacy, and netiquette

In the third part, children encountered a scenario where they received a message from an unfamiliar person inviting them to a party and requesting a photo. After the message, an open-ended question prompted children to consider how they would react. This task relates to affordances and tests the ability to react to unwanted online contact. The coding was based on whether the child would share a photo and the reasons behind their decision. Furthermore, children were presented two social media posts. The first showed a publicly shared telephone number, and the second a bikini photo shared only with friends. This task relates to online privacy and evaluates the child's awareness of appropriate sharing practices. The coding criteria assessed whether each post was considered appropriate considering the provided explanations. regarding the bikini photo, children could argue its appropriateness based on it being shared only with friends or its inappropriateness due to its revealing nature, even among friends.

In the fourth part, children were presented two WhatsApp conversations about climate change. This task relates to netiquette and involves the critical evaluation of how interpersonal mediated communication affects others. In each

chat, one person denies climate change, and the other supported its reality. In the second chat, the person who is arguing that climate change is problematic becomes insulting. After both chat screens, an open question prompted children to identify any problematic aspects in the conversation. The coding scheme scored whether the chat was problematic and the accompanying explanations. Only the second chat conversation with aggressive elements should have been considered problematic.

3.3.3 Content creation and production: Affordances, content quality, and ownership

The fifth part involved five tasks about content creation and production. The first task centred on strategies to make a GIF go viral when shared online with a broader audience. This task relates to content quality and tests the ability to attract attention and generate impact online. Successful strategies included using hashtags, sharing with friends, and requesting reposts. The second task focused on alternative ways of sharing a presentation beyond email, with correct answers involving programs for file sharing and cloud computing. In the third task, children were asked to improve a presentation slide. Examples of correct improvements were changing font type, reducing the amount of text, using colours, and adding visuals. In the fourth task, children were instructed to create and upload a new slide featuring an animal video. They were provided a link to a website offering free-to-use videos for both commercial and personal use. The task was scored based on their ability to (1) create a new slide, (2) insert an animal video, and (3) save and upload the file. The third and fourth task relate to affordances and test the ability to use multimodality. The final task involved selecting a copyright-free image containing a polar bear and melting ice. This task relates to ownership and test the ability to use online content covered by copyright. The scoring was based on whether a copyright-free image was uploaded.

3.4 The final performance tests

After carefully addressing the issues identified in the initial performance tests, an enhanced and final version was developed. Two more general changes were implemented. First, the test was divided into two modules. The first module focuses on information navigation and processing skills and content creation skills, and the second module focuses on communication and interaction skills. Second, there was a more balanced distribution of skills tasks. In the pilot, a relatively large amount of time was spent on information navigation and processing skills and on content creation skills. The number of similar tasks was reduced, allowing the inclusion of skill indicators not fully covered in the pilot.

The validation procedure included feedback from the research team and scholars from six country partners (Estonia, Finland, Germany, Italy, Poland, and Portugal). The final sample included countries that rank high, medium, and low on the Digital Economy and Society Index (DESI). This composite index is used by the European Commission to assess and compare the digital performance of European Union countries. Pilot testing involved small groups of two to three children in each country. The final performance test instrument is presented in the supplementary. The next section outlines specific adjustments made to the pilot test.

3.4.1 Module 1: Information navigation and processing skills

Changes were made to information navigation and processing skills by focusing all tasks on Greta Thunberg. The overarching theme of climate change was chosen for the entire test, reflecting its widespread discussion in schools across all participating countries. In the pilot test, the topic of Netflix turned out to be too centred on native English-speakers,

given the varying availability of information across countries where the service is used which meant that this was more a test of comfort with the English language than of information navigation and evaluation skills. Furthermore, a more straightforward coding process was implemented to make cross-national comparisons easier. For example, in the final test, children list the search queries they use for each search attempt. For the same reason, multiple-choice options were added for some questions. For example, the initial open question about the purpose of posts now includes predefined answer options. Answer options are also provided for the task in which children account for a specific time range in their search.

Furthermore, to ensure all skill indicators of the yDSI received adequate attention, additional tasks were simplified, and new skill indicators related to evaluation were incorporated. In the final test, children indicate which website they used to find the answer, select the most reliable website from a list of search results, and select what makes a website trustworthy from provided multiple-choice options. Finally, children are asked which of five existing websites available in all countries in the local language is least likely to provide reliable information about climate change.

3.4.2 Module 1: Content creation and production skills

For content creation and production skills, the slide improvement task changed. In the final test, children are required to create a slide focused on climate change, adhering to specific guidelines: using an image as a template, converting its colour to black and white, adding a title, listing three major causes of climate change in bullet points, and including a pollution-related video. Like in the pilot test, a 15-minute maximum limit was implemented. This restriction, coupled with clear task instructions, aims to provide better guidance to children during the test.

Furthermore, the task related to making content go viral was refined for better alignment with the test's theme and continuity. Children are asked to share their creation with as many people as possible. Rather than an open-ended question format, the task now presents options and asks to select the two options that make widespread dissemination most likely.

3.4.3 Module 2: Communication and interaction skills

Communication and interaction skills involve three parts: (1) receiving and sharing information with others; (2) interacting with others; and (3) intimate conversations with friends. In the first part, children are asked to identify which of four posts should not be shared without permission, aligning better with the test's overall theme and aiming to minimise ambiguity compared to the previous bikini photo task, as children could argue that it was either appropriate because it was only shared with friends or inappropriate since it was too revealing. The task involving a message from an unknown person has been revised to streamline responses and make the task more age appropriate (e.g., younger children do not get invited to parties). Instead of open questions yielding varied answers, children select the two most appropriate steps to take when a discussion turns nasty with sexist comments.

In part two, the task on how to contact friends is extended to better capture yDSI items. Children are now prompted to consider different scenarios—such as discussions with a teacher and classmates, close friends, or an expert—and select the most suitable medium for each. A task about Zoom settings during a session where a teacher is speaking has been introduced, both for the child him- or herself and others. Finally, a task on contacting an expert about COVID-19 via email is added.

In part three, the WhatsApp conversations changed. The fact that someone was a climate change denier proved to be controversial. This was seen as wrong by children and thus confused the results which were supposed to relate to recognizing when someone is bullied online and not the veracity of the content of messages. The new conversations therefore focus on a school project. Messages in the conversation are numbered and are referred to in answer options, allowing children to select inappropriate parts or choose the option 'none of them', thereby reducing cognitive demand.

4. Findings

This study focuses on developing performance tests that can be applied across various European countries to assess children's digital skills. The results show that our tests effectively differentiate between three dimensions of digital skills: information navigation and processing, communication and interaction, and content creation and production. For example, variations in performance between girls and boys were observed depending on the specific skill assessed. The performance tests are also used as teaching materials in class. The current contribution shows the lessons learned in developing performance tests to measure three dimensions of digital skills in different European countries. Findings from this study can be used to inform future test applications.

4.1 Designing performance tests

First, important to emphasise is that technical and operational skills underpin all tasks. Although we designed design tasks specifically oriented to information navigation and processing, communication and interaction, or content creation and production skills, it is not possible to rule out that all skills are to some extent needed to perform. An important lesson learned was the necessity of aligning topics with children's online experiences and lived realities to enhance their motivation in completing tasks. This study particularly focused on ensuring topics were suitable for a wide age range (12 to 17 years old) across various European countries. Choosing universal themes (e.g., climate change or COVID-19) ensured that search task topics are available internationally and applicable across age groups.

The design of a coding scheme is important to generate comparable results but proved to be a difficult endeavour for performance tests of digital skills. Issues arose in determining how to assess the quality of online search performance. To illustrate, a broad search query does not necessarily yield an incorrect answer, sparking debates over whether it was possible to develop objective criteria (e.g., specific keywords, number of search attempts) for successful task performance. Designing a coding scheme also required balancing the complexity of skill indicators and ease of use, especially for large-scale standardised skills assessments. It is important to allocate sufficient time for thorough training with the research team to ensure consistent understanding and application of the criteria across all evaluations.

This test used general survey software; unlike tests designed in a closed test environment, no technical expertise was needed to develop a platform that simulates real-world ICT applications. A disadvantage of performance tests in an open internet environment is the influence of search engine results on skill-related actions. Search engine results can vary based on personalized algorithms, making it more difficult to ensure consistent and reliable measurement of digital skills across individuals.

Additionally, skills related to specific apps or platforms may not always be transferable; for instance, search result filtering settings vary across search engines. Furthermore, not every participant uses the same apps or platforms,

and the popularity of these tools can vary significantly between countries. A lesson learned was to let participants choose their preferred search engine when answering fact-based questions.

Designing tasks for communication and interaction, as well as content creation and production skills, proved challenging due to their context-specific nature and reliance on situational relevance. Context helps to resolve ambiguities and ensure consistent measures, especially in cross-national performance tests. The difficulty lies in how to make it as realistic as possible in an open internet environment without programming a platform or a social media timeline. A lesson learned was to involve children early in the process and take children's level of understanding and experience as a starting point. For instance, initial chat message designs by researchers did not always reflect typical peer conversations as perceived by the children, highlighting the need for adjustments. Communication skill tasks often result in scenario-based questions to capture the interaction element. Generally, balancing real-life authenticity with research control is inherently challenging when developing performance tests. Tasks completed in an open internet environment are authentic but lack control over the differences in internet resources and other confounding factors. Although the developed tasks try to replicate real-life scenarios, their validity depends on whether they are realistic and well designed by the researchers.

4.2 Implementing performance tests

The concept of digital skills is broad, making it challenging to design a test that comprehensively assesses all skill dimensions. Because the administration of tasks takes time, it is not feasible to measure all skill dimensions in one performance test. Additionally, performance testing is cognitively demanding, particularly for children, as sustained attention may diminish if tasks are overly time-consuming. Both the complexity and completion time of the test are important to carefully manage. Tests with no time limits bear the risk that some participants spend too much time on certain tasks. In the current study, performance testing could not take longer than one school hour, limiting how extensively each skill can be measured.

Before implementing performance tests, it is important to hold expert consultations and cognitive interviews with the participant group. Designing information navigation tasks – which we expected to be relatively easy– proved to be difficult because solutions needed to be available in the native language of all participating countries, yet not too easily found in the search results. Various rounds of adjustments were necessary to measure information navigation skills cross-nationally. Expert reviews identified potential weaknesses in task instructions, while cognitive interviews provided insights into children's thought processes. These interviews revealed how children react and reason, improving performance tests. For example, while children understood the purpose of the chat messages, they pointed out that these texts did not reflect how a conversation between peers usually goes in real life. A key lesson was to use cognitive interviews (in addition to an expert round) to understand task interpretation and the need to conduct these interviews in all countries involved for unique perspectives.

In general, explicit instructions are critical for children, reducing the cognitive load of processing information. A lesson learned was to split two-pronged questions (for example, by letting the child answer first if he or she would send a photo and then asking to provide the explanation). Last, an unforeseen challenge was the quality of internet connections at schools, causing difficulties like uploading presentations, despite the availability of computers with internet access.

4.3 Analysing performance tests

Performance testing is time- and labour-intensive resulting in small sample sizes. One solution is to integrate additional questions and let the participant do some coding. For instance, ask the child to list the search terms used. Although it saves effort and time for the researcher, it is more demanding for the child. To balance this, a combination of open-ended and closed tasks was used.

Coding of the performance tests is also labour-intensive. In tasks related to communication and interaction skills, the correct answers to tasks are often subject to interpretation, underscoring the importance of pretesting performance tests within each participating country. For example, in our study, the participating European countries deemed it correct to have cameras on during online classroom conversations. However, cultural differences might influence this view as turning cameras on could be seen as controversial. Additionally, the 'other' option was often selected, indicating a need for more detailed guidelines. Open-ended questions, while adding depth to the test, yielded wide-ranging responses, suggesting extensive testing to anticipate possible answers. A drawback of providing more options is that children might not have considered these options themselves and the test in this format might teach them about these rather than test their existing knowledge. Nevertheless, providing precoded categories appeared valuable when working cross-nationally, though leaving an open category for unexpected answers is also essential.

Finally, tasks should focus on a single action, ensuring dependencies between tasks are minimised. For example, the inability to find a copyright-free image should not prevent participants from doing an uploading task. Another lesson was to restrict the number of coders per country to one or two and ensure that all coders are trained before starting the analysis.

5. Conclusion

Ongoing debates exist about the exact dimensions of digital skills and how they should be measured. Scholars generally agree that digital skills are multidimensional (Jin et al., 2020). However, little is known about how to measure a broader range of digital skills through performance testing, especially in cross-national studies involving children. This study addresses test development and application procedures to improve the performance test quality. By developing and cross-nationally testing compatible tasks, we tackled specific issues in performance test development beyond the known challenges of being time- and labour-intensive.

Our study expands knowledge on how to design effective performance tests, encouraging other researchers to assess digital skills directly. Carefully designed tests measure the actual behaviours and real-life technology engagement, providing a valid assessment of digital skills free from self-assessment biases (Aesaert & Van Braak, 2015; Pagani et al., 2016). These developed tests can be used by other researchers to assess digital skills, covering a broader range of dimensions such as information navigation, communication, and content creation. However, important areas to consider are the constraints of various types of performance tests and the associated coding and analysis procedures.

Acknowledgments

Add here.

Funding

Add here.

Conflict of Interests

Add here.

Data Availability

Add here.

Supplementary Material

Supplementary material for this article is available online in the format provided by the author (unedited).

References

Aesaert, K., & Van Braak, J. (2014). Exploring factors related to primary school pupils' ICT self-efficacy: A multilevel approach. *Computers in Human Behavior*, *41*, 327-341. https://doi.org/10.1016/j.chb.2014.10.006

Aesaert, K., & Van Braak, J. (2015). Gender and socioeconomic related differences in performance based ICT competences. *Computers & Education*, *84*, 8–25. https://doi.org/10.1016/j.compedu.2014.12.017

Aesaert, K., Van Nijlen, D., Vanderlinde, R., & Van Braak, J. (2014). Direct measures of digital information processing and communication skills in primary education: Using item response theory for the development and validation of an ICT competence scale. *Computers & Education*, *76*, 168–181. https://doi.org/10.1016/j.compedu.2014.03.013

Aesaert, K., Voogt, J., Kuiper, E., & van Braak, J. (2017). Accuracy and bias of ICT self-efficacy: An empirical study into students' over-and underestimation of their ICT competences. *Computers in Human Behavior*, *75*, 92–102. https://10.1186/s40536-016-0029-z

Alkan, M., & Meinck, S. (2016). The relationship between students' use of ICT for social communication and their computer and information literacy. *Large-Scale Assessments in Education*, *4*(1), 1–17. https://10.1186/s40536-016-0029-z

Allmann, K., & Blank, G. (2021). Rethinking digital skills in the era of compulsory computing: Methods, measurement, policy and theory. *Information, Communication & Society*, *24*(5), 633–648. https://doi.org/10.1080/1369118X.2021.1874475

Bawden, D. (2001). Information and digital literacies: A review of concepts. *Journal of Documentation*, *57*(2), 218–259. https://doi.org/10.1108/EUM0000000007083

Bilal, D., & Gwizdka, J. (2018). Children's query types and reformulations in Google search. *Information Processing & Management*, *54*(6), 1022–1041. https://doi.org/10.1016/j.ipm.2018.06.008

Castek, J., Zawilinski, L., McVerry, G., O'Byrne, I., & Leu, D. J. (2011). The new literacies of online reading comprehension: New opportunities and challenges for students with learning difficulties. In C. Wyatt-Smith, J. Elkins, & S. Gunn (Eds.), *Multiple perspectives on difficulties in learning literacy and numeracy* (pp. 91–110). New York: Springer.

Claro, M., Preiss, D.D., San Martín, E., Jara, I., Hinostroza, J.E., Valenzuela, S., Cortes, F., & Nussbaum, M. (2012). Assessment of 21st century ICT skills in Chile: Test design and results from high school level students. *Computers & Education*, *59*(3), 1042–1053. https://doi.org/10.1016/j.compedu.2012.04.004

Coiro, J. (2011). Predicting reading comprehension on the Internet: Contributions of offline reading skills, online reading skills, and prior knowledge. *Journal of literacy research*, *43*(4), 352–392. https://doi.org/10.1177/1086296X11421979

Cortesi, S., Hasse, A., Lombana, A., Kim, S., & Gasser, U. (2020). *Youth and digital citizenship+ (plus): Understanding skills for a digital world*. Berkman Klein Center Research Publication No. 2020-2. https://doi:10.2139/ssrn.3557518.

Engelhardt, L., Naumann, J., Goldhammer, F., Frey, A., Horz, H., Hartig, K., & Wenzel, S.F.C. (2021). Development and evaluation of a framework for the performance-based testing of ICT skills. *Frontiers in Education*. https://doi.org/10.3389/feduc.2021.668860

Eshet-Alkali, Y., & Amichai-Hamburger, Y. (2004). Experiments in digital literacy. *CyberPsychology & Behavior*, 7(4), 421–429.

Fraillon, J. (2018). International large-scale computer-based studies on information technology literacy in education. In: J. Voogt, G. Knezek, R. Christensen, & K.-W. Lai (Eds.), *Second handbook of information technology in primary and secondary education* (1161–1179). Cham, Switzerland: Springer.

Fraillon, J., & Ainley, J. (2010). *The IEA international study of computer and information literacy (ICILS)*. Melbourne, AU: Australian Council for Educational Research. Available at: http://www.iie.cl/wp-content/uploads/2013/11/2013-007-AE-ICILS-2013-Descripcion_detallada_del_proyecto.pdf

Frerejean, J., Velthorst, G. J., van Strien, J. L., Kirschner, P. A., & Brand-Gruwel, S. (2019). Embedded instruction to learn information problem solving: Effects of a whole task approach. *Computers in Human Behavior*, *90*, 117–130. https://doi.org/10.1016/j.chb.2018.08.043

Gui, M., & Argentin, G. (2011). Digital skills of internet natives: Different forms of digital literacy in a random sample of northern Italian high school students. *New Media & Society*, *13*(6), 963–980. https://doi.org/10.1177/1461444810389751

Haddon, L., Cino, D., Doyle, M.A., Livingstone, S., Mascheroni, G., & Stoilova, M. (2020). *Children's and young people's digital skills: A systematic evidence review*. Leuven, KU Leuven: ySKILLS. Available at: https://zenodo.org/records/4274654#.X-pMceSWysc

Hatlevik, O.E., & Christophersen, K.A. (2013). Digital competence at the beginning of upper secondary school: Identifying factors explaining digital inclusion. *Computers & Education*, *63*, 240–247.

Hinostroza, J. E., Ibieta, A., Labbé, C., & Soto, M. T. (2018). Browsing the internet to solve information problems: A study of students' search actions and behaviours using a 'think aloud'protocol. Education and Information Technologies, *23*, 1933–1953. https://doi.org/10.1007/s10639-018-9698-2

Huggins, A. C., Ritzhaupt, A. D., & Dawson, K. (2014). Measuring information and communication technology literacy using a performance assessment: Validation of the student tool for technology literacy (ST2L). *Computers & Education*, *77*, 1–12. https://doi.org/10.1016/j.compedu.2014.04.005

Iordache, C., Mariën, I., & Baelden, D. (2017). Developing digital skills and competences: A quick-scan analysis of 13 digital literacy models. *Italian Journal of Sociology of Education*, *9*(1), 6–30. https://doi.org/10.14658/pupj-ijse-2017-1-2

Jin, K.Y., Reichert, F., Cagasan Jr, L.P., de la Torre, J., & Law, N. (2020). Measuring digital literacy across three age cohorts: Exploring test dimensionality and performance differences. *Computers & Education*, 157, 103968. https://doi.org/10.1016/j.compedu.2020.103968

Kaarakainen, M. T., Saikkonen, L., & Savela, J. (2019). Information skills of Finnish basic and secondary education students: The role of age, gender, education level, self-efficacy and technology usage. *Nordic Journal of Digital Literacy*, *13*(4), 56–72.

Kiili, C., & Leu, D. J. (2019). Exploring the collaborative synthesis of information during online reading. *Computers in Human Behavior*, *95*, 146–157. https://doi.org/10.1016/j.chb.2019.01.033

King, M.F., & Bruner, G.C. (2000). Social desirability bias: A neglected aspect of validity testing.*Psychology & Marketing*, *17*(2), 79–103.

Kolle, S. R. (2017). Global research on information literacy: A bibliometric analysis from 2005 to 2014. *The Electronic Library*, *35*(2), 283–298.

Litt, E. (2013). Measuring users' internet skills: A review of past assessments and a look toward the future. *New Media & Society*, *15*(4), 612–630. https://doi.org/10.1177/1461444813475424

Livingstone, S., Mascheroni, G., & Stoilova, M. (2021). The outcomes of gaining digital skills for young people's lives and wellbeing: A systematic evidence review. *New Media & Society*, *25*(5), 1176–1202. https://doi.org/10.1177/14614448211043189

Nygren, T., & Guath, M. (2019). Swedish teenagers' difficulties and abilities to determine digital news credibility. *Nordicom Review*, *40*(1), 23–42.

Pagani, L., Argentin, G., Gui, M., & Stanca, L. (2016). The impact of digital skills on educational outcomes: Evidence from performance tests. *Educational Studies*, *42*(2), 137–162. https://doi.org/10.1080/03055698.2016.1148588

Palczyńska, M., & Rynko, M. (2021). ICT skills measurement in social surveys: Can we trust self-reports? *Quality & Quantity*, *55*(3), 917–943.

Porat, E., Blau, I., & Barak, A. (2018). Measuring digital literacies: Junior high-school students' perceived competencies versus actual performance. *Computers & Education*, *126*, 23–36. https://doi.org/10.1016/j.compedu.2018.06.030

Rodríguez-de-Dios, I., Van Oosten, J.M., & Igartua, J.J. (2018). A study of the relationship between parental mediation and adolescents' digital skills, online risks and online opportunities. *Computers in Human Behavior*, *82*, 186–198. https://doi.org/10.1016/j.chb.2018.01.012

Siddiq, F., Gochyyev, P., & Wilson, M. (2017). Learning in Digital Networks–ICT literacy: A novel assessment of students' 21st century skills. *Computers & Education*, *109*, 11–37. https://doi.org/10.1016/j.compedu.2017.01.014

Siddiq, F., Hatlevik, O.E., Olsen, R.V., Throndsen, I., & Scherer, R. (2016). Taking a future perspective by learning

from the past: A systematic review of assessment instruments that aim to measure primary and secondary school students' ICT literacy. *Educational Research Review*, *19*, 58–84. https://doi.org/10.1016/j.edurev.2016.05.002

Talja, S. (2005). The social and discursive construction of computing skills. *Journal of the American Society for Information Science and Technology*, *56*(1), 13–22. https://doi.org/10.1002/asi.20091

Willis, G. B. (2005). *Cognitive Interviewing: A tool for improving questionnaire design*. Thousand Oaks, CA: Sage Publications.

About the Authors

Photo and Biography