Computers & Education 203 (2023) 104830

Contents lists available at ScienceDirect

Computers & Education

journal homepage: www.elsevier.com/locate/compedu

Digital competence test for learning in schools: Development of items and scales

Margus Pedaste[a], Külli Kallas[a,*], Aleksandar Baucal[b]

[a] Institute of Education, University of Tartu, Jakobi 5, 51005, Tartu, Estonia
[b] Department of Psychology, University of Belgrade, Cika Ljubina 18-20, 11000, Belgrade, Serbia

ARTICLE INFO

Keywords:
Secondary education
Elementary education
Evaluation methodologies
Distance education and online learning
Information literacy

ABSTRACT

Digital competence is one of the key competences in modern society. COVID-19 related remote schooling revealed that the level of digital competence for learning is not at an expected level. While schools try to overcome this issue, there is a need to understand the level of students' digital competence by assessing it. However, previous attempts to design a comprehensive digital competence test have not been entirely successful. One of the reasons might be that the focus of these tests has been too general. Therefore, the aim of this study was to create a test for assessing primary and lower secondary school students' digital competence for learning in ten dimensions identified based on earlier studies. This digital competence for learning assessment test (Digitest) was carried out with 836 third to ninth grade students from Estonian schools. IRT analysis and confirmatory factor analysis were conducted to establish the potential of the test to measure digital competence for learning and discover which latent variables can be differentiated with it. The results showed that nine dimensions describing attitudes, skills and behaviours can be distinguished with the Digitest and that the test items have good fit to assess digital competence for learning. However, the study also revealed that two higher-order dimensions of digital competence for learning could be identified: motivational and cognitive-behavioural. Thus, empirical data collected with the Digitest is initiating discussions for assessing students' digital competence in a holistic way but also more general discussions on the concept of competence. This can help educators put greater emphasis on areas where students need further improvement.

1. Introduction

Digital technology is constantly being developed further, and it is bringing us more opportunities and freedom than we have ever had. Being digitally competent means that we can understand media, search and critically evaluate information and communicate with others through a variety of different means (Ferrari, 2012). More specifically, "digital competence (DC) is the set of knowledge, skills, attitudes (thus including abilities, strategies, values and awareness) that are required when using ICT and digital media to perform tasks; solve problems; communicate; manage information; collaborate; create and share content; and build knowledge effectively, efficiently, appropriately, critically, creatively, autonomously, flexibly, ethically, reflectively for work, leisure, participation, learning, socialising, consuming, and empowerment" (Ferrari, 2012, p. 3). As being digitally competent is not something we are born with, education (along with other factors, such as family background (Hatlevik, Ottestad, & Throndsen, 2015) plays a crucial role in shaping

* Corresponding author. Institute of Education, University of Tartu, Jakobi 5, 51005, Tartu, Estonia.
E-mail addresses: margus.pedaste@ut.ee (M. Pedaste), kulli.kallas@ut.ee (K. Kallas), abaucal@f.bg.ac.rs (A. Baucal).

https://doi.org/10.1016/j.compedu.2023.104830
Received 15 January 2022; Received in revised form 4 May 2023; Accepted 7 May 2023
Available online 26 May 2023
0360-1315/

students' digital competence that is needed to leverage their learning process. In the light of that, DC is classified by the European Commission as one of the eight key competences for lifelong learning in European schools (Key competences for lifelong learning in the European schools, 2018). Moreover, some European countries have included DC in their national curricula. For example, the Estonian national curriculum states that DC is one of the eight key competences that the schools in Estonia focus on (National curriculum for basic schools, 2021). In the Nordic countries (Denmark, Finland, Sweden, Norway), DC is also present in national curricula through different concepts, such as "IT and Media" in Denmark, "Multiliteracies and DC" in Finland, "Digital skills" in Norway and DC in Sweden (Godhe, 2019). What these examples illustrate is that authors agree on the importance of developing students' DC. At the same time, these curricula focus on DC necessary in all spheres of life and not specifically on digital competence for learning (DCL). This might be the reason why it has been difficult to reach a common understanding about what DC is and how it could be measured (see Weinert, 2001, for a discussion on different approaches to the concept of competence in the educational context). In this study we want to place emphasis on DC required in school and define DCL as digital competence required in learning situations.

Recently, the COVID-19 pandemic related remote schooling experience revealed that knowledge of using digital resources is rather limited in education (Carretero, Napierala, Bessios, Mägi, Pugacewicz, Ranieri et al., 2021) and therefore, DC needed for carrying out effective learning (meaning: DCL) are not on an expected level. However, assessing DC or DCL for developmental purposes comprehensively is a subject of debate within the scientific community. Previous attempts to assess DC or DCL have not provided us with a good solution. According to various literature reviews, the existing instruments often focus on only some aspects of the competence (e.g., the unidimensional instrument of Jin, Reichert, Cagasan, de la Torre, & Law, 2020) or use self-assessment, which might be unreliable (see Laanepere, 2019; Saltos-Rivas, Novoa-Hernández, & Serrano Rodríguez, 2022; Siddiq, Hatlevik, Olsen, Throndsen, & Scherer, 2016). Assessment of DC or DCL seems to be a particularly understudied area at the level of primary and lower secondary school (see Siddiq et al., 2016); in this age group, self-assessments might be even more unreliable (see Palczyńska & Rynko, 2021). Ford, Kozey, and Negreiros (2012) also emphasize that assessment instruments also need to consider the age group of the learners. Therefore, we focus more specifically on developing an instrument for assessing DCL in primary and lower secondary schools. DCL could be understood as DC in a specific context, and the narrower concept and age group allow us to expect a higher possibility of validating the assessment instrument empirically. To the best of our knowledge, studies undertaken so far do not provide a valid method to assess this age group's DCL as a multidimensional construct covering not only knowledge and skills but also attitudinal and behavioural aspects in the learning context. Without assessment, however, it is not possible to determine whether the learning goals in schools are being met and educators lack a reliable way of knowing if students are learning what they are supposed to be learning. Thus, we aimed to develop and psychometrically validate a DCL assessment test for primary and lower secondary school students, an age group heavily affected by the change of remote schooling needs during the COVID-19 pandemic and, to the best of our knowledge, lacking a reliable multidimensional contextualized assessment instrument. With this study we hope to contribute to ongoing discussions about the dimensionality of digital competence and introduce ideas about reframing the dimensions of digital competence.

1.1. *Defining digital competence*

When defining key competences, the Organisation for Economic Co-operation and Development (OECD) also defined the concept of competence as being "more than just knowledge and skills. It involves the ability to meet complex demands, by drawing on and mobilising psychosocial resources (including skills and attitudes) in a particular context" (The Organisation for Economic Co-operation and Development, 2005, p. 4). There are many other publications that also emphasize the broader view of competences (see for an overview Le Deist & Winterton, 2005; Weinert, 2001). In our study we rely on Weinert's (2001) overview of the concept of competence, where he introduces several approaches to the concept. We view digital competence as an example of a specialized competence, but not only as a specialized cognitive competence but a competence where cognitive aspects are related to motivational and behavioural ones. More specifically, we anticipate, as Weinert suggests in a general context, that for meaningful application of digital technologies for learning it is important to apply (behavioural aspect, action tendencies as defined by White, 1959) both cognitive (skills and knowledge, abilities to do something) and motivational (willingness to do something, or action need as framed by White, 1959) dimensions of competence. Similarly, Schneider (2019) argues based on Weinert's studies that competence could be described through cognitive and motivational elements and that both of these could consist of several dimensions: e.g., cognitive abilities, skills, knowledge, strategies, routines necessary for mastering specific demands, expectations and performance criteria (specific performance disposition) could be classified as cognitive elements; and the subjective estimation of personal performance resources and related motivational action tendencies could be classified as motivational elements. In addition, some researchers (e.g., Key Competences for Lifelong Learning in the European Schools, 2018) also suggest that these higher-level competence constructs introduced by Weinert are key competences and meta-competences, meaning that the former are transferable (e.g., language competences needed in every learning situation) and the latter are facilitating the learning process (strategies of thinking, learning, planning). In our study we did not focus on key competences or meta-competences but rather on more content- or context-specific competences needed in the digital (or digitally supported/enabled) learning process.

Based on a literature review, Schneider (2019) identified even more general dimensions of competence that were described as incompatible with each other: behaviour vs. ability, inherited vs. learned, observable vs. non-observable, motivational vs. cognitive, quality vs. state, specific disposition vs. general disposition, cognitivist construct vs. behaviouristic construct, objectivist view vs. constructivist view, construct vs. non-construct. Several of these could also be applicable in the context of digital competence for learning. Thus, it seems that in general, one could identify two higher-order factors and many more specific factors describing a competence. However, it is not so clear how these higher-order competence dimensions could be defined. Therefore, our research is also attempting to find out if these dimensions could be categorized as motivational dimensions and cognitive dimensions or if one could even distinguish a third higher-order factor describing the behavioural dimensions of a competence.

More specifically, in our study we aimed to design an assessment instrument suitable for distinguishing not only cognitive (knowledge and skills) but also motivational (attitudes) and behavioural aspects (action tendencies) of the digital competence for learning, a competence which refers to not only knowledge of how to use digital devices but also attitudes toward them (that guide our behavioural decisions). This is in line with the study of Ilomäki, Paavola, Lakkala, and Kantosalo (2016, p. 671), who reported that digital competence (often also referred to as "digital literacy") consists of "the skills and practices required to use new technologies in a meaningful way and as a tool for learning, working and leisure time, understanding the essential phenomena of digital technologies in society as well as in one's own life, and the motivation to participate in the digital world as an active and responsible actor". According to these definitions and following Weinert's (2001) notion of specialized (cognitive) competences, motivational action tendencies and action competences, DCL could be defined as *the attitudes, skills and practices required for using digital technologies in a meaningful way for learning*.

DCL as a specialized competence has not been widely studied and therefore we also rely on DC in the development of the concept and assessment instrument for DCL. DC is defined as one of the key competences in learning. The European Commission carried out the DIGCOMP project for further identifying the key components of DC, resulting in the DigComp 2.0 framework (based on the previous framework by Ferrari, 2012, 2013) describing five areas of DC: information and data literacy, communication and collaboration, digital content creation, safety, and problem solving (Vuorikari, Punie, Carretero, & Van Den Brande, 2016). The newer version of the framework is DigComp 2.1, which has the same five competence areas but breaks them down into 21 sub-areas (Claro et al., 2012). In 2022 DigComp 2.2 was published, which added examples of new emerging themes to the framework, including separate attention to skills, knowledge and attitudes, allowing deeper understanding of the DC concept and bringing in the importance of attitudes in DC (Vuorikari, Kluzer, & Punie, 2022). DigComp can be used to assess one's own DC and, at the same time, DC assessment tests are also created on the basis of DigComp dimensions (Vuorikari et al., 2016). The latest version of the DigComp 2.2 framework has not yet been used for developing assessment instruments, although examples of relevant skills, knowledge and attitudes on different dimensions could be used for that purpose. In the meantime, the instruments created have usually been focusing on self-assessment and older age groups that might be more skilled at self-assessment and reflection: e.g., the DigCompSAT project developed a tool that was tested in the 16–65-year-olds' age group (see Clifford, Kluzer, Troia, Jakobsone, Zandbergs, Vuorikari et al., 2020). It seems that assessing the DC of primary school and lower secondary school students is still an area to be discovered in further studies. Therefore, in our study we focused on developing a DCL test for younger learners.

1.2. Assessing digital competence for learning

As stated earlier, assessing DC or DCL is still a topic of discussion among researchers because many assessment instruments (1) rely on self-assessment, (2) do not provide a way to follow students' skills, knowledge and attitudes throughout the school years, (3) have limitations in giving a comprehensive overview of the respondents' competences in different dimensions, and (4) are focusing on too narrow context.

First, the assessment often relies on self-assessment, usually on Likert-type scales (e.g., Chou & Chiu, 2020; Kong, Wang, & Lai, 2019; Lee, Chen, Li, & Lin, 2015; Yang, Tlili, Huang, Zhuang, & Bhagat, 2021; Zhang & Zhu, 2016). This, however, does not seem to be the ideal option: as Siddiq et al. (2016) noted in their review of assessment instruments, in the majority of the studies, students' self-reported levels of competence were not in line with their actual performance. The same was concluded by Porat, Blau, and Barak (2018) in their study with junior high school students where the researchers reported that in the educational context, confidence in one's digital abilities is only partially associated with students' actual performance of digital tasks. Palczyńska and Rynko (2021) also reported in their more general study on investigating the trustworthiness of self-reports in information and communication technology (ICT) skills measurement that people with higher ICT skills are more likely to overreport having these skills. The authors also found that overreporting one's skills decreases with age (Palczyńska & Rynko, 2021), which is another indicator that relying on self-assessment in primary and secondary school may not be reliable. The same was concluded by Jin et al. (2020), who suggested that performance-based digital skills assessment would more likely provide accurate data. Furthermore, as competences develop mostly through real-life experiences, the best way for assessing them could be with the help of assignments drawn from real-life situations (Reichert, Zhang, Law, Wong, & de la Torre, 2020; Siddiq et al., 2016).

The second challenge we have noted is that there are not many studies on the assessment of DC or DCL that focus on younger students. The aforementioned DigCompSAT project focused on the group of people older than 16. Siddiq et al. (2016) investigated 66 publications reporting DC or DCL assessment instruments and only nine of them were focused on children below the 6th grade. More international studies have been conducted in order to get a more wholesome picture of international ICT and digital literacy levels (e.g., Framework for ICT Literacy (ETS) in 2006; International Computer and Information Literacy Study (ICILS) 2013 and 2018; Program for International Student Assessment (PISA) in 2018, see from Chung & Yoo, 2021); however, these are also focusing on assessing students in upper secondary education, leaving out primary and lower secondary school students.

The third challenge seems to involve all areas of DC or DCL – the studies tend to focus on the assessment of only skills or abilities (e.g., Aesaert, van Nijlen, Vanderlinde, & van Braak, 2014; Jun, Han, Kim, & Lee, 2014; Kiili et al., 2018; Key Competences for Lifelong Learning in the European Schools, 2018udes (Asil, Teo, & Noyes, 2014) or leave out some specific dimensions of the competence (e.g., Heitink, 2018). In relation to that, it is still not clear in the scientific community as to whether the concept of DC or DCL is unidimensional or comprising several dimensions. Many authors agree that digital literacy (or in our case, DC or DCL) is rather unidimensional (Aesaert et al., 2014; Claro, Preiss, San; Jin et al., 2020; Kim, Ahn, & Kim, 2019a, 2019b; Martin, Jara, Hinostroza, Cortes et al., 2012; Reichert et al., 2020) whereas other authors conclude that in digital literacy there are more underlying dimensions

(Guitert, Romeu, & Baztán, 2021; Porat et al., 2018; Siddiq, Gochyyev, & Wilson, 2017; Wilson, Gochyyev, & Scalise, 2017). At the same time, it has also been noted that in the case of multidimensional structures, the dimensions are often highly correlated (Ihme, Senkbeil, Goldhammer, & Gerick, 2017; Jin et al., 2020; Siddiq et al., 2017), so we can argue whether we should be presenting the concept as unidimensional or there is indeed a logic behind the strong internal connections, as the concept's different dimensions are naturally connected to each other. Furthermore, performance estimation for multiple dimensions of DC or DCL provides more meaningful feedback for educators, allowing them to plan more focused classroom interventions (Wilson, Scalise, & Gochyyev, 2019). For the above-mentioned reasons, it seems more appropriate to prefer ways to assess students' DC or DCL that – rather than focusing merely on overall competence – allow us to separately evaluate students' performance in different areas.

The fourth topic to mention is that sometimes the instruments focusing on both skills and attitudes are developed for assessment in a very specific context: e.g., Fernández-Montalvo, Peñalva, Irazabal, and López (2017) focused on investigating the effectiveness of a digital literacy programme with particular attention paid to internet use and risks focusing on skills, knowledge and also attitudes; however, the nature of the study resulted in having a very specific target group. In our study, we also focused on a specific context, but it is less narrow and could be extended to all learning contexts.

In conclusion, as also reported by Siddiq et al. (2016) and Godaert, Aesaert, Voogt, and van Braak (2022) in their systematic literature reviews, developing a valid instrument for assessing primary school students' digital competence in its multidimensionality and using performance assessment in addition to self-evaluations is a challenge that we take at the core of the current study. Having such an instrument is especially important if we also aim to monitor the development of those abilities throughout the school years.

Attempts towards development of this type of instrument – and specifically with a focus on DCL – have been made by (Pedaste et al., 2021), who analysed Estonian data from the international EU Kids Online self-report survey (Smahel et al., 2015; Sukk & Soo, 2018) and the Estonian National Digital Competence Test focusing on performance assessment (created on the basis of the DigComp framework). The suggestions of the study by Pedaste et al. (2021) for revising the framework for the assessment of digital competence at primary school and lower-secondary school level were taken into consideration in our study.

In conclusion, it seems that after several decades of conducting studies on the topic, the research community is still at a point where we do have room for developing a valid instrument for assessing students' DC, or even more specifically DCL, in a comprehensive multidimensional way, keeping in mind common psychometric properties and focusing on primary and lower-secondary school students. Hence, the main goal of our study was to create a test (called "Digitest") for assessing primary and lower secondary school students' digital competence for learning in ten dimensions and to evaluate the quality of the test. Moreover, assessing DCL (also in the long term) allows schools and policy-makers to make informed decisions to ensure the continuous development of students' digital knowledge, skills and attitudes. For researchers, the current study can provide further evidence on the dimensionality of DCL and fill a research gap in how to assess it in a meaningful way.

The key research question of this study was whether the Digitest could be used to assess primary and lower secondary school students' DCL. In order to find the answer to the main research question several specific sub-questions were formulated in accordance with the common measures used in psychometrics to understand the quality of tests.

The quality of the items must be addressed to ensure that the level of difficulty of the test is suitable for the age group. Therefore, the first sub-question is as follows.

1. What is the quality of each individual item in Digitest?

The Digitest should be suitable for assessing DCL in primary and lower secondary school. Differential item functioning allows us to make a decision as to whether the test items are suitable for both younger and older students. For that reason, the second sub-question is the following.

2. Whether each item enables a fair or biased assessment of students' DCL in different grades based on the Differential Item Functioning (DIF) analysis?

The Digitest has been created with the intent to assess DCL in ten dimensions. Discovering the latent variables of the test will allow us to decide if the test can indeed enable us to assess the desired ten dimensions; therefore, the third sub-question is as follows.

3. What are the latent variables that can be differentiated with Digitest?

2. Methods

2.1. Developing the test items and compiling the test

For developing the Digitest, an expert group of three people from the Centre for Educational Technology at the University of Tartu created the test items. The items were developed to cover the dimensions defined based on the synthesis of dimensions described by Pedaste et al. (2021) and Adov (2022). One of them specified nine dimensions of skills and knowledge, attitudes, and behaviour; the other specified four dimensions of attitudes. Authors' framework was used as the basis for the test because it covered the areas of the widely used DigComp framework, but we extended it by also including attitudes from the study by Adov (2022) and focused more specifically on digital competence in learning settings. Thus, even if the names of the specified dimensions do not reflect specific focus on the context of learning, this was nevertheless the case at the content level of the items (see Appendix A). Pedaste et al. (2021)

identified nine dimensions that could be differentiated in assessing DCL: performing operations with digital tools, communicating in the digital world, creating digital materials, programming digital content, assessment of coping in digital environments, comparative assessment of DC, legal behaviour in the digital world, protecting oneself and others in the digital world, and attitudes towards technology. The dimensions originate from the DigComp framework, and the authors contributed in ascertaining the suitable dimensions for the learning context. Furthermore, the authors also suggest capturing the diversity of attitudes. Here we implemented the work of Adov (2022), who found, based on the analysis of several frameworks, theories and authors' own empirical data, that it might be reasonable to differentiate between four different categories of attitudes. The first is *attitudes towards behaviour*, which consists of (1) anxiety, referring to a degree to which an individual experiences negative feelings about using technology (Social Cognitive Theory (SCT); Bandura, 1997; Compeau & Higgins, 1995), and (2) performance expectancy, referring to a degree to which an individual believes that using the system will help the person to attain gains in job performance (Unified Theory of Acceptance and Use of Technology (UTAUT); Venkatesh, Morris, Davis, & Davis, 2003). The second dimension is *social aspects*, which is divided into (1) social influence, indicating a degree to which an individual perceives that his/her technology usage is important to others (UTAUT; Venkatesh et al., 2003), and (2) facilitating conditions, indicating an individual's belief in the existence of a technology support system (UTAUT; Venkatesh et al., 2003). The third dimension is *perceived control*, which is describing self-efficacy as belief in one's abilities to execute the actions required (SCT; Bandura, 1997). And the fourth and final dimension is *behavioural intention*, which indicates an individual's readiness to perform a behaviour (Theory of Reasoned Action, Theory of Planned Behaviour, Technology Acceptance Model 2, UTAUT; Ajzen, 2011). The statements about attitudes towards digital technology in the four dimensions were formulated in the current study in accordance with Adov (2022), as previous research had also demonstrated a link between students' attitudes and performance and shown that attitudes should be taken into consideration for this reason (Aydin, 2021). However, as Adov (2022) focused on attitudes related to using mobile devices for learning, the original statements for attitudes towards digital device usage for learning were created in our study. The origin of each dimension can be seen in Fig. 1.

Furthermore, the aim of the Digitest was also to represent the areas brought forward by the DigComp framework. Therefore, the six dimensions concerning knowledge and skills in the Digitest and 21 specific competences of DigComp 2.2. Framework were matched in order to make sure that the Digitest captures the complexity of the widely used digital competence framework. The comparison can be seen in Table 1. It does not cover comparison of the Digitest items focusing on behaviour and attitudes because they are not so evident in the descriptions of the DigComp. As revealed in the analysis, some of the Digitest items represent multiple DigComp sub-dimensions, which could also be a reason why DigComp dimensions have not been empirically shown before. Moreover, a few of the DigComp dimensions are not straightforwardly embodied in the test items. However, the connection can be made on a more general level –

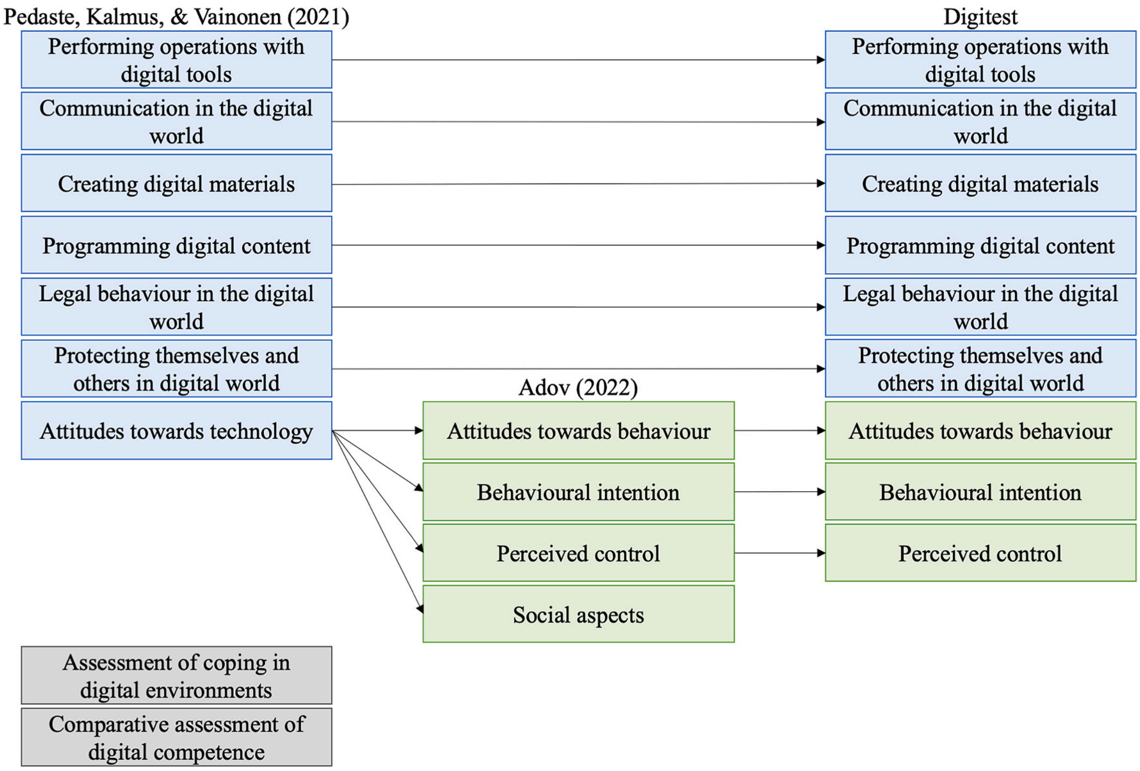

Fig. 1. Compilation of the dimensions of the Digitest (blue boxes indicate the Digitest dimensions originating from the study by Pedaste et al. (2021); green boxes refer to those originating from Adov (2022); and grey ones refer to dimensions that were not included in the current study). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Comparison between dimensions of Digitest items and DigComp framework.

Knowledge/skills dimension of Digitest	Digtest item	DigComp sub-dimension (based on the conceptual model in Vuorikari et al., 2022)
Creation of digital materials/content	CREAT1	1.3 managing data, information and digital content (information and data literacy)
	CREAT2	3.1 developing digital content (digital content creation)
	CREAT3	3.1 developing digital content (digital content creation) 5.1 solving technical problems (problem solving)
	CREAT4	3.1 developing digital content (digital content creation)
Programming digital content	PROG1	3.4 programming (digital content creation)
	PROG2	3.4 programming (digital content creation)
	PROG3	3.4 programming (digital content creation)
Protecting themselves and others in the digital world	PROT1	4.1 protecting devices (safety)
	PROT2	4.1 protecting devices (safety)
	PROT3	4.1 protecting devices (safety) 4.2 protecting personal data and privacy (safety)
	PROT4	4.3 protecting health and well-being (safety)
Legal behaviour in the digital world	LEG1	2.2 sharing through digital technologies (communication and collaboration)
	LEG2	2.6 managing digital identity (communication and collaboration) 3.3 copyright and licences (digital content creation) 4.2 protecting personal data and privacy (safety)
	LEG3	3.3 copyright and licences (digital content creation)
	LEG4	3.3 copyright and licences (digital content creation)
Communicating in the digital world	COMM1	2.6 managing digital identity (communication and collaboration)
	COMM2	2.6 managing digital identity (communication and collaboration) 4.2 protecting personal data and privacy (safety)
	COMM3	2.1 interacting through digital technologies (communication and collaboration) 2.2 sharing through digital technologies (communication and collaboration) 2.3 engaging citizenship through digital technologies (communication and collaboration) 2.4 collaborating through digital technologies (communication and collaboration)
	COMM4	2.1 interacting through digital technologies (communication and collaboration) 2.4 collaborating through digital technologies (communication and collaboration) 2.5 netiquette (communication and collaboration) 4.4 protecting the environment (safety)
Performing operations with digital tools	OPER1	1.1 browsing, searching and filtering data, information and digital content (information and data literacy) 1.3 managing data, information and digital content (information and data literacy) 5.2 identifying needs and technological responses (problem solving)
	OPER2	1.1 browsing, searching and filtering data, information and digital content (information and data literacy) 1.2 evaluating data, information and digital content (information and data literacy)
	OPER3	3.1 developing digital content (digital content creation)
	OPER4	3.1 developing digital content (digital content creation) 3.2 integrating and re-elaborating digital content (digital content creation) 5.3 creatively using digital technology (problem solving)

DigComp dimension 4.4 “Protecting the environment” could be tied to test item COMM4, where the students have to think about the benefits of online meetings, potentially a more sustainable choice compared to face-to-face meetings, and DigComp dimension 5.4 “Identifying digital competence gaps” is in fact the main goal of the test, as students might not be objective when judging their capabilities, as stated earlier. Furthermore, attitude dimensions are not separately represented in DigComp; therefore, they are also not represented in Table 1.

After compiling the first version of the test, it was further improved in three pilot studies. Revisions were made after each phase before conducting the next pilot study.

The first pilot study was conducted with four students in one-on-one settings. Typical students from primary school were selected because it was expected that younger students might have more issues completing the test. The students were asked to think aloud while taking the test and the researcher was also asking questions whenever concerns were noted (e.g., when the student was thinking very long on one assignment). After the first pilot study, necessary clarifying changes (mainly wording) were introduced to the test.

The next phase in the preparation of the test was conducting the second pilot study (N = 185) which aimed to find the items most suitable for assessing students’ skills, knowledge and attitudes for the test as well as decrease the overall number of items, thus also reducing the time needed for taking the test. In the second pilot study, observers from the research group were also present, taking notes of the students’ possible misunderstandings. The students were guided to ask all questions and express all opinions they would have when taking the test. Moreover, comprehension of the items by the students was the focus this time as well, and criteria for assessing open-ended answers were developed. Cohen’s kappa was used in the pilot phase to come to a suitable level of agreement when assessing the open-ended answers, and finally, items for which the level of agreement was moderate or stronger were selected: 18 items had a Kappa value above 0.8 and four items above 0.7. The difficulty of the test items and their alignment with the total test score as well as estimated discrimination of students were assessed using Item Response Theory (IRT) analysis (if the item’s statistical indicators were out of the desired range, the item was analysed further by the experts and changes in the item content were made as needed). Confirmatory Factor Analysis (CFA) was used to test how items in different dimensions of DCL are suitable for forming latent variables. If some of the items had low regression scores and affected the model fit negatively then we considered replacement or revision of those items. Finally, based on the second pilot study, six test items were revised and one new item was added.

The revised test was used in the third pilot study (N = 426; ages 9–15; 16 students from the 3rd grade, 112 from the 4th grade, 56 from the 5th grade, 87 from the 6th grade, and 155 from the 8th grade) in which the difficulty and the quality measures of the items were once again analysed. For the final version of the test (see Appendix A), seven items were removed and three items assessing attitudes were added based on the IRT and CFA of the third pilot study. The test version created after the third pilot study is the focus of the current study and will be used in the following analyses.

2.2. Sample of the study

For this study, the data collected from November 2021 to March 2022 were used. Thirteen Estonian schools (62 classes of students from those schools) participated in the data collection. The schools were purposively selected to cover a wide variety of students according to three criteria. First, we took into account students' results in national academic tests (math, language, science). Second, we used data about their digital readiness evaluation (assessment by students, teachers and parents) and national digital competence assessment. Third, we applied the results of the national well-being survey (assessment by students, teachers and parents). Based on each of the three criteria (academic achievement, digital readiness, perceived well-being), schools were divided either into a high-level, middle-level or low-level group so that they could be described by their performance patterns (e.g., high, low, high); however, only the schools that belonged to the high-level or low-level group were selected for the sample. Thus, each school was categorized into one of eight groups based on its pattern: 1) high, high, high, 2) high, high, low, 3) high, low, high, 4) low, high, high, 5) high, low, low, 6) low, low, high, 7) low, high, low, 8) low, low, low. At least one school was selected from each of the eight groups and two schools were selected from the extreme groups (1 and 8). After that, a call was sent to schools from various groups to have a larger sample to meet the needs specified in the project in the context of which the study was conducted. The study was conducted with the approval of the Ethics Committee of the University of Tartu. Informed consent was obtained from students and their parents. No incentives were offered to either the schools or students participating in the study. However, all schools were provided with feedback about DC in their classes.

The Digitest was one of the assessment instruments used in the study; it was taken by 836 students, 316 from the 3rd grade (age 9–10 years), 291 from the 6th grade (age 12–13 years) and 229 from the 9th grade (age 15–16 years). All the classes were mixed-gender classes and the size of the classes varied from 8 to 31 students.

2.3. The procedure of conducting tests

The test was conducted in an electronic environment used for most state-level tests in Estonia —the Examination Information System (EIS) developed in Estonia by Foundation Innove. It was completed in the Estonian language (the language of instruction) using desktop computers, laptops or tablets.

University students involved in the research group together with students' regular teachers introduced the test and monitored the process of taking the test. It was done during the school day. The university students and teachers were allowed to help with the technical questions (e.g., how to move to the next question). Six different types of items were used in the final version of the Digitest: (1) multiple-choice questions with only one correct option, (2) multiple-choice questions with more than one correct option, (3) tasks to form a sequence of phases, (4) matching items, (5) marking something in a picture, and (6) open-ended questions.

For taking the test, participating students were first registered by the schools and then they could log into the system, where they could answer all the questions sequentially, as it was not possible to move back to the previous question. Students did have the opportunity to skip questions, and the university students and teachers were instructed to encourage students to skip tasks when they saw that the student could not complete the task at hand. This practice revealed that tasks concerning programming skills were often skipped by the third-graders, and in many cases, by the sixth-graders as well. A similar pattern was observed for several items assessing operational skills, in the last section of the test. One of these (OPER4) was often skipped even by ninth-graders (62 out of 229), probably because it was quite a demanding task – students had to download a picture, edit it, and upload it again. However, a decrease in test-taking motivation (see Silm et al., 2020 for an overview) might also be an explanation. The number of missing values for items assessing attitudes showed decreased test-taking motivation among the third- and sixth-graders for items at the end of the test—the ones focusing on attitudes.

There was no time limit for taking the test. However, the students' regular schedules and routines nevertheless may have limited their motivation to complete the test. For example, if a student working at a slower pace needed more time than the regular time scheduled for a lesson and this interfered with the scheduled breaks, the student may have chosen to leave the test unfinished. On average, the test was completed in 44 min (the time ranged from 11 to 91 min, SD = 10.1).

2.4. Data analysis

Data analysis was performed in three phases following three research questions.

The first phase was aimed at evaluating the quality of each item used in the pilot and main study. Assuming that all items measure the same domain (DCL), we used one-parameter Item Response Theory (1PL IRT, i.e. the Rasch measurement model). We preferred 1PL IRT analysis to 2PL for two reasons: (a) it allows more rigorous evaluation of the quality of items because it requires each item to have good discrimination, contrary to the 2PL model adjusting discrimination of items to the data, and (b) it is more suitable for identifying any shortcomings of the items and provides a better ground for their improvement (Fox & Bond, 2015). The analysis was performed with Winsteps 5.3.0.0 because it enables most detailed evaluation of the quality of the items (see the Winsteps manual, http://www.winsteps.com/winman/index.htm?diagnosingmisfit.htm). The quality of each item in the Digitest was evaluated based on several criteria: (1) how much the estimated discrimination index differs from the required value of 1; (2) how much each item fits to the expectation of the 1PL model based on infit and outfit (expected to be in the range 0.7–1.3); (3) how good is item-scale correlation (>0.3 for good items, and >0.2 for acceptable value); (4) to which extent distribution of item difficulty across the scale (with average item difficulty set on 0) enables good measurement of different levels of DCL; and (5) quality of scoring key for each item, that is, whether different scores (full credit and half credit scores) fit to the expected value.

The second phase focused on evaluating whether the Digitest items enable fair measurement of DCL in two grades. The analysis was performed with Winsteps and was based on differential item function (DIF). Evaluation of items for potential significant DIF size was based on the Mantel-Haenszel methods and the logit-difference method. Items with DIF size higher than 0.5 and with statistically significant DIF size were identified as items that function differently in two grades (Fox & Bond, 2015).

The main aim of the third phase was to validate dimensionality of the Digitest based on the Confirmatory Factor Analysis (CFA) performed with MPlus 8.0 (Muthén & Muthén, 1998). Evaluation of different factor models was based on a set of different indices. We considered the model to be acceptable if the value of the normed Chi square index was below 3 and good if it was below 2 (Kline, 2016). Secondly, the fit indices were calculated, and, according to Bond, Yan, and Heene (2021), the model was considered acceptable if the root mean square error of approximation (RMSEA) was $\leq$0.05, the comparative fit index (CFI) $\geq$ 0.95, and the Tucker-Lewis index

Table 2
The dimensions and items of the Digitest.

Dimension		Dimension description	Test items
Attitudes towards digital device usage (AT)	Social aspects (AT_SA)	The degree to which the learner perceives that it is important to others that he or she should use technology (Venkatesh et al., 2003; e.g., value of peers' use of digital tools; motivation to do what others expect; perception of others' support when using digital tools)	AT_SA1 AT_SA5 AT_SA2
	Perceived control (AT_PC)	The learner's perception of their possibility to make decisions in using technology and their self-efficacy in using technology (Scherer, Siddiq, & Teo, 2015; e.g., students could decide whether to use digital devices; believing in one's own ability to use digital devices)	AT_PC1 AT_PC2 AT_PC3 AT_PC4 AT_PC5
	Behavioural attitudes (AT_BATT)	Positive or negative feelings towards the technology affect its use (see Ajzen, 2012; e.g., anxiety towards using digital tools; belief in the value and simplicity of digital tools)	AT_BATT1 AT_BATT2 AT_BATT3 AT_BATT4 AT_BATT5 AT_BATT6
	Behavioural intention (AT_BINT)	Learner's readiness to perform a behaviour (Ajzen, 2011; e.g., students prefer digital devices or environments to other ways of doing things)	AT_BINT1 AT_BINT2 AT_BINT3 AT_BINT4
Skills to use digital devices, materials, and content and to create digital content	Creation of digital materials/content (CREAT)	Knowledge and skills required to create digital texts or visual materials (Pedaste et al., 2021)	CREAT1 CREAT2 CREAT3 CREAT4
	Programming digital content (PROG)	Knowledge and skills required for formulating problems and their solutions so that the solutions are represented in a form that can be carried out by an information-processing agent (Wing, 2006; e.g., specify step-by-step the actions required for achieving a goal or interpreting the programs written by others)	PROG1 PROG4 PROG3 PROG5 PROG2
	Communicating in the digital world (COMM)	Knowledge and skills required for secure online communication in the digital world (Pedaste et al., 2021)	COMM1 COMM2 COMM3 COMM4
	Performing operations with digital tools (OPER)	Technical knowledge and skills acquired by using digital devices to solve problems (Pedaste et al., 2021; e.g., common activities required for managing and learning with digital devices)	OPER1 OPER2 OPER3 OPER4
Behaviour in digital learning environments	Protecting themselves and others in the digital world (PROT)	The respondent's ability to protect themselves and others from the threats in the digital world (Pedaste et al., 2021)	PROT1 PROT2 PROT3 PROT5 PROT4
	Legal behaviour in the digital world (LEG)	Following the rules in the digital world (Pedaste et al., 2021)	LEG1 LEG5 LEG2 LEG3 LEG4

(TLI) $\geq$ 0.95. In addition, we expected to see weighted root mean square (WRMR) index close to 1.0. All items were treated as categorical because the scales were rather short and sometimes even dichotomous (correct/incorrect). Where the theoretical model was not supported, Exploratory Factor Analysis (EFA) was conducted. Moreover, composite reliability for all the dimensions was calculated.

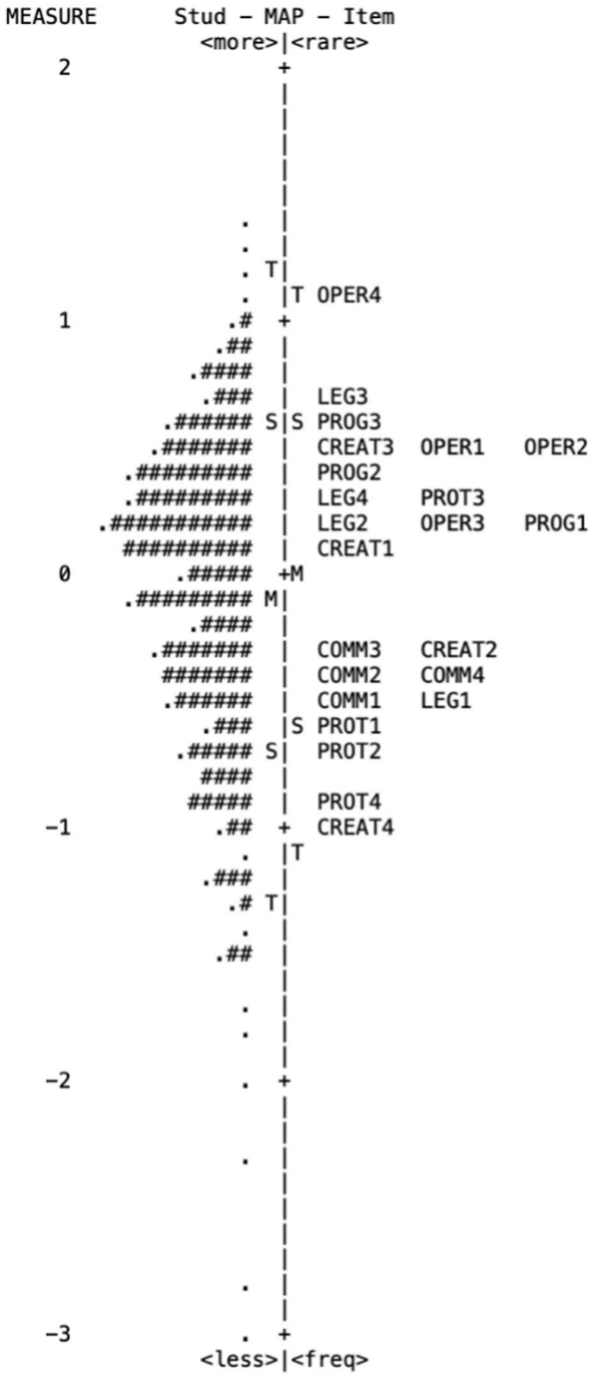

Fig. 2. The Item-Person map (Wright Map) presenting the difficulty of Digitest items and distribution of student abilities
M: student IRT measure (left side); average difficulty of items (right side)
S: one standard deviation from the average student IRT measure (left side) and items (right side)
T: two standard deviations from the average student IRT measure (left side) and items (right side).

3. Results

3.1. Description of the digitest and descriptive data on students' scores

The final version of the Digitest includes 41 items in total: 18 items to assess students' digital attitudes, 15 items to assess skills and 8 items to describe behaviour (the specific dimensions and items are in Table 2 below, and the final test used for data collection with the scoring key is in Appendix A). The test in Appendix A is the English translation of the test; the test was conducted in Estonian. The English version has not yet been validated since the current article focuses on the data obtained from Estonia.

The overall quality of the test was good based on IRT analysis—the average item difficulty was −0.08, the student reliability was 0.86 and the item reliability was 0.88. In Fig. 2, the difficulty per item is presented. The Wright map shows that the average difficulty of the Digitest items is in accordance with the average student IRT measure. Students with either high or low levels of DCL can be adequately measured by Digitest since the higher end of the distribution of the students is on the same level as the highest levels of the item thresholds, and the same can be said about the lower end of the scale.

The test described students' digital attitudes on a continuous scale of 1–6. Different scales were used to assess the characteristics describing skills and behaviour, the values of which varied depending on the item (see Appendix A). Some of the questions were computer-assessable (24), some human-assessable (13) and some hybrid-assessable (4). In the case of hybrid-assessable items, the assessor could also mark as correct the answers that the computer had marked as incorrect (e.g., if the student had marked the correct answer differently, for example where a sequence of letters was expected without spaces, but students added spaces or commas between the letters). The evaluation of the questions was carried out by one evaluator, as the suitability of the evaluation guidelines was already ensured in the pilot phase of the study based on calculations of Cohen's weighted kappa.

3.2. The quality of each individual item of the digitest and the potential of the test to measure DCL

At the item level, digital attitudes and questions assessing skills and behaviour were analysed separately because they were considered to be different types—the former were measured with the self-evaluation scale and the latter with test items. The IRT test results for the statements describing digital attitudes are presented in Table 3. The results of the IRT analysis of the questions assessing students' digital skills and behaviour are presented in Table 4. Based on IRT analysis, the item infit and outfit scores were sometimes

Table 3
Indices of the quality of Digitest attitude items evaluated based on the 1PL IRT analysis.

Item[a]	Measure[b]	Mean infit[c]	Mean outfit[c]	Corr.[d]	Estim. Discrim.[e]
AT_BATT1	−0.87	**1.35**	1.21	0.39	1.04
AT_BATT2	−1.02	**1.38**	**1.31**	0.37	1.03
AT_BATT3	−0.09	1.19	1.22	0.39	0.89
AT_BATT4	−0.06	1.02	1.02	0.47	1.01
AT_BATT5	−0.83	**1.34**	**1.71**	0.31	0.99
AT_BINT1	−0.05	0.90	0.93	0.57	1.22
AT_BINT2	0.5	1.18	1.20	0.48	0.88
AT_BINT3	0.52	1.15	1.24	0.49	0.90
AT_BINT4	0.47	1.15	1.16	0.46	0.84
AT_PC1	−0.41	0.92	0.87	0.52	1.11
AT_PC2	−0.08	0.95	1.01	0.49	1.06
AT_PC3	−0.15	0.71	0.70	0.55	1.17
AT_PC4	−0.14	**0.60**	**0.61**	0.64	1.25
AT_PC5	0.05	**0.65**	**0.69**	0.56	1.17
AT_SA1	0.68	**1.31**	**1.40**	0.24	0.35
AT_SA2	0.39	1.14	1.19	0.40	0.68
AT_SA3	0.23	0.86	0.89	0.45	0.98
AT_SA4	0.86	1.02	1.16	0.27	0.70

Values outside of the acceptable range are marked in bold.

[a] Item names; the first letters (AT) show that the items measure attitudes, and the letter combination after the underscore shows which attitude dimension each item evaluates (BATT – behavioural attitudes, BINT – behavioural intention, PC – perceived control, SA – social aspects) and the number of the item in the test.

[b] Probability of choosing a statement (average value is 0, higher values indicate higher frequency and lower values lower frequency).

[c] Mean infit and mean outfit refer to the Infit MNSQ and Outfit MNSQ indices, which suggest how well students' estimated scores (based on item difficulty and students' IRT measure) fit to real student scores (items with infit and outfit indices in the range of .7–1.3 are considered good items).

[d] Correlation between item score and respondent IRT measure estimated based on the 1PL IRT model (items with a correlation >.2 are considered satisfactory items; items with a correlation >.3 are considered good items).

[e] Estimated item discrimination, i.e. what the item discrimination index would be if the data were analysed using a 2PL IRT model (items with estimated discrimination in the range of .5–2.0 are considered to be items with satisfactory discrimination).

Table 4
Indices of the quality of Digitest skills and behavioural items evaluated based on the 1PL IRT analysis.

Item[a]	Type[b]	Measure[c]	Mean infit[d]	Mean outfit[d]	Corr.[e]	Estim. Discrim.[f]
LEG1	6	−0.5	1.30	1.28	0.56	1.23
LEG2	6	0.24	1.21	**1.35**	0.40	1.06
LEG3	6	0.74	0.70	**0.69**	0.60	0.94
LEG4	6	0.25	1.30	**1.41**	0.49	1.06
CREAT1	1	0.11	0.85	0.89	0.56	1.06
CREAT2	1	−0.31	0.90	0.95	0.59	1.13
CREAT3	1	0.48	1.01	1.13	0.43	0.98
CREAT4	2	−0.99	0.88	0.76	0.61	1.06
PROG1	3	0.24	**1.46**	**2.71**	0.41	0.71
PROG2	1	0.4	**1.34**	**1.46**	0.44	1.22
PROG3	1	0.63	**1.35**	**1.46**	0.39	1.16
OPER1	7	0.49	0.83	0.77	0.70	1.01
OPER2	7	0.51	**0.61**	**0.56**	0.64	1.69
OPER3	5	0.2	0.77	0.80	0.68	1.13
OPER4	7	1.11	0.77	0.68	0.53	1.01
COMM1	1	−0.53	0.82	0.90	0.61	**−0.09**
COMM2	6	−0.42	**1.37**	**1.37**	0.54	**−0.03**
COMM3	1	−0.28	0.79	**0.69**	0.69	1.41
COMM4	1	−0.37	0.89	0.88	0.60	1.12
PROT1	1	−0.63	**1.60**	**1.63**	0.47	1.18
PROT2	1	−0.71	**0.62**	0.70	0.59	0.87
PROT3	1	0.3	0.72	0.73	0.56	1.00
PROT4	1	−0.93	0.88	0.90	0.57	1.09

Values that are outside of the acceptable range are marked in bold.
[a] Item names; the letters indicate which skill or behaviour of the DC dimensions the question describes (LEG – legal behaviour in the digital world, CREAT – creating digital materials, PROG – programming digital content, OPER – performing operations with digital tools, COMM – communicating in the digital world, PROT – protecting oneself and others in the digital world) and the number of the item in the test.
[b] Item type: 1 – open-ended question; 2 – forming a sequence of phases; 3 – programming; 4 – multiple-choice question with only one correct option (radio button); 5 – matching items; 6 – multiple-choice question with more than one correct option (check boxes); 7 – set of operations.
[c] Item difficulty measure (average value is 0, higher values indicate more difficult items).
[d] Mean infit and mean outfit refer to the Infit MNSQ and Outfit MNSQ indices, which suggest how well students' estimated scores (based on item difficulty and students' IRT measure) fit to real student scores (items with infit and outfit indices in the range of .7–1.3 are considered to be good items).
[e] Correlation between item score and respondent IRT measure estimated based on the 1PL IRT model (items with a correlation higher than .2 are considered to be satisfactory items; items with a correlation higher than .3 are considered to be good items).
[f] Estimated item discrimination, i.e. what the item discrimination index would be if the data were analysed using a 2PL IRT model (items with estimated discrimination in the range of .5–2.0 are considered to be items with satisfactory discrimination).

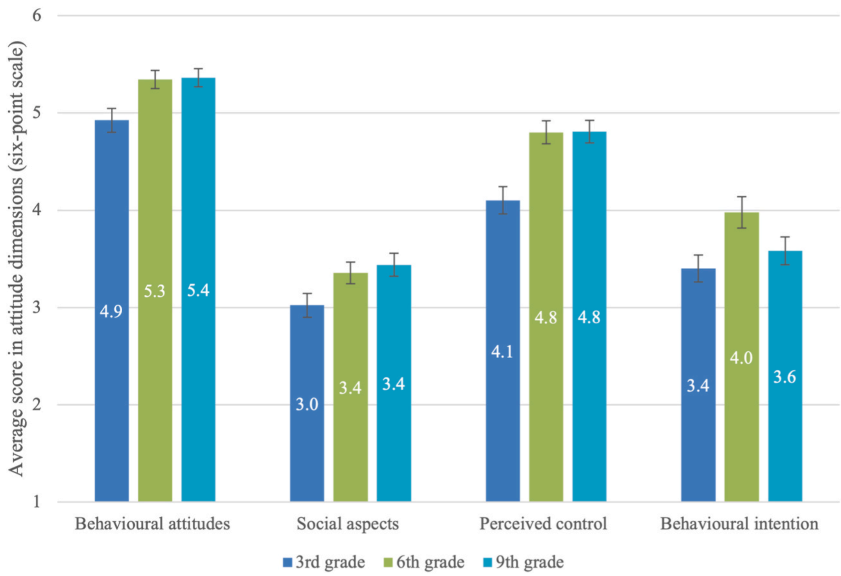

Fig. 3. Student scores for attitude questions based on average values with 95% confidence intervals (higher scores indicate more positive attitudes).

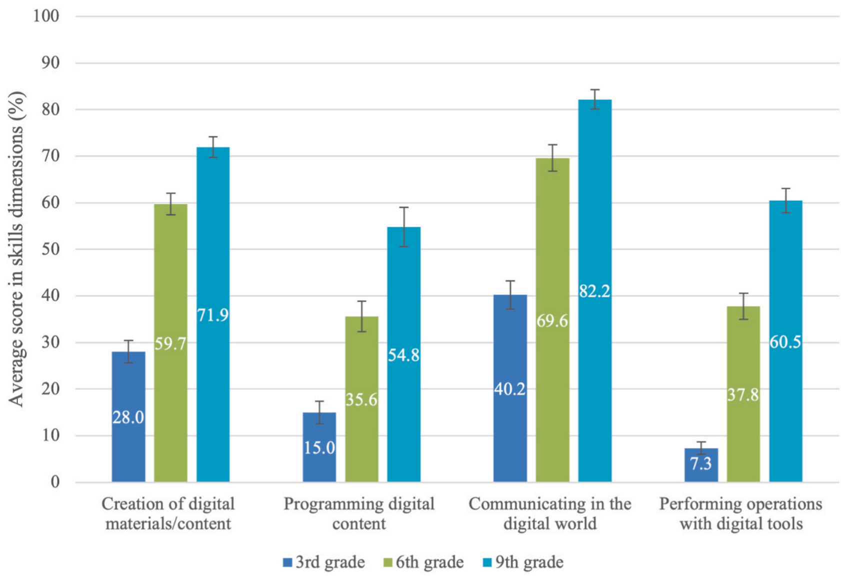

Fig. 4. Student scores in the skills dimensions based on average values with 95% confidence intervals (both presented in percentages).

out of the expected range, but this is usually not considered a significant issue if the other quality indicators are acceptable. Expected correlation was above the suggested value 0.3 for almost all items (in two cases it was above 0.2, which is considered acceptable) and in the case of estimated discrimination, there were only two items outside of the suggested range of values (items COMM1 and COMM2). Thus, in conclusion, the IRT analysis did not provide strong reasons for revising the test by leaving out some items.

For descriptive statistics and correlation data, see Figs. 3–5 (by age groups; the average values of dimensions of skills and

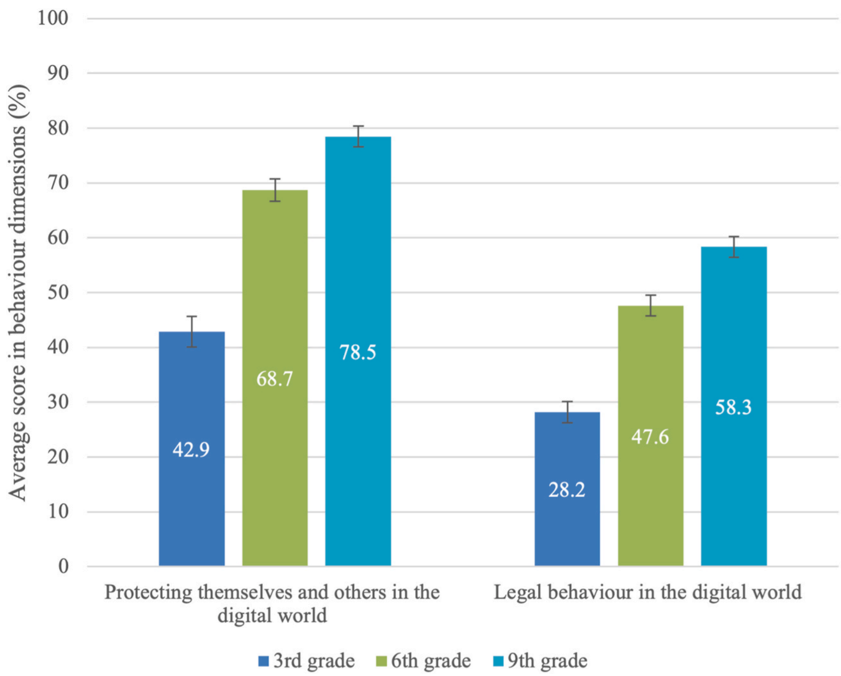

Fig. 5. Student scores in the behaviour dimensions based on average values with 95% confidence intervals (both presented in percentages).

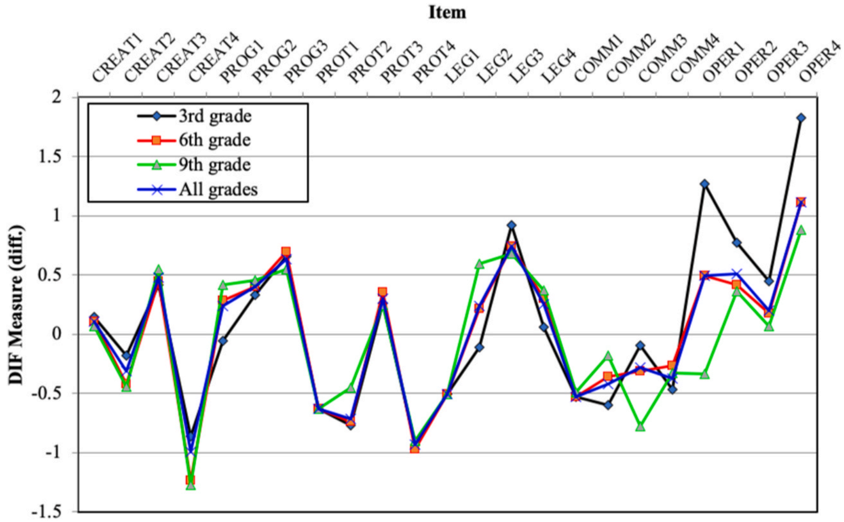

Fig. 6. Results of IRT DIF-analysis to evaluate Digitest digital skills and behaviour questions by different student grades (3rd grade, n = 316, black line; 6th grade, n = 291, red line, 9th grade, n = 229, green line; the mean of all grades is indicated by a blue line). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

behaviours are represented in percentages, since the items have different maximum values) and Appendix B. The results show that social aspects have the lowest scores (the most negative) in attitude questions, being the only dimension in which all students evaluated their attitude as average or below. Students' skill scores vary depending on the grade, and younger students especially have lower scores on items assessing programming and operational skills. Of the behavioural dimensions, the legal behaviour scores are lower than those for protecting themselves and others in the digital world. Additionally, it is worth mentioning that even though the average scores of 3rd grade students in skills and behaviour dimensions are significantly lower than those of 6th or 9th grade students, for most dimensions (with the exception of "creation of digital materials/content" and "performing operations with digital tools") there were also 3rd grade students who performed on a similar level compared to 9th-graders, as the standard deviation values were

Table 5
Statistically significant differences in difficulty measure by items in the cases where the DIF contrast is considered meaningfully large (at least 0.5; student groups represent the viewed age group (Grade 3, 6 or 9).

Item	Student group 1	Difficulty measure 1	Student group 2	Difficulty measure 2	DIF Contrast
LEG2	3	−0.11	9	0.60	−0.70
COMM3	3	−0.10	9	−0.78	0.68
OPER1	3	1.27	6	0.49	0.77
OPER1	3	1.27	9	−0.34	1.60
OPER1	6	0.49	9	−0.34	0.83
OPER4	3	1.83	6	1.11	0.71
OPER4	3	1.83	9	0.88	0.95

Table 6
Fit indices of different factor structures.

Model	χ^2	df	χ^2/df	RMSEA	CFI	TLI	WRMR
9-factor model	1109.759	558	1.99	0.034	0.972	0.969	1.152
3-factor higher-order model (attitudes, skills and behaviours)	1156.081	582	1.99	0.034	0.971	0.969	1.219
2-factor higher-order model (attitudes; skills and behaviour)	1158.310	584	1.98	0.034	0.971	0.969	1.223
1-factor higher-order model (all factors loading to one higher-order "DCL" factor)	1839.410	585	3.14	0.051	0.937	0.932	1.619
3-factor model (attitudes, skills and behaviours)	1696.236	591	2.87	0.047	0.944	0.940	1.545
2-factor model (attitudes; skills and behaviour)	1693.169	593	2.86	0.047	0.944	0.941	1.546
1-factor model (all items)	3808.621	594	6.41	0.080	0.838	0.828	2.398

relatively high.

3.3. Differential item functioning in digitest

With the second research question we attempted to find out whether each test item enables a fair or biased assessment of students' DCL in different grades based on the DIF analysis. Viewing the DIF analysis results (see Fig. 6), some differences across grades occurred, although in general, all items were equally suitable for use, regardless of age group.

The remarkable and statistically significant contrast between grades is present in items LEG2, COMM3, OPER1 and OPER4 (see Table 5). Item LEG2 concerns the usage of photos, which also includes other people in addition to the student. In this case, the item was relatively easier for the 3rd grade students than for the 9th grade students. The reason could be that, according to the national curriculum, internet privacy related topics are addressed mainly in 4th to 6th grade and some aspects concerning regulations that apply to the internet are already addressed in 3rd grade, making it a less relevant topic for older students.

The rest of the items were relatively more difficult for the younger students. COMM3 asks about benefits of digital learning environments for collaboration at distance. The better results of older students might be explained by their more extensive experience in using different online learning environments because distance learning was used in their cases more often and they had more teachers who implemented a larger variety of environments. OPER1 and 4 are tasks in which each student's technical skills are assessed. OPER group items are expected to be more difficult for younger students.

3.4. Latent variables and internal consistency of the digitest scales

In order to determine the latent variables differentiable with Digitest, we tested whether the CFA model fitted the data of the entire sample. The analysis with all items had some issues because the model fit was not acceptable. Closer inspection of the items revealed that one item (AT_BATT4) in the factor AT_BATT had a negative regression; it was left out from models and will be excluded from the Digitest in further analyses and should be removed from the test when it is further used, although its content was in line and in the same direction with other items described by the same factor. This item assessed students' opinion about the difficulty of learning in digital environments. The selected 40 items in 10 factors had factor loadings above 0.3; however, two of the items from the social aspects dimension had the models' lowest factor loadings (0.335 for AT_SA1 and 0.334 for AT_SA4). Due to this, internal consistency was calculated for the attitude dimensions using composite reliability, and the values were as follows: social aspects (0.52), perceived control (0.85), behavioural attitudes (0.79), and behavioural intention (0.66). Next, EFA was conducted to understand the factor structure of the attitude items. The results showed that in the case of both three- and four-factor models the items describing social aspects did not load into one factor but mostly had quite similar factor loadings when loading into two factors. Therefore, based on the initial 10-factor model factor loadings, EFA and composite reliability calculations, the attitude dimension of social aspects was dropped from the model.

Several models with nine dimensions of the digital competence were then tested for data fit (the models' fit indices can be seen in Table 6). The analysis revealed that the 9-factor model and the 2- and 3-factor higher-order factor models all fitted the data similarly well. However, as we were defining competence based on Weinert's (2001) theory, where competence consisted of cognitive-behavioural and motivational elements, the 2-factor higher-order model was chosen as the best fit for the data and the previous theory (the final model can be seen in Fig. 7). We noticed that the correlation between the two factors was rather low (0.459) and therefore there was no reason to assume another higher order factor (i.e. general factor) that would be measured by two factors. However, we noted that the 3-factor higher-order factor model should also be discussed and considered in further studies because the model's fit with the data was equally good. Still, our findings also show that the 3-factor model has an issue (the correlation of latent higher order variables describing skills and behaviour was 1.035) that was not identified in the 2-factor model. Therefore, these findings suggest that DCL can be described based on a model including a higher-order factor structure and not a lower-order factor structure of one, two or three factors.

The models' fit with the data was also tested separately for 3rd, 6th and 9th grade students and the analyses confirmed acceptable fit indices in all grades (see Table 7).

Finally, internal consistency of the scales described by the nine factors of the final 2-factor second-order factor model was calculated. Composite reliabilities were from 0.65 to 0.91 for different latent variables, which is at least at an acceptable level above 0.60 (Hajjar, 2018), although in a few dimensions it was rather low, meaning that in those dimensions more items should be developed and tested in further studies.

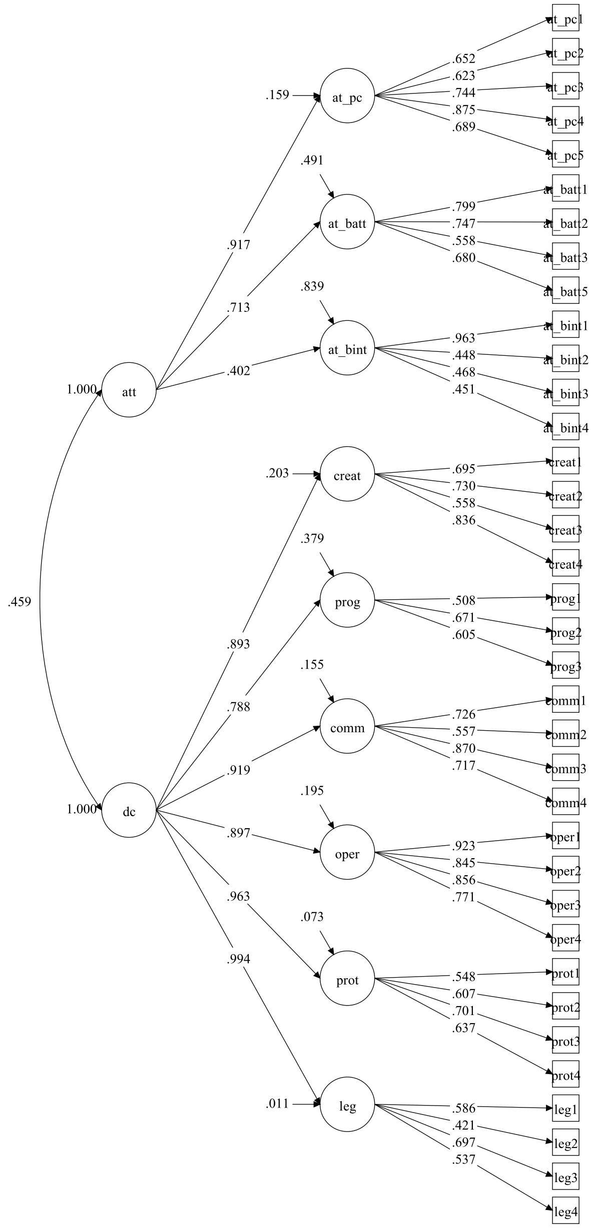

Fig. 7. Factor model of DCL based on data from the second phase of the study (n = 836; factors: att – attitudes towards digital device usage, dc – skills to use digital devices and behaviour in digital learning environments, at_pc – perceived control, at_batt – behaviour-related attitudes, at_bint – behavioural intention, creat – creation of digital materials, prog – digital content programming, comm – communication in the digital world, oper – digital operations, prot – legal behaviour in the digital world, leg – protection of oneself and others in the digital world). The figure shows the standardized results where the variances of the latent features are fixed. All features were considered categorical.

Table 7
Fit indices of 2-factor higher-order model (attitudes; skills and behaviour) for 3rd, 6th and 9th grade data.

Model	χ^2	df	χ^2/df	RMSEA	CFI	TLI	WRMR
3rd grade	726.644	584	1.24	0.028	0.930	0.924	1.005
6th grade	752.076	584	1.29	0.031	0.923	0.917	1.024
9th grade	891.170	584	1.53	0.048	0.843	0.830	1.195

4. Discussion and implications

Previous studies about DC have been clear to state that although DC is an important concept in the society and therefore in learning, there is still no systematic and meaningful way to assess it in schools in different dimensions in the context of learning. The need for assessment has risen since the remote schooling experience due to the COVID-19 situation brought evidence that the DC (more precisely DCL) of the students is not at the desired level (Carretero et al., 2021) and we need to take steps to improve it systematically. It is of our specific interest to focus on DCL because the more general focus on DC has proved a challenge in determining practically useful multidimensionality in the studies focusing on assessment instruments (see Siddiq et al., 2016 and Godaert et al., 2022 for reviews). The present study was designed to determine if the test designed to assess DCL (Digitest) is suitable for assessing the DCL of primary and lower secondary school students. The test was created on the basis of DCL assessment suggestions outlined by Pedaste et al. (2021), who developed their test based on the DigComp framework and EU Kids Online framework. In addition, the study by Adov (2022), specifying attitudes towards technology use, was taken into account in theoretically developing the dimensions of DCL. The idea of linking the more skills, knowledge and action oriented DigComp-based approach with attitudes towards technology use can be attributed to Weinert (2001), who described several approaches to conceptualize competence. Based on the ideas of Weinert we anticipated that DCL could be seen as a specialized competence that consists of higher-level dimensions describing different knowledge and skills, behavioural and attitudinal dimensions of the competence.

The main research question was whether the Digitest could be used to assess primary and lower secondary school students' DCL; in order to draw conclusions, three sub-questions were formulated in the current study. Next, we discuss the findings according to the sub-questions.

The first sub-question was about evaluating the quality of the items in the Digitest and the test's potential to measure DCL. For assessing the Digitest's suitability for its purpose, the quality and potential of the test items were analysed. The data are showing a variety of difficulty levels when it comes to the test items. This is a valuable finding, as the items are also in an accepted range of difficulty even in the different grades (from the 3rd to the 9th grade) in the focus of the current study; therefore, we can conclude that the test items are suitable for the target group. We have thus contributed to solving the shortage of psychometrically validated instruments for assessing DC (or more specifically DCL) at the level of primary and lower-secondary schools (see Godaert et al., 2022; Siddiq et al., 2016). Although the reviews mentioned above were studying DC, we believe that in the case of young students, learning-related DC is even more important as it enables students to leverage their learning process and allows teachers to use a wider toolset to enhance learning.

The second sub-question was formulated to see whether each item enables a fair or biased assessment of students' DCL in different grades, as the Digitest was designed to be able to assess the competence of primary and lower secondary school students. According to the data, no major differential item functioning across grades exists for the items measuring students' DCL. One item is significantly more difficult for students in older grades, and a few are relatively more difficult for the younger students; however, viewing the difficulty measures separately does not reveal alarmingly extreme values. Moreover, the data indicate that the targeting of this sample's digital skill level has been well accomplished by the current version of the Digitest. Looking at the results, the 9th grade students are performing better than younger students on items addressing the dimension of operational skills. This is somewhat surprising because technical skills are only the basis for developing other higher-level skills (see Kallas & Pedaste, 2022). However, the results showed that for younger students it was easier to complete the assignments about internet privacy. These results can be explained by the ages at which these topics are taught according to the national curriculum in Estonia (Põhikooli riiklik õppekava, 2021). At the same time, the majority of the test items were relatively more difficult for the younger students, which may either indicate the need for a revision of these items to make them more appropriate for 3rd grade students in the future or point to the need for further test development so that we can overcome the limitations which 3rd grade students may have, such as a lower level of reading speed and comprehension. Thus, the DIF-test did not reveal any major differential item functioning across grades. Furthermore, as students' average scores with standard deviations were compared, it could be seen that some of the 3rd grade students were equally competent as the 9th grade students, which leads us further to believe in the suitability of the instrument. These are encouraging findings, as the same test items are mostly suitable for both younger and older students and assessing students on higher and lower levels of DCL is possible with the test. Additionally, overall differences in the performance of different age group students paired with the result that some of the younger students presented abilities equal to their more experienced peers is raising further questions about how and what kind of digital abilities are actually taught in schools and opens up possibilities for further research on the matter.

The third sub-question was aimed at discovering which latent variables could be differentiable with the Digitest. According to the model, two higher-order and nine sub-dimensions of DCL were described with the Digitest. The model from the data fits the framework suggested by Pedaste et al. (2021) and is also consistent with Weinert's (2001) definition of competence, where competence consists of cognitive and motivational aspects, and further confirms the idea that DC consists of several dimensions, as also suggested in several other studies (Guitert et al., 2021; Porat et al., 2018; Siddiq et al., 2017; Wilson et al., 2017).

As the nine factors load into two higher-order factors showing us the overall attitude and, on the other hand, skills and behaviours, the Digitest allows us to assess the latent variables separately, helping to provide more meaningful feedback to educators, similarly to the suggestions of Wilson et al. (2017). However, here the results are calling for further discussion and additional studies. Weinert (2001) introduced, based on a review, a number of different approaches to defining the general concept of competence. However, according to our understanding these approaches are somewhat overlapping, resulting in different types of higher-order factors. The same was found in our study – it is clear that, compared with a lower-order factor structure, a higher-order factor structure of DCL is more in line with the empirical data collected with the Digitest. The difficulty lies in specifying the number of higher-order factors.

Based on Weinert's (2001) review one (ability) could be defined, but more likely there are two or three higher-order factors (motivational (see also White, 1959) or, in our context, attitudinal dimensions and then knowledge and skills either separately or in one single higher-order factor together with behavioural factors). Comparison of different higher-order factor models in our study revealed that the two-factor higher-order model is the best because all residuals and factor loadings of both higher-order and lower-order factors were in the expected range in the case of this model. However, although there were some issues with the residuals and factor loadings in the case of the 1- and 3-factor higher-order models, the fit indices were the same for all models. This might show that there is some room for improvement of the test items to clarify which higher-order factor structure is the most suitable. From the theoretical viewpoint these three structures describing DCL could be viewed as equally good. We might explain digital competence as a general construct combining nine more specific competences or a construct combining these nine specific competences into two or three more general competences that cannot, according to our analyses, be combined into an even higher level general DCL. Weinert (2001) did not favour one approach to others but suggested that favouring a particular approach might depend on the context. In the context of DCL we can see that clarification of the higher-order factor structure might be necessary for supporting researchers and practitioners in the assessment and development of DCL. However, as in the case of all of these three higher-order factor structures, the basis of the factor structure consists of nine more specific dimensions, which should be used in further studies. Thus, the main theoretical contribution of our study lies in the empirical validation of the notion that DCL could be described through nine dimensions that are organized into one to three higher-order dimensions.

The main practical contribution of our study lies in answering the general research question: whether the Digitest could be used to assess primary and lower secondary school students' DCL. According to the evidence described above we can conclude that the Digitest is suitable for assessing nine dimensions of DCL of students in primary and lower secondary school. Lack of psychometric validation of instruments assessing DC has been brought forward by Godaert et al. (2022) and earlier by Siddiq et al. (2016) in systematic literature reviews. In addition, our test is more specific compared to the existing ones because it is focusing on DC in the context in learning – DCL. At the same time, Digitest is allowing us to assess a wide variety of dimensions also described in DigComp framework but does not rely on self-assessment as the previous instruments often do according to the reviews by Siddiq et al. (2016) and Godaert et al. (2022). One more contribution of the study is giving a possibility to assess DCL throughout quite large age range with the same instrument. This is especially valuable in longitudinal studies and in comparing DCL of different age groups. Previous studies have mostly been conducted on narrower age groups (see Siddiq et al. (2016) for the review). Furthermore, Digitest is designed to reflect real-life problems, which is also the suggested by Reichert et al. (2020) and Siddiq et al. (2016). However, the assessment should be done with caution, as there were also some limitations to the study.

As the Digitest consisted of tasks where students' skills and behaviours were evaluated by experts (open-ended questions) and automatically by the testing environment, the attitudes section of the test was based on self-assessment. For more accurate data, information about attitudes should also be collected through independent measurement (e.g., using biometrical data to measure emotion arousal) or observations. Another limitation is that all conclusions in the study are based on Estonian students, and as DC or DCL is taught differently in different countries, results from other countries may vary. One more aspect to consider is that the students had the option of skipping questions during the test. Older students rarely took that option, but younger students chose to skip questions more often, resulting in a higher number of missing values. However, these missing values could be interpreted in several ways—perhaps the student was not able to complete the assignment, they may have had low motivation to complete the test, or there may not have been sufficient time to complete the test even if the student actually could, as the students' working pace varies. This means that conducting the Digitest with younger students may require a longer time allowance for completing the tasks and further motivating the students to try to complete all the tasks in the test. Furthermore, as the test assumes that the respondents read and understand the assignments, students with lower reading skills may struggle with completing the given tasks on time (this limitation could be more relevant for younger students). This limitation, however, could be removed by, for example, adding audio guides to the test. The data in our study indeed did not show reading or writing fluency to affect the quality of the Digitest in assessing DCL because there was only one item where a bit more writing was needed and for which the contrast between younger and older students was remarkable and statistically significant according to the DIF analysis. Thus, this flexibility of timing that we used in conducting the test should be also allowed in applying the Digitest in the future. Finally, we also have to note that some of the Digitest items were not clearly contextualized with a focus on learning situations, although the competence assessed by these could have been assessed in the context of learning as well. This means that some items might measure students' DC rather than DCL. Therefore, there is still some room for further improvement of the Digitest.

One of the long-term goals of the development of the test is to provide a way to compare the skills of students from different countries and follow their progress in becoming more digitally competent through their school years. Accordingly, in the next phase of the study, emphasis should be placed on assessing the invariance across grades to gain knowledge of the long-term development of students' DCL through re-testing each student several times throughout their school years and following their individual results. To achieve international comparability, the non-Estonian versions of the Digitest should be validated.

For now, the contribution of the current study lies in providing a psychometrically validated test that allows researchers and practitioners to assess students' DCL in nine dimensions and compare the results with students from the same age group. The results are valuable both for students, who will get a clearer view of their own skills, behaviours and attitudes, and for educators, who can see how digitally competent their students are compared to other students from the same age group and can then put more emphasis on areas where students need further improvement. Our findings might help the research community to develop new ways to assess DCL. At the moment, there is abundant room for further progress in determining whether the Digitest is a good basis for creating a test which will help to monitor the development of DCL throughout the primary and secondary school.

Data availability

Data will be made available on request.

Acknowledgements

We would like to thank all schools, teachers and students who have given their time to participate in our study. This study has been supported by the DIGIVARA5 project "The effect of using digital learning materials for learning and teaching in the context of Estonian basic school (1.05.2020−30.04.2023)" financed by the Ministry of Education and Research in Estonia.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.compedu.2023.104830.

References

Adov, L. (2022). *Predicting teachers' and students' reported mobile device use in STEM education: The role of behavioural intention and attitudes*. Thesis https://dspace.ut.ee/handle/10062/76359.

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

Ajzen, I. (2011). The theory of planned behaviour: Reactions and reflections. *Psychology and Health, 26*(9), 1113–1127.

Ajzen, I. (2012). The theory of planned behaviour. In P. Van Lange, A. Kruglanski, & E. Higgins (Eds.), *Higgins, handbook of theories of social psychology* (Vol. 1, pp. 438–459). United Kingdom: SAGE Publications Ltd. https://doi.org/10.4135/9781446249215.n22.

Asil, M., Teo, T., & Noyes, J. (2014). Validation and measurement invariance of the computer attitude measure for young students (CAMYS). *Journal of Educational Computing Research, 51*(1), 49–69. https://doi.org/10.2190/EC.51.1.c

Aydin, M. (2021). *A multilevel modeling approach to investigating factors impacting computer and information literacy: ICILS korea and Finland sample. Education and information technologies*. https://doi.org/10.1007/s10639-021-10690-1

Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.

Bond, T. G., Yan, Z., & Heene, M. (2021). *Applying the Rasch model: Fundamental measurement in the human sciences* (4th ed.). Routledge.

Carretero, S., Napierala, J., Bessios, A., Mägi, E., Pugacewicz, A., Ranieri, M., et al. (2021). *What did we learn from schooling practices during the COVID-19 lockdown? Insights from five EU countries*. https://doi.org/10.2760/135208

Chou, Y. C., & Chiu, C. H. (2020). The development and validation of a digital fluency scale for preadolescents. *Asia-Pacific Education Researcher, 29*, 541–551. https://doi.org/10.1007/s40299-020-00505-1

Chung, J., & Yoo, J. (2021). *Skills for life: Digital literacy*. Inter-American Development Bank. https://doi.org/10.18235/0003368

Claro, M., Preiss, D. D., San Martín, E., Jara, I., Hinostroza, J. E., Valenzuela, S., et al. (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

978-92-76-27592-3. In Clifford, I., Kluzer, S., Troia, S., Jakobsone, M., Zandbergs, U., DigCompSat, V. R., et al. (Eds.). Luxembourg: Publications Office of the European Union, Article JRC123226. https://doi.org/10.2760/77437.

Compeau, D. R., & Higgins, C. A. (1995). *Computer Self-Efficacy: Development of a Measure and Initial Test. MIS Quarterly, 19*(2), 189. https://doi.org/10.2307/249688

Fernández-Montalvo, J., Peñalva, A., Irazabal, I., & López, J. J. (2017). Effectiveness of a digital literacy programme for primary education students/Efectividad de un programa de alfabetización digital para estudiantes de educación primaria. *Cultura Y Educación, 29*(1), 1–30. https://doi.org/10.1080/11356405.2016.1269501

Ferrari, A. (2012). *Digital competence in practice: An analysis of frameworks*. Publications Office. https://data.europa.eu/doi/10.2791/82116.

Ferrari, A. (2013). *Digcomp: A framework for developing and understanding digital competence in europe*. Publications Office. https://data.europa.eu/doi/10.2788/52966.

Ford, L., Kozey, M. L., & Negreiros, J. (2012). Cognitive assessment in early childhood: Theoretical and practice perspectives. In D. P. Flanagan, & P. L. Harrison (Eds.), *Contemporary intellectual assessment: Theories, tests and issues* (3rd ed., pp. 585–622). New York, NY: Guildford Press.

Fox, C. M., & Bond, T. G. (2015). *Applying the Rasch model: Fundamental measurement in the human sciences* (3rd ed.). New York: Routledge.

Godaert, E., Aesaert, K., Voogt, J., & van Braak, J. (2022). Assessment of students' digital competences in primary school: A systematic review. *Education and Information Technologies, 27*(7), 9953–10011.

Godhe, A.-L. (2019). Digital literacies or digital competence: Conceptualizations in nordic curricula. *Media and Communication, 7*(2), 25–35. https://doi.org/10.17645/mac.v7i2.1888

Guitert, M., Romeu, T., & Baztán, P. (2021). The digital competence framework for primary and secondary schools in Europe. *European Journal of Education, 56*(1), 133–149. https://doi.org/10.1111/ejed.12430

Hajjar, S. (2018). Statistical analysis: Internal-consistency reliability and construct validity. *International Journal of Quantitative and Qualitative Research Methods, 6*(1), 46–57.

Hatlevik, O. E., Ottestad, G., & Throndsen, I. (2015). Predictors of digital competence in 7th grade: A multilevel analysis. *Journal of Computer Assisted Learning, 31*(3), 220–231. https://doi.org/10.1111/jcal.12065

Heitink, M. (2018). *Eliciting teachers' and students' technological competences: Assessing technological skills in practice*. Enschede, Nederland: Doctoral dissertation, University of Twente. https://doi.org/10.3990/1.9789036546850

Ihme, J. M., Senkbeil, M., Goldhammer, F., & Gerick, J. (2017). Assessment of computer and information literacy in ICILS 2013: Do different item types measure the same construct? *European Educational Research Journal, 16*(6), 716–732. https://doi.org/10.1177/1474904117696095

Ilomäki, L., Paavola, S., Lakkala, M., & Kantosalo, A. (2016). Digital competence – an emergent boundary concept for policy and educational research. *Education and Information Technologies, 21*(3), 655–679. https://doi.org/10.1007/S10639-014-9346-4

Jin, K. Y., Reichert, F., Cagasan, 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*. https://doi.org/10.1016/J.COMPEDU.2020.103968

Jun, S., Han, S., Kim, H., & Lee, W. (2014). Assessing the computational literacy of elementary students on a national level in Korea. Educational Assessment. *Evaluation and Accountability, 26*, 319–332. https://doi.org/10.1007/s11092-013-9185-7

Kallas, K., & Pedaste, M. (2022). How to Improve the Digital Competence for E-Learning? *Applied Sciences, 12*(13), 6582. https://doi.org/10.3390/app12136582.

Key competences for lifelong learning in the European schools. Retrieved from https://www.eursc.eu/BasicTexts/2018-09-D-69-en-1.pdf, (2018).

Kiili, C., Leu, D. J., Utriainen, J., Coiro, J., Kanniainen, L., Tolvanen, A., & Leppänen, P. H. T. (2018). Reading to learn from online information: Modeling the factor structure. *Journal of Literacy Research, 50*(3), 304–334. https://doi.org/10.1177/1086296X18784640

Kim, H. S., Ahn, S. H., & Kim, C. M. (2019a). A new ICT literacy test for elementary and middle school students in Republic of Korea. *The Asia-Pacific Education Researcher, 28*(3), 203–212. https://doi.org/10.1007/s40299-018-0428-8

Kim, H. S., Ahn, S. H., & Kim, C. M. (2019b). A new ICT literacy test for elementary and middle school students in Republic of Korea. *Asia-Pacific Education Researcher, 28*(3), 203–212. https://doi.org/10.1007/s40299-018-0428-8

Kline, R. B. (2016). *Principles and practices of structural equation modelling* (4th ed.). Guilford Press.

Kong, S. C., Wang, Y. Q., & Lai, M. (2019). Development and validation of an instrument for measuring digital empowerment of primary school students. In *CompEd 2019 - proceedings of the ACM conference on global computing education* (pp. 172–177). https://doi.org/10.1145/3300115.3309523

Laanepere, M. (2019). *Recommendations on assessment tools for monitoring digital literacy within UNESCO's digital literacy global framework* (1st ed.). UNESCO Institute for Statistics (UIS). https://doi.org/10.15220/2019-56-en

Le Deist, F. D., & Winterton, J. (2005). What is competence? *Human Resource Development International, 8*(1), 27–46. https://doi.org/10.1080/1367886042000338227

Lee, L., Chen, D. T., Li, J. Y., & Lin, T. B. (2015). Understanding new media literacy: The development of a measuring instrument. *Computers and Education, 85*, 84–93. https://doi.org/10.1016/j.compe du.2015.02.006

Muthén, L. K., & Muthén, B. O. (1998). *Statistical analysis with latent variables user's guide*. Retrieved from www.StatModel.com. (Accessed 12 January 2022).

National core curriculum for basic education. (2021). Finnish national agency for education. https://www.oph.fi/en/education-and-qualifications/national-core-curriculum-basic-education.

Palczyńska, M., & Rynko, M. (2021). ICT skills measurement in social surveys: Can we trust self-reports? *Quality and Quantity, 55*, 917–943. https://doi.org/10.1007/s11135-020-01031-4

Pedaste, M., Kalmus, V., & Vainonen, K. (2021). Dimensions of digital competence and its assessment in basic school. Eesti Haridusteaduste Ajakiri. *Estonian Journal of Education, 9*(2), 212–243. https://doi.org/10.12697/EHA.2021.9.2.09.

Põhikooli riiklik õppekava. (2021). *National curriculum of basic schools*. Retrieved from https://www.riigiteataja.ee/akt/123042021010. (Accessed 12 November 2021).

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

Reichert, F., Zhang, D., Law, N. W. Y., Wong, G. K. W., & de la Torre, J. (2020). Exploring the structure of digital literacy competence assessed using authentic software applications. *Educational Technology Research & Development, 68*(6), 2991–3013. https://doi.org/10.1007/s11423-020-09825-x

Saltos-Rivas, R., Novoa-Hernández, P., & Serrano Rodríguez, R. (2022). How reliable and valid are the evaluations of digital competence in higher education: A systematic mapping study. *Sage Open, 12*(1). https://doi.org/10.1177/21582440211068492

Scherer, R., Siddiq, F., & Teo, T. (2015). Becoming more specific: Measuring and modeling, teachers' perceived usefulness of ICT in the context of teaching and learning. *Computers & Education, 88*, 202–214. https://doi.org/10.1016/j.compedu.2015.05.005

Schneider, K. (2019). What does competence mean? *Psychology, 10*(14), 1938.

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

Silm, G., Pedaste, M., & Täht, K. (2020). The relationship between performance and test-taking effort when measured with self-report or time-based instruments: A meta-analytic review. *Educational Research Review, 31*, 100335.

Smahel, D., Machackova, H., Mascheroni, G., Dedkova, L., Staksrud, E., Ólafsson, K., et al. (2015). *EU Kids online 2020: Survey results from 19 countries*. EU Kids Online. https://doi.org/10.21953/lse.47fdeqj01ofo

Sukk, M., & Soo, K. (2018). *EU Kids Online'i Eesti 2018. aasta uuringu esialgsed tulemused [Preliminary results of the EU Kids Online Estonia 2018 survey]*. Retrieved from https://sisu.ut.ee/sites/default/files/euko/files/eu_kids_online_eesti_2018_raport.pdf. (Accessed 12 January 2022).

The Organisation for Economic Co-operation and Development. (2005). The definition and selection of key competencies. Executive summary. *The DeSeCo Project*. Retrieved from February 24 2022 http://www.oecd.org/dataoecd/47/61/35070367.pdf.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 425–478.

Vuorikari, R., Kluzer, S., & Punie, Y. (2022). *DigComp 2.2: The digital competence framework for citizens*. Luxembourg: EUR 31006 EN, Publications Office of the European Union. https://doi.org/10.2760/115376JRC128415, 978-92-76-48882-8.

Vuorikari, R., Punie, Y., Carretero, S., & Van Den Brande, L. (2016). *DigComp 2.0: The digital competence framework for citizens*. Jrc-Ipts. https://doi.org/10.2791/11517

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen, & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–66). Göttingen: Hogrefe.

White, R. H. (1959). Motivation reconsidered: The concept of competence. *Psychological Review, 66*, 297–333.

Wilson, M., Gochyyev, P., & Scalise, K. (2017). *Modeling data from collaborative assessments: Learning in digital interactive social networks*. https://doi.org/10.1111/jedm.12134

Wilson, M., Scalise, K., & Gochyyev, P. (2019). Domain modelling for advanced learning environments: The BEAR Assessment System Software. *Educational Psychology, 39*(10), 1199–1217. https://doi.org/10.1080/01443410.2018.1481934

Wing, J. M. (2006). Computational thinking. *Communications of the ACM, 49*(3), 33–35.

Yang, J., Tlili, A., Huang, R., Zhuang, R., & Bhagat, K. K. (2021). Development and validation of a digital learning competence scale: A comprehensive review. *Sustainability, 13*(10), 5593. https://doi.org/10.3390/su13105593

Zhang, H., & Zhu, C. (2016). A study of digital media literacy of the 5th and 6th grade primary students in Beijing (2), 181–208 *Asia-Pacific Education Researcher, 25*(4), 579–592, 10. 1007/s40299-016-0285-2.