Assessment of students' digital competences in primary school: a systematic review

Godaert Eline[a], Aesaert Koen[b], Voogt Joke[c], van Braak Johan[a]

[a] Department of Educational Studies, Ghent University, Ghent, Belgium

[b] Educational Effectiveness & Evaluation, KU Leuven, Leuven, Belgium

[c] Department of Educational Studies, University of Amsterdam, Amsterdam, The Netherlands

Abstract

Although there is a growing body of literature that recognizes the importance of being digitally competent today, there have been few empirical investigations into the assessment of primary school students' digital competences. This study presents a systematic review of the empirical research on the assessment of primary school students' digital competences. In total, 14 studies were selected and reviewed. The purpose of this review is twofold. First, the areas of digital competence that were measured by the assessment instruments were labeled according to the European Digital Competence Framework. Results showed that most studies evaluated digital competences as 'information and data literacy', 'communication and collaboration' and 'creation of digital content'. Less attention is paid to the assessment of the competence areas 'safely and responsible use' and 'problem solving'. Besides, the emphasis of most instruments is rather on the measurement of skills. Attention towards knowledge and attitudes, as important aspects of competences, remain underexposed. Second, an analysis of the provided evidence of the quality of assessment instruments for measuring primary school students' digital competences is given, based on the Research Centre for Examination and Certification Framework. Results indicate different approaches to increase the quality of the assessment instruments, but there is generally poor reporting of the psychometric properties of the tests. Based on these results, suggestions for further research and practice are discussed.

Keywords*: primary education; digital competences; assessment; systematic review; DigComp 2.1. framework; RCEC framework*

Assessment of students' digital competences in primary school: a systematic review

1. Introduction

In recent decades, national and international education policy has stressed the importance of young learners becoming digitally competent. This has resulted in the establishment of numerous reference frameworks that governments use as input to introduce digital competences in their national curriculum (Ilomäki et al., 2016), such as the Technology & Engineering Literacy Framework (National Center for Education Statistics [NCES], 2014), or the DigComp 2.1. Framework. The latter framework, which describes digital competences as a necessary transversal key competence in terms of lifelong learning (Carretero et al., 2017), is widely used in the European Union. Digital competences are generally defined as a multi-layered and complex construct, in which higher-level thinking skills are supported by technical and practical digital knowledge and skills (Aesaert & van Braak, 2015; Carretero et al., 2017; Ilomäki et al., 2016; Siddiq et al., 2016).

It is usually assumed that citizens' digital competences should be developed at school at an early stage (Aesaert et al., 2013; Ilomäki et al., 2016) as young people are, at that point, already confronted with a lot of digital information (e.g. watching movies, playing games) and devices (computers, TVs, tablets, and smartphones) (Juhaňák et al., 2019). However, young children generally have little awareness of the risks associated with the use of digital technology (Chaudron et al., 2018). In other words, both the frequency of exposure and the risk in the use of technology is already strongly present with young children (Ainley, 2018; Morgan, 2020). Previous research (e.g. Chaudron et al., 2018) suggests that the use of digital technology at an early age has a significant influence on the development of students' digital competence and upon the development of their autonomy in digital technology use. Consequently, primary education should therefore be seen as a natural place to develop digital competences (Aesaert et al., 2013; Kong et al., 2019).

Before children enter primary school, it is expected that their out-of-school-repertoire of digitally related knowledge, skills, and attitudes expands as schools integrate the development of these digital competences in their curriculum. However, the study of Aesaert & van Braak (2015) shows that at the end of primary school, students are still having trouble with certain digital competences, such as judging, and assessing the relevance of digital information, providing digital content in a socially acceptable and comprehensible way, and performing complex searching activities. Similarly, Claro et al. (2012) report that communicating, such as writing an e-mail, requires a different level of digital competence compared to searching digital information. This undermines the widely accepted belief in the existence of digital natives as this generation is not that computer savvy as it is assumed (Aesaert & van Braak, 2018; Kirshner & De Bruyckere, 2017).

Along with the increasing focus on digital competences as formal learning objectives, reliable and valid assessment instruments are needed to monitor students' progress in mastering this competence (Siddiq et al., 2016) and to allow schools to analyze students' digital competences and adapt the curriculum to these needs (Calvani et al., 2012). Although research on the assessment of digital competences has been growing in recent years, it still remains scarce (Siddiq et al., 2016). Moreover, it is characterized by two persistent issues: a conceptual and a methodological challenge. *Conceptually*, it is challenging that most existing assessment instruments do not fully cover the complexity of digital competence as a large majority of them focuses mainly on one or two sub-competences such as information and data literacy, communication and collaboration and digital content creation, while safely and responsible use and problem solving remain underexposed (Siddiq et al., 2016). In other words, it is an added value to also highlight the underexposed sub-competences of digital competence in assessment instruments. *Methodologically*, in a first interpretation, conceptual ambiguities need to be minimized because they can threaten the content and construct validity of digital competence (Blömeke et al., 2015; Kahn & Zeidler, 2016; Weinert, 2001). It is impossible to operationalize a construct in a valid way if there is conceptual ambiguity. In a second interpretation, research indicates that most instruments measure digital competences in an indirect way (e.g. ICT self-efficacy) (Aesaert et al., 2014; Aesaert & van Braak, 2018; Merrit et al., 2005; Siddiq et al., 2016) or use student observations in small samples. Although indirect measures are easily administered to large samples, questions have been raised about the validity and reliability of these methods. Validity problems are visible because such instruments depend on students' ability to make judgements about their own competences (Aesaert & van Braak, 2015). The reliability is also questionable because such assessment instruments only provide proxies of the actual competences (Law et al., 2009; Litt, 2013). While direct assessment based on observations, offers a suitable alternative to assess digital competences in a valid way, they suffer from practical problems such as being time-consuming, being expensive, difficult to replicate and to administer in large samples. Third, there is generally a lack of explicit documentation of the psychometric characteristics of assessment instruments (Siddiq et al., 2016). Yet, the presence of this information is viewed as a central concern for the trustworthiness of the instrument and the credibility of the results obtained by using an assessment instrument (Wilson & Sloane, 2000).

In a review study of Siddiq et al. (2016) the scarcity of high-quality assessment instruments for primary school students is already highlighted. In this study, Siddiq et al. (2016) aim to provide future directions for the assessments of digital competences in primary and secondary education based on a systematic review of the literature from 1990 until 2014. The present study can be seen as a follow-up study in which the conceptual and methodological challenges of assessing digital literacy is further investigated. Because of increased attention for digital literacy in many countries (Fraillon et al., 2018) we argue that it is useful to investigate recent studies (2014 - 2020) on the assessment of students' digital competences, based on two theoretical frameworks, the Digital Competence Framework (Carretero et al., 2017) and framework of the Research Centre for Examination and Certification (RCEC) framework (Hemker et al., 2019; Sanders et al., 2016). We focus in this review on primary education as digital competences need to be developed at an early age. The investigation of both conceptual and methodological challenges is a continuing concern. The lack of correspondence between indicators (operationalized constructs) and the constructs (definitions or concepts) are problematic with respect to the validity of measurement instruments. As a multitude of characteristics is to be considered in digital competence assessment, further research is needed to examine valid assessment of digital competences and to identify the characteristics of available assessment instruments regarding their validity and reliability.

To address the challenges mentioned above, the present study systematically reviews the literature concerned with instruments for assessing primary school students' digital competences. The two aims of this study are 1) to identify the measured components of digital competences, using the European Digital Competence Framework (DigComp 2.1.) (Carretero et al., 2017) and 2) to identify actions taken to pursue the quality of an assessment instrument, using the Research Centre for Examination and Certification (RCEC) framework (Hemker et al., 2019; Sanders et al., 2016).

2. Theoretical framework

2.1. Defining digital competences

The concept of digital competences is widely used in scientific publications and policy documents. Related concepts (e.g. ICT fluency, information literacy, ICT literacy, digital literacy, technological literacy, ICT competencies, internet skills) are interchangeably used to describe a set of knowledge, skills and attitudes related to the use of technology (Ala-Mutka, 2011; Ilomäki et al., 2016; Law, et al., 2009). Most of today's definitions support the idea that digital competences are not only about technical capabilities (see e.g. Aesaert et al., 2014; Bawden, 2001; Huggins et al., 2014; Martin & Madigan, 2006), but also concerns the ability to use, understand, evaluate and analyze information in multiple formats from a variety of digital sources and apply digital technologies for creation and communication (Casey & Bruce, 2011; Ilomäki et al., 2016, Kim et al., 2019). However, some researchers have criticized the conceptualization of digital competences and call it a "loose concept: one that is not well-defined, still emerging, with meanings varying based on users from different approaches" (e.g. Ilomäki, et al., 2016, p.656; Le Deist & Winterton, 2005).

2.2. Digital competences according to the DigComp 2.1. Framework

Considering the complexity of the construct digital competence and the need to understand the processes and abilities that lead to successful real-life situations, it becomes increasingly important that assessment is based on theoretical models (Klieme et al., 2008). Both national (e.g. NCES, 2018) and international (e.g. Fraillon et al., 2013) frameworks exist to clarify the conceptualization of digital competences and to describe the outcome measures of students' capacities to use technology. The emphasis on the growing importance of digital competences contrasts with a lack of conceptual clarity and insufficient insight into the various components of digital competence. As a progression for conceptual clarity, in 2016 the revised DigComp 2.1. Framework was created (Carretero et al., 2017), based on the analysis of 15 frameworks for digital competences that have been synthesized and reviewed by relevant stakeholders (e.g. PIAAC, PISA 2013 2015, ICILS 2013) and experts (Siddiq et al., 2016). DigComp 2.1. describes which competences are required today to use digital technologies in a confident, critical, cooperative, and creative way to effectively perform in activities related to work, learning, and participation in our digital society. According to the authors, this framework provides a clear and complete blueprint for conceptualizing digital competences and their included components. In this study, we use the operationalization of DigComp 2.1. to define 'digital competence', over other frameworks such as ICILS (Fraillon et al., 2018), because it provides generalizability across different age groups and regions. The framework can also be used for comparisons on an international level (Wild & Heuling, 2021). According to the Digcomp 2.1. framework (Carretero et al., 2017), digital competences are seen as not just technical, but as key transversal skills, knowledge, and attitudes. In this framework basic technical skills, e.g. turning a computer on and off, are less explicitly addressed.

The Digcomp 2.1 framework consists of four dimensions (Carretero et al., 2017) of which the first two dimensions are important for this research. Both dimensions consist of five competence areas and the corresponding competences. The first dimension consists of five areas of competence, i.e., information and data literacy; communication and collaboration; creation of digital content; safety; and problem solving. The second dimension consists of the specific competences covered by a competence area and their description. The other dimensions of DigComp 2.1. consist of the proficiency levels and examples of use and supports the development of learning and training materials. As we focus on the conceptualization of digital competences, these dimensions are not considered in this study. An overview of the first two dimensions of the framework is shown in Appendix C.

The first area of competence, *information, and data literacy*, refers to the ability to browse, search, filter, evaluate and manage data, information, and digital content. *Communication and collaboration*, is the next overarching competence area and contains the competences interacting, sharing and collaborating through digital technologies, engaging in citizenship through digital technologies, netiquette (applying behavioral norms while using digital technologies) and managing digital identity. The third area, *digital content creation*, refers to developing, integrating, and re-elaborating digital content, understanding how copyrights and licenses apply to data, digital information and digital content and programming (in other words: to solve problems or perform given tasks by developing an algorithm for a computing system). The fourth area of the DigComp 2.1. Framework is *safety,* which refers to protecting devices, personal data, and privacy in digital environments, being able to protect the physical and psychological well-being while using digital technologies and being aware of the environmental impact of digital technologies and their use. The final area, *problem solving*, is about being able to solve technical problems, to identify digital needs and technological responses, to use digital technologies creatively, and to identify digital competence gaps (Carretero et al., 2017).

2.3. Measuring digital competences

2.3.1. Assessment of students' digital competences

To monitor primary school students' development in mastering this complex competence, reliable and valid instruments are needed (Siddiq et al., 2016). Until now, most test developers have focused on assessing the digital competences of secondary school students (e.g. Claro et al., 2012; Hatlevik & Christophersen, 2013), whereas the assessment of primary school students' digital competences has been barely investigated (Kong et al., 2019; Siddiq et al., 2016).

First, the assessment of competences is hindered if there is no clear conceptualization of what we are aiming to measure. In that case, both the construct validity and the content validity are compromised (Leutner et al., 2008). As noted by Blömeke et al. (2015), the assessment of competences is also challenging because competences are context- and situation-related. Therefore, it is not feasible to develop a single test that covers all (Aesaert & van Braak, 2015; Siddiq, et al., 2016). In general, it requires an assessment context that is as authentic as possible. The assessor must face the difficult task of designing real (class) situations and measuring observable behavior (Koeppen et al., 2008; Straetmans & Sanders, 2001).

However, reliability requirements typically imply a large sample which mainly leads to indirect measurement or assessment tools such as selected-response assessments (objective measurement), which can be quickly administered and scored. A widely used method of indirect measurement is assessment through students' self-reported confidence about their performance in using, evaluating, and analyzing digital technology (Aesaert & van Braak, 2018; Litt, 2013; Siddiq et al., 2016). However, such self-reporting instruments do not provide an accurate measure of digital competences (Ballantine et al., 2007; Merritt et al., 2005; van Deursen & van Dijk, 2011). Students might overestimate or underestimate their own digital competences (Porat et al., 2018; Spisak, 2018), making this indirect way of measuring susceptible to validity problems (Merrit et al., 2005). Another type of indirect measurement is objective measurement where traditional item formats are used, for example multiple choice questions, dragging and dropping questions, or simple correct-wrong questions. One of the challenges to objective measurement is that these traditional formats do not represent how students use technology in practice. They do not fully capture the authenticity of the contexts in which students normally demonstrate their digital competences (Aesaert et al., 2014; Siddiq et al., 2016).

To address the shortcomings of indirect measurement, an increasing number of researchers (see e.g. Aesaert et al., 2014; Ainley et al., 2016; Heitink, 2018; Kaarakainen et al., 2018; Porat et al., 2018; van Deursen & van Dijk, 2016) focus on a more direct and innovative way of assessing digital competences. This means that the level of digital competences is grounded in the analysis of the actions carried out and directly monitored. More specifically, a researcher observes students while they complete real-life tasks on a computer, which in turn are analyzed (Aesaert & van Braak, 2018). The difficulty of these measurements, in the context of large-scale assessment, is to digitize the observations in a reliable logging and encoding system (Aesaert & van Braak, 2015; Litt, 2013). New trends in international large-scale assessment are characterized by two attributes: the test content should reflect real-world use of digital technologies and the test should make use of dynamic and multimodal opportunities by computer-based environments, e.g. hyperlinks or watching videos (Fraillon, 2018). An example in low secondary education is ICILS (International Information and Computer Literacy Study 2013; 2018), using a closed computer-controlled environment to measure the digital competences of secondary school students (Fraillon et al., 2014; Fraillon, et al., 2018). New challenges include ensuring task independence, securing construct validity, maintaining standardized test administration procedures, and providing automatic feedback from the computer-based environment (Fraillon, 2018).

2.3.2. The quality of an assessment instrument

Next to the type of assessment instrument, reporting the quality of an assessment instrument is seen as a central concern for the validity and reliability of the instrument and the credibility of the results obtained through its use (Wilson & Sloane, 2000). The results in the review study by Siddiq et al. (2016) indicate a poor reporting of the psychometric properties of the tests (Wilson & Sloane, 2000). Only if the quality of test is high enough to produce accurate and unbiased indicators of students' digital competences, the results can be used for intended purposes, e.g. adaptive teaching or measuring school effects.

Quality assurance systems are available for a long time. These systems have their origins in evaluating psychological tests (e.g. COTAN, in Evers et al. (2010)). The RCEC-framework is an example of an analytical assessment system that specifically focuses on the evaluation of the quality of educational tests, and examinations (Sanders et al., 2016) and is in large part inspired by the COTAN-system of the Dutch Committee on Tests and Testing (COTAN) (1969).The RCEC-framework contains six criteria and provides a set of questions to be answered by giving a rating on a three-point scale: insufficient—sufficient—good (see Appendix D for an overview). According to the RCEC-framework (Sanders et al., 2016) an assessment instrument firstly should only have one purpose and use. In other words, it should be clear what exactly is tested and why it was tested. The second criterium considers the quality of the assessment material of the tasks and exercises and the instructions for the participants and assessors. Third, representativeness includes both the content, the composition, and the level of difficulty of the test or exam. The fourth criterium, reliability, is about whether the scores that students achieve on a test or exam are trustworthy. The reliability can be quantified with a reliability coefficient, the percentage of misclassifications and the standard measurement error. Another criterium is how the norm (or standard) of the test has been determined and maintained. Part of this is the determination of the benchmark, such as the boundary between 'passed or sufficient' and 'failed or insufficient'. Finally, administration and security provide the reviewer insight into the data storage, software, and security of the system. The rationale for selecting the RCEC-framework is that it is specifically tailored for evaluating the quality of educational tests and the provided guidance questions invite to consider assessment instruments in depth. It improves understanding of both the substantive and the organizational aspects of assessment, as well as the technical aspects. The framework not only focuses on the intrinsic quality of the instrument itself, but also on the development of the assessment instrument and how accurately the development process was conducted. Moreover, the RCEC-framework offers an approachable, clear, and comprehensive overview of quality indicators not only to review assessment instruments, but also to use as a guide to develop new assessment instruments. In sum, the RCEC-framework is a valuable review system to report the quality of the tests to foster the development of improved assessment of digital competence.

Based on the above findings, this review aimed to synthesize the existing research on the measurement of digital competences of primary school students by investigating content and test characteristics of existing assessment instruments, with DigComp 2.1. and RCEC as frames of reference. The aims can be translated to two research objectives:

- RO 1: To provide an overview of the measured components of digital competences in existing assessment instruments for primary school students
- RO 2: To provide an overview of the available evidence of the quality criteria of assessment instruments for measuring primary school students' digital competences

To reach this aim, we first investigate which components of students' digital competences are measured in existing studies, based on the DigComp 2.1. Framework. Second, we analyze how the different studies take actions to monitor the quality of the assessment instrument based on the RCEC-framework.

3. Method

To investigate which areas of digital competences are measured by existing assessment instruments and to analyze the provided evidence of the quality of assessment instruments, the standard steps of a systematic review process were followed. Published studies were identified using a search strategy based on the recommendations of the PRISMA (Preferred Reporting Item for Systematic reviews and Meta-Analyses, Moher et al., 2010) statement. First, studies were searched through scanning relevant databases. Next, publications were only included in the analysis if they met the inclusion and exclusion criteria. Third, the data from the selected studies were abstracted and analyzed. Analysis was based on two frameworks to answer both research questions: DigComp 2.1. Framework (RO 1) and RCEC-framework (RO 2). All steps of the review process were discussed in an interdisciplinary research team consisting of four experts within the field of educational measurement, digital curricula, and assessment of digital competences.

3.1. Literature search

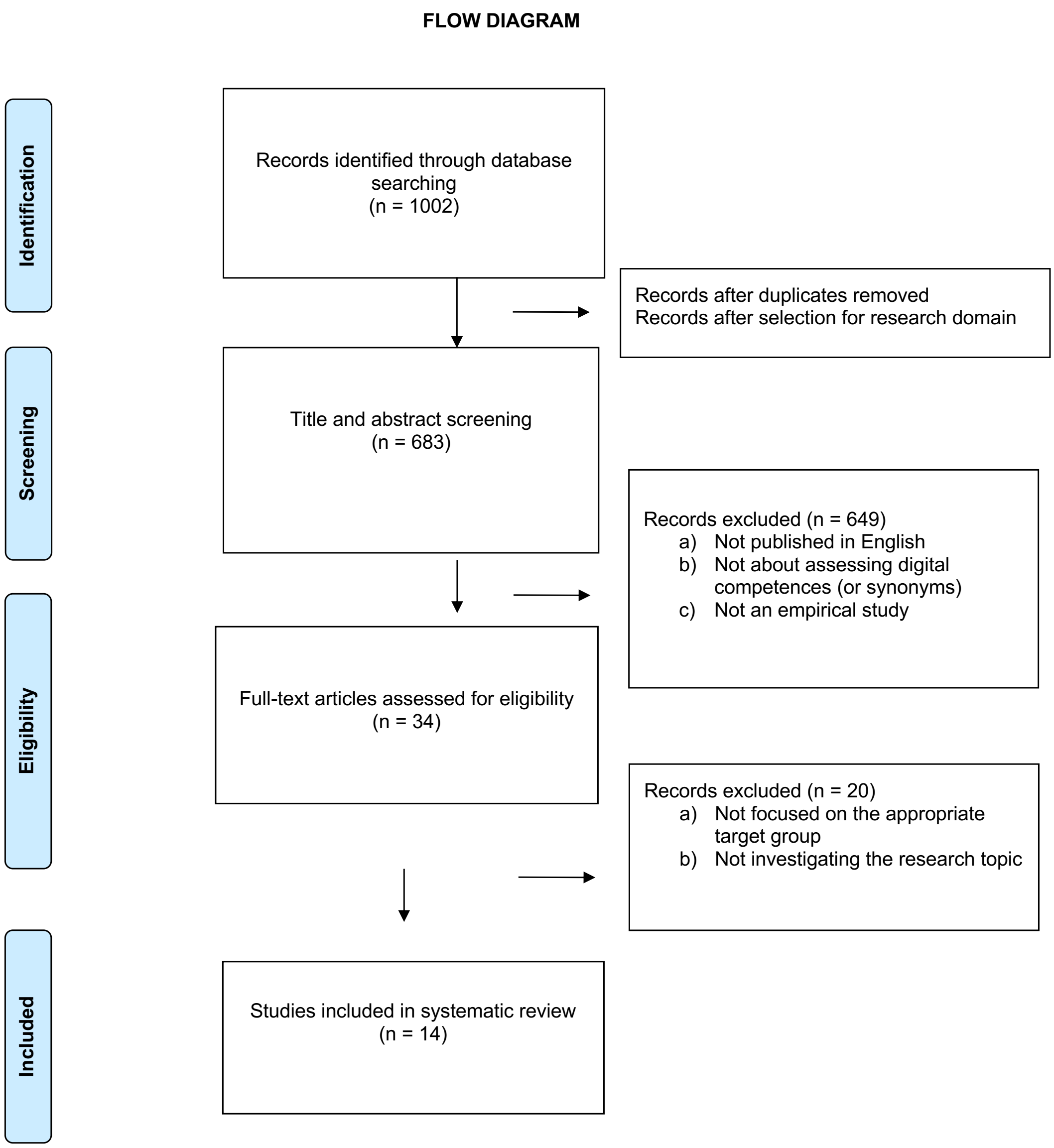
During the period of January-May 2020 a literature search was conducted using the educational research databases Web of Science (WoS), Google Scholar and Scopus. The search was performed using a Boolean logic combining key concepts from the research question. Search strings consisted of a combination of terms that were scanned for in titles, keywords, abstracts, and full text using the 'OR' operator between the synonyms and the 'AND' operator between the different searches. The search engine can be found in Appendix A. The following search terms were used:

- **Digital competence:** ICT competence, digital literacy, ICT literacy, digital capacities, ICT capacities, digital knowledge, digital skills, digital attitudes, ICT knowledge, ICT skills, ICT attitudes, ICT fluency, information and computer literacy, technological competence
- **Measure:** assessment, measurement, instrument/test development, assessing, measuring, testing
- **Primary education:** primary education, primary school, elementary school, students, pupils, K-6 and alternatives of K-6 (1st-grade, first-grade, grade 1, grade one, 2nd-grade, second-grade, grade 2, grade two, 3rd-grade, third-grade, grade 3, grade three, 4th-grade, fourth-grade, grade 4, grade four, 5th-grade, fifth-grade, grade 5, grade five, 6th- grade, sixth-grade, grade 6, grade six)

Literature was only consulted if it was published after 2014 to avoid overlap in time with a review study published by Siddiq et al. (2016) (review study on measuring instruments for digital competences (selection from 1990 - 2014)). Moreover, only search terms in English were used, to obtain English-language research from academic journals.

3.2. Study selection

To select the different studies, a protocol based on the recommendations of the PRISMA statement was used (Moher et al., 2010). The study selection process occurs in four phases represented in a flow diagram (see Figure 1).



From: Moher D, Liberati A, Tetzlaff J, Altman DG, The PRISMA Group (2009). *P*referred *R*eporting *I*tems for *S*ystematic Reviews and *M*eta-*A*nalyses: The PRISMA Statement. PLoS Med 6(7): e1000097. doi:10.1371/journal.pmed1000097

For more information, visit www.prisma-statement.org.

Figure 1 *Overview of the Search Protocol based on the PRISMA Statement*

The initial search of the three databases consisted of 1002 studies. After removing duplicates and refining for the research domain, 683 records remained. The screening phase included two phases. In the first screening phase, the first author screened the remaining studies by judging the title and abstract. Studies were excluded if they were not published in English, and were not reporting an empirical study (e.g. editorial). Second, the remaining 34 studies were fully read. After fully reading all these studies, some were retained based on the inclusion and exclusion criteria. Studies were excluded if a) they did not focus on the appropriate target group, namely: primary school students (up to 13 years) and b) the research topic: assessment of digital competences.

Based on these eligibility criteria, the first author selected 20 relevant studies. After fully reading the articles, 14 studies remained. In case of doubt, the members of the interdisciplinary team assessed the relevance of these studies. The following methodological criterion was considered for large-scale samples to include studies of sufficient quality: samples of at least 150 respondents. We have set the minimum to 150 respondents because we consider this number to be plausible for the quality of the instrument. We do acknowledge that the larger the sample, the more accurate the results and the higher the quality of the data.

The following study characteristics were formally abstracted: name of the test, authors, publication year, country, grade/age, type of test, measured competences, duration of the test, paper-and-pencil test or computer-based assessment, scoring, aspect of competence (skills, knowledge, or attitudes) and quantity of items. An overview of the descriptive information of the papers can be found in Appendix E. To capture all important information of the selected studies, each of the 14 studies were coded by the first author using a developed coding scheme (data extraction form, see Appendix B).

3.3. Data deduction and analysis

First, the studies were clustered by categorizing the competences that are (or are intended to be) measured by the testing instrument according to the five competence areas of the DigComp 2.1 Framework. To illustrate the following procedure, a conceptual framework discerning different components of digital media literacy (the construct the authors, Zhang and Zhu (2016), aim to measure) is compared to the underlying definitions to the definitions of the different components of the second dimension of the DigComp 2.1. Framework. In case of compliance between the content of both definitions, we marked it in Table 2 (see result section) by a cross. If some information based on the assessment instrument did not fit with the characteristics of this framework, extra information was added in Table 2 (see e.g. 'additional competences': basic technical skills or focus on attitudes).

Second, the reported quality actions of all assessment instruments were classified according to the RCEC-framework. All the characteristics of the assessment instruments were listed in Table 3, following the guiding questions of the RCEC-framework (See Appendix D).

During a third phase, an independent coder reviewed the categorization of all 14 studies. In case of disagreement, a consensus was reached between the reviewers by discussion.

In a fourth phase, for each competence or sub-competence of the DigComp 2.1. Framework, and for all components of the RCEC-framework, all relevant findings were described to make comparisons and joint interpretations.

In the final phase, the findings were summarized and compared to provide an overview of the measured components of digital competences and quality actions in existing assessment instruments.

4. Results

4.1. Descriptive information about the assessment instruments

The selected assessment instruments have been developed in ten different countries over three continents: Asia, Europe, and Australia. All studies were published between 2014 and 2020. The sample sizes varied between 150 and 40072 pupils. Most studies (n = 11) cover the assessment of digital competences in the last (two) year(s) of primary education. Three studies focus on both primary and secondary education. The average age of the pupils was 12 years. Most studies examine digital skills, with or without an extra section on knowledge questions, and measure attitudes via a separate questionnaire. Concerning the test format, five studies use only self-reporting tools, one study uses only selected-response assessment formats (objective assessment) based on multiple choice questions and constructed answer options and one study only integrates performance-based assessment (items are simulated and authentic software applications are used). Most studies (n = 8) combine different types of assessment, e.g. traditional assessment based on multiple choice questions and constructed answer options in combination with performance-based assessment. Performance-based tasks or objective assessment formats are usually dichotomously scored (correct or false) and self-reporting tools generally use Likert-type scales to score the different answers. Furthermore, the analysis suggests that computer-based assessment tends to be more prevalent than paper-and-pencil assessment. Some studies combine both. Finally, the number of items/questions/assignments ranges from 12 to 111 and the duration of the tests (from 8 to 100 minutes) varies highly. Table 1 gives an overview of the sample characteristics. A more comprehensive overview of the descriptive information is given in Appendix D.

Table 1: Overview of the sample characteristics

Author(s)	Publication date	Country	Data collection	Sample size	Grade	Age
Aesaert, Van Nijlen, Vanderlinde, van Braak	2014 2015	Belgium (Flanders)	2012	560	6	11 – 12
Asil, Teo, Noyes	2014	Singapore	2014	503	6	10-13
Jun, Han, Kim, Lee	2014	Korea	2012	40072	1 - 6	6-12
Kim, Kil, Shin	2014	Korea	2012	11767	4-6	9-11
Lee, Chen, Li, Lin	2015	Singapore	2014	574	4 - 11	10-17
Zhang & Zhu	2016	China	Not mentioned	796	5-6	9-13
Fernández-Montalvo, Peñalva, Irazabal, López-Goñi	2017	Spain	2011	309	6	11
ACARA	2018	Australia	2017	5,439	6	11-12
Heitink	2018	The Netherlands	2017	1036 (407 primary school 629 secondary school)	5-8	10-13
Kiili et al.	2018	Finland	2014-2015	426	6	12-13
Kim, Ahn, Kim	2019	Korea	2016	6383 primary school 9183 Middle school	4-9	10-15
Kong, Wang, Lai	2019	China	Not mentioned	328	4-6	Not mentioned
Chou & Chiu	2020	Taiwan	Not mentioned	666	6	11-12
Lazonder Walraven Gijlers Janssen	2020	The Netherlands	2016-2017-2018 (longitudal study)	151	5-6	8-11

4.2. Provided overview of the integrated digital competences into existing assessment instruments for primary school students (RO 1)

Figure 2 provides an overview of the measured competences according to the DigComp 2.1. Framework. In the following sections, we present the results addressing the different competence areas.

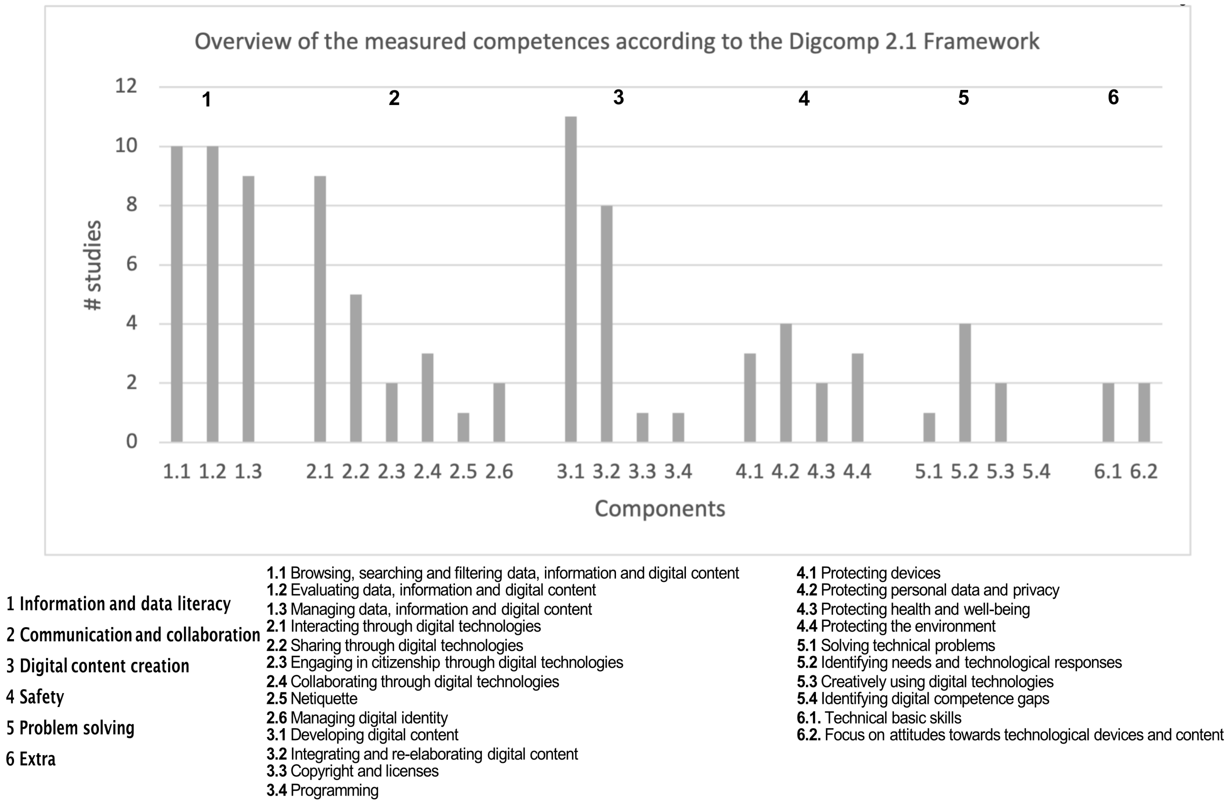


Figure 2 *Overview of the measured competences according to DigComp 2.1.*

4.2.1. Competence area 1: Information and data literacy

Eleven of the assessment instruments measured competences defined in the area of information and data literacy. Within this category, the first competence 'Browsing, searching and filtering data, information and digital' content was measured most frequently (1.1, see Table 1 for the meaning of these numbers; n = 10), together with 'Evaluating data, information and digital content' (1.2; n = 10) and 'Managing data, information and digital content' (1.3; n = 9). Eight of the eleven tests measured all three competences. All tests, except for one, were administered in an online test environment (two of them in combination with a paper-and-pencil test). Self-reported assessment (e.g. "*I am able to search for information or materials related to my homework"* Zhang & Zhu, 2016) were most common, whether or not in combination with traditional answer formats (such as multiple choice) or/and performance-based assessment (e.g. "*Pupils can use a search engine by entering more correct search terms derived from a task or question"*, Aesaert et al., 2014). Most studies focused on measuring skills (knowing how to use it and being able to use it; n = 7, e.g. "*The student can apply the filters and options of search engines during a search";* Heitink, 2018) or measuring both skills and knowledge/conceptual understanding (n = 3). Three instruments did not assess this competence area and one rather focused on the attitudes towards technical basic skills (Asil et al., 2014), one emphasized the meaningfulness and impact of digital technologies and students' creativity and competence belief (Kong, et al., 2019) and finally, one study (Fernandez-Montalvo et al., 2017) investigated the effectiveness of a digital literacy program with particular attention for internet use and risks (like managing and protecting digital identity).

4.2.2. Competence area 2: Communication and collaboration

Within the second competence area, most tests measured the competence 'Interacting through digital technologies' (2.1; n = 9). This sub-competence was mainly measured in an online environment using self-report, in which students indicate the extent to which they consider themselves "*capable of communicating, for example, with family, friends or classmates"* (see e.g. Zhang & Zhu, 2016, p. 590). In three studies, students are asked to *show* how they communicate (2.4) by for example an email as a part of a performance-based task (see e.g. Kiili et al., 2018). The same trends could be noticed for the competence 'Sharing through digital technologies' (2.2; n = 5). The other sub-competences were less presented in the assessment instruments. For example, the competence of 'Engaging in citizenship through digital technologies' (2.3) was addressed in only two of the testing instruments. This sub-competence was operationalized as 'Citizenship participation' (e.g. "*I am able to participate in a discussion about civic or political issues online";* Zhang & Zhu, 2016, p. 590*)* or 'Digital citizenship' (e.g. "*I can encourage my classmates and family members for safe, legal, and responsible information and technology usage*"; Chou & Chiu, 2020, p. 7). To our knowledge, 'Managing digital identity' (2.6) only appeared in two studies, measured in an indirect way, and focused on self-reported skills or knowledge (e.g. *"Do you know what digital identity is?"*; Fernandez-Montalvo et al., 2017, p. 9). Finally, the sub-competence 'Netiquette' (2.5) was only measured once in an indirect way as a part of a paper-and-pencil assessment (e.g. *"I know how to identify the type of content that can be shown (such as violent content or sexually explicit content").*

4.2.3. Competence area 3: Digital content creation

The competence area of digital content creation appeared in twelve studies. Eleven studies reported results regarding the sub-competence 'Developing digital content' (3.1). This sub-competence was mostly measured in an indirect way through self-report. The coverage of this sub-competence also revealed a large variation in e.g. performance-based assessment: from a basic level e.g. "*Students had to do text editing (adding bullet points to a text to present it more clearly)*" (Heitink, 2018) to rather higher-level skills e.g. *"creating a webpage to promote a sport event"* (Australian Curriculum, Assessment, and Reporting Authority [ACARA], 2018). Eight studies included the sub-competence of 'Integrating and re-elaborating digital content' (3.2) and it also varied in test formats on the one hand more authentic tasks like designing a slideshow (including the choice for content, structure, layout and format) (Lazonder et al., 2020) and on the other hand more indirect ways of measurement (e.g. self-report): "*I can use models and simulations to explore complex systems and issues*" (Chou & Chiu, 2020). There was only one study that covers the field of 'Copyright and licensing' (3.3), measured in a direct way using performance-based tasks about copyright and attribution requirements for content on the internet (ACARA, 2018); and 'Programming' skills (3.4) (e.g. "setting commands for a robot", Jun et al., 2014). There was not a single test that evaluates all competences within this area of competence.

4.2.4. Competence area 4: Safety

As illustrated in Table 1, the competence area 'Safety' was measured less than previous competences. Overall, six assessment instruments had focused on this competence. The first sub-competence 'Protecting devices' (4.1, n = 3) was conceptualized and operationalized as, for example, '*critical consuming literacy'* (*the individuals' ability to question, criticize and challenge the credibility of media contents*) (Lee et al., 2015) or "*I am able to protect my computer from being infecting by viruses (such as installing a virus scan program)*" (Zhang & Zhu, 2016) both measured as self-reported skills. The second sub-competence 'Protecting personal data and privacy' (4.2) was measured in three of the selected studies. For example, in an experimental study of Fernandes-Montalvo et al. (2017), focusing on measuring the effectiveness of a digital literacy program, the safe use of the internet and sharing or protecting personal data was the main topic. Finally, it is noteworthy that only two tests reported the measurement of the competence 'Protection of health and well-being' (4.3).

Besides, only a few tests included the assessment of competences in the field of 'Protecting the environment' (4.4, n = 3). For example, in a self-reporting questionnaire of Kong et al. (2019), focusing on attitudes and skills, this sub-competence was conceptualized as 'the impact of digital technologies' and measured as an attitude *'digital technologies make a difference in the world'.*

4.2.5. Competence area 5: Problem solving

Less than half of the studies had reported on measuring the competences within the 'problem solving' competence area (see Table 1). *Problem solving strategies*, conceptualized as 'Identifying needs and technological responses' (5.2) are for example operationalized in *"I can utilize digital tools to identify and define authentic problems for investigation"* (Chou & Chiu, 2020) and mostly measured in an indirect way using self-reporting questionnaires in an online environment. Furthermore, a limited number of tests (n = 2) concerned competence (5.3) 'Creatively using digital technologies'. It seems similar to the competence 'Digital content creation' but was more focused on the use of digital technologies in a creative way to solve problems. An example can be found in a study of Kong et al. (2019) as "*I can build new solutions to day-to-day problems with digital technologies*.". Competence 5.4 'Identifying gaps in digital competences' was not mentioned in any assessment instrument.

4.2.6. Additional competences

Four assessment instruments measured competences that are not fully covered by the DigComp 2.1. Framework or were difficult to categorize. Two of them were only focusing on 'Attitudes towards digital technologies and devices' (6.2) whereby students had to indicate the extent to which they agree or disagree (Likert Scale) with several statements e.g. "I use the computer to help me to do my work better" (Asil et al., 2014). Two assessment instruments also focused on 'Basic (or simple) technical skills' (6.2) like e.g. 'Understanding PC components' (Jun et al., 2014).

In sum, the competence areas 'information and data literacy' (n = 11), 'communication and collaboration' (n = 10) and 'creation of digital content' (n = 12) were the most frequently measured competence areas, while there were fewer tests that covered the areas of 'safety' (n = 6) and 'problem solving' (n = 5). Table 1 gives an overview of the categorization and results of all studies included in this review.

Table 2: Overview of measured competences according to the DigComp 2.1. Framework

	Aesaert et al., 2014	Asil et al., 2014	Jun et al., 2014	Kim et al., 2014	Lee et al., 2015	Zhang and Zhu, 2016	Fernandez-Montalvo et al., 2017	ACARA, 2018	Heitink, 2018	Kiili et al., 2018	Kim et al., 2019	Kong et al., 2019	Chou & Chiu, 2020	Lazonder et al., 2020
1.Information and data literacy														
1.1 Browsing, searching and filtering data, information and digital content	X		X	X		X		X	X	X	X		X	X
1.2 Evaluating data, information and digital content	X			X	X	X		X	X	X	X		X	X
1.3 Managing data, information and digital content	X			X	X	X		X	X	X	X		X	
2 Communication and collaboration														
2.1 Interacting through digital technologies	X		X	X	X	X		X		X	X		X	
2.2 Sharing through digital technologies	X				X	X		X					X	
2.3 Engaging in citizenship through digital technologies						X							X	
2.4 Collaborating through digital technologies					X	X		X						
2.5 Netiquette						X								
2.6 Managing digital identity						X	X							
3 Digital content creation														
3.1 Developing digital content	X		X	X	X	X		X	X		X	X	X	X
3.2 Integrating and re-elaborating digital content	X				X	X		X	X	X			X	X
3.3 Copyright and licenses								X						
3.4 Programming			X											
4 Safety														
4.1 Protecting devices					X	X		X						
4.2 Protecting personal data and privacy						X	X	X					X	
4.3 Protecting health and well-being							X	X						
4.4 Protecting the environment								X				X	X	
5 Problem solving														
5.1 Solving technical problems			X											
5.2 Identifying needs and technological responses				X							X	X	X	
5.3 Creatively using digital technologies												X	X	
5.4 Identifying digital competence gaps														
6 EXTRA														
6.1. Technical basic skills			X		X									
6.2. Focus on attitudes towards technological devices and content		X										X		

4.3. Provided evidence of the quality criteria of assessment instruments for measuring primary school students' digital competences (RO 2)

In the next step, the 14 studies were analyzed based on the RCEC-framework. How the test developers tried to meet the criteria to guarantee the quality of the assessment instrument are described below. An overview is provided in Table 3. See also Table 1 and Table 3 for more information about the quality of the instruments.

4.3.1. Purpose and use

First, all studies described the target group for which the assessment instrument is intended. However, in most cases, the indication of the target group was limited to stating the age or grade of the students (e.g. students at the end of primary school) for whom the test will be used, except for two studies (see Fernandez-Montalvo et al., 2017; Lazonder et al., 2020) for which conditions were formulated (e.g. students had to participate in an experimental intervention program). In this study, the test was only provided for those test participants who have completed a certain trajectory. Furthermore, no relevant prior knowledge of the participants in the selected study was described. Next, the main purpose of the instruments was clearly stated, e.g. measuring the effectiveness of a digital literacy program, assessing computational literacy, examining digital fluency. Finally, all assessment instruments were used to monitor current stages of skills and competence. In other words, all test instruments in this systematic review were low-stake and had no direct influence on the school career of the participants. Also, other practical implications were described, e.g. to evaluate and refine the effectiveness of the classroom initiatives/the curriculum, to inform teachers and schools or make them aware of the level of the pupils, to promote professional development, to compare with other countries. The results could also include theoretical implications, e.g. to revise and validate assessment instruments for future research and to investigate which variables do correlate with the dependent variable.

4.3.2. Quality of the test and examination material

This criterion concerns the quality of the test material: the tasks or assignments, the instructions for the candidates and assessors and the scoring system. Interestingly, two of the 14 studies did not provide information about the development or validation of the test and examination material. To optimize the content validity of the assessment instrument and make necessary adaptations (readability, adapted to the level of the target group), ten studies outlined a pilot study with a sample of students and/or experts as a starting point. For example, in a study by Lee et al. (2015) questions were revised based on think aloud, resulted in for example the avoidance of acronyms (e.g. “OMG” or “LOL”). In all the studies, all the participants completed the same questions aside from the order in which the questions were offered. In only a few of the studies, it was mentioned that the questions/items were provided a fixed sequence. Next, the data collection procedure was mostly described in detail and outlines a strict and standardized way of doing research. Furthermore, in all of the studies – except for two - an objective scoring system was used (e.g. correct/incorrect with a scoring rubric, Likert scale). If participants got open-ended questions, the answers were scored by using a rubric and the interrater reliability was calculated. Finally, to make sure that the participant understood the instructions, in four computer-based assessments an avatar was added to give instructions and to guide the participant through the online environment. For some of the assessment instruments, the instructions were being explained by a test administrator while giving an introduction before administering the test. The final part was strongly linked to the criterion 'administration and security'.

4.3.3. Representativeness

Representativeness refers to the degree to which the content, composition and difficulty of the test or exam align with the competences it intends to measure. In other words: does the instrument measure what it is supposed to measure? Most studies started from a domain analysis or literature review to make informed decisions on various aspects of the assessment instrument, e.g. strengthening operational decisions, finalizing measurement scales. In most of the studies (n = 10), a blueprint, competence profile, test matrix, or examples of items were presented to give an operationalization of the measured construct. In the remaining studies, you could determine the construct to be measured through the conceptual framework, but we did not get any insight into the operational part. For example, in Jun et al. (2014) the authors aimed to measure the construct ‘computational literacy’ consisting of three components: fundamental concepts (knowledge) and skills referred to students' knowledge of computers and skills in using them. In this study, only examples of items were provided, but no overview of all questions. In contrast, the study of Aesaert et al. (2014), provided a complete test matrix whereby each item is a translation of a digital skill. To make sure that the difficulty of the items was adjusted to the target group, most of the studies included a pilot study of students and experts. Next, most of the studies primarily checked the dimensionality, i.e., whether the authors checked that the test only measured the dimension it was intended to measure. This is a form of internal validity check (e.g. via advanced tests such as IRT, PCA and CFA). Besides, five studies checked whether the tests could be used for different groups by using measurement invariance analyses such as Differential Item Functioning (DIF). Finally, some authors measured the local independence and model-data fit to optimize the quality of the construct.

4.3.4. Reliability

The reliability could be quantified with a reliability coefficient and the standard error of measurement. It gives the test developer insight into the degree of measurement error of the instrument. There are several ways to measure the reliability of the instrument. In the selection of the studies, the following were provided: Cronbach's alpha, TIF, IFF, composite reliability, Kuder-Richardson Formula 20, Cohens Kappa. In all the studies, the reliability coefficient was given. Eight of the 14 studies calculated the Cronbach's Alpha for the different scales of the test instrument. Usually, the authors explained this score (good, acceptable, worse) and the scale was adjusted if necessary. In only three studies the reliability calculation procedure was explained (e.g. in Fernandez-Montalvo et al. (2017), the internal consistency (Cronbach's Alpha >.70) and the correlation between the items was investigated.)

4.3.5. Standard setting, norms, and equating

This criterion assesses how the norm (or standard) of the test has been determined and maintained. A distinction can be made between absolute (criterion-referenced) and relative (norm-referenced) standard-setting. In the case of absolute standard-setting, the results of the test indicate how a student performs a task or skill or understands concepts, concerning a fixed minimum standard or criterion (Ornstein & Hunkins, 2018), e.g. students must achieve 5/10 to pass. In the case of relative standard-setting, a students' performance on a test is compared with the results of other students who are his or her peers. This means that the norm or standard is variable (Ornstein & Hunkins, 2018). Only four of the 14 studies, contain absolute standards, calculated with the Bookmark method (n = 2) or the Angoff-method (n = 1), or prescribed according to the national minimum mastery level of the Australian curriculum (n = 1). The Angoff-method is a method whereby a group of experts determines how difficult it is to score correctly on the items to set the norm. The Bookmark-method is a method whereby a group of experts determines whether test items should be mastered in a specific level. In this process, items are ordered from easy to difficult. In a study of ACARA (2018) the provided performance standards are the national standards, representing the expectation of what typical YEAR 6 students should achieve by the end of the respective years. In these four cases, the standard-setting specialists are properly selected and trained. Interestingly, only two of the 14 studies, contain relative norms, whereby groups are compared by each other (e.g. differences between states and territories).

4.3.6. Administration and security

Information about the administration procedure and how to assure a secure administration, must be available for the proctor (the person who will be monitoring the test or exam). This information should be available in a clear form on paper or digitally. This is important in terms of a standardized data collection or to prevent fraud. 13 assessment instruments are administered in a computer-based way, whether combined with a paper-and-pencil assessment instrument. One study only contains a paper-and-pencil assessment. Only six studies report sufficient information on the administration of the test, e.g.: 'Each task starts with a pop-up window, containing the instruction for the task' (Aesaert et al., 2014); 'Test administrators were responsible for running the student tutorial supervising student participation and monitoring student progression between each section/module' (ACARA, 2018); Four assessment instruments are administered from a secured and closed online/offline computer-based environment in order to standardize the administration and to make a comparison between students possible. Information about data protection is generally not provided.

5. Conclusion and discussion

As primary school students still encounter problems with higher level digital competences (Aesaert & van Braak, 2015; Claro et al., 2012) and several researchers (e.g. Aesaert et al., 2013; Chaudron et al., 2018; Ilomäki et al., 2016) support the idea that digital competence should be developed at an early stage, this research focused on digital competences in the context of primary school. More specifically, this study investigated existing assessment instruments for measuring digital competences at the end of primary school. The assessment of digital competences among students seems necessary to further develop teaching practices, to launch digital programs to prepare the next generation for professional practice and to inform educational policy. According to Pagani et al. (2016) digital competences are pivotal for academic achievement in general. Moreover, they are strongly linked with other core competences like collaboration skills (NCERL, 2003). Next, studies have shown that there is a strong relationship between learning and assessment, implying that what is assessed influences what is learned (e.g. Alderson & Wall, 1993). Baartman et al. (2007) state that the quality of the assessment is seen as an important condition for the quality of education as a whole. In other words, reliable and valid tools are needed to assess the extent to which students master this complex competence and to monitor their progress. However, to date, only a few assessment instruments are available to measure primary school students' digital competences. Therefore, this study aimed to identify the measured components of digital competences, using the DigComp 2.1. framework (Carretero et al., 2017) and to identify actions taken to pursue the quality of an assessment instrument using the RCEC-framework (Hemker et al., 2019; Sanders et al., 2016).

To investigate the first research objective, the five competence areas of the DigComp 2.1. Framework were used to label the outcome variables (i.e., digital competences) of the selected studies. Our analysis indicates a relatively large amount of experience with the development of assessment instruments for the competence areas of information and data literacy, communication and collaboration and content creation. Although other competence areas such as safety (e.g. digital citizenship) or problem solving (e.g. identifying needs and technological responses) are also relevant competence areas for education, the knowledge on which the design of assessment instruments for these areas can be based, is not yet sufficiently developed on an international level. These results could be expected, as some competences are easier to assess in a valid and reliable way than others, or there is more expertise in measuring them (Bawden, 2001; Siddiq et al., 2016). Aesaert et al. (2014) and Siddiq et al. (2016) have pointed at a lack of feasibility in designing a test that covers all areas of competence. From the analysis of the competences based on the DigComp 2.1. Framework, we can conclude that the selected assessment instruments cover the different competence areas of this Framework.

This study has shown that there is only one competence that we could not include in the framework, namely basic technical skills. For the lack of a basic technical skills component, we assume that the authors in the selected studies expected that students already master the basic skills and usually proceed immediately to assess more advanced stages (and therefore do not focus on just testing basic skills). Indeed, basic technical skills are often considered to be a prerequisite to explore digital contexts in different ways (Li & Ranieri, 2010) or the first step to a more advanced level of digital competences (Ala-mutka, 2011; Markauskaite, 2007; Pöntinen & Räty-Záborszky, 2020).

Furthermore, the evidence of this review study shows that digital competences are mainly operationalized in terms of skills. A likely explanation is that it is more meaningful for students to complete functional assignments in relevant and recognizable contexts (see e.g. Heitink, 2018) than to gauge their knowledge about digital technologies and their use. This does not exclude an implicit survey of knowledge: the demonstration of skills is supported by different forms of knowledge (e.g. Taxonomy of Bloom; van de Kamp, 2012): factual knowledge (e.g. description of a search engine; difference between 1MB and 1GB), conceptual knowledge (the functioning of word processor), procedural knowledge (how to integrate a video file into a text file), metacognitive knowledge (which different search strategies can I apply?). These types of knowledge support an insightful application of digital skills. It is also possible to measure students' attitudes towards the use of digital technology as interest in and attitude towards technology is an important factor in improving learning outcomes and making educational decisions (Knezek & Christensen, 2008). The DigComp 2.1. framework presents an overarching definition and framework which contributes to conceptual quality in what digital competence means. However, our review shows that none of the studies complied with all components of the DigComp 2.1. framework, mainly measure skills and not knowledge and attitudes. Moreover, the framework, used in this study, does not discuss the position of basic technical skills. Realizing conceptual clarity, an important goal for realizing content and construct validity when measuring digital competence, continuous to be challenging.

To address the second research objective, the RCEC-framework was applied, covering six criteria to give an overview of provided evidence to ensure quality of the assessment. This review states the reporting of the key characteristics and efforts to promote the quality of the assessment, is not consistently extensive or qualitative. In this respect, we affirm the conclusion of Siddiq et al. (2016) concerning the limited reporting of both reliability and validity arguments, and thereby challenging the evaluation of replicability and minimizing the possibility of using the tools in future studies (Duncan et al., 2014). All studies in this review do have a clear purpose and use (i.e. measuring a certain construct or competence), but how the test was developed and how the authors tried to guarantee the quality is not always described comprehensively. However, this research exposes different approaches to increase the quality, reliability, and representativeness of the instrument, but there are also studies which for example do not perform preliminary research that is usually seen as crucial for construct validity. To conclude, steps to enhance validity and reliability might involve the development of 1) a test matrix; 2) the development of the assignments and the test environment; and 3) the preliminary tests to validate the developed instruments. The systematic approach guarantees transparency of the development process and replicability for any future assessments.

In this study, we used the four-phase flow diagram from the PRISMA statement to minimize bias in the design and conduct of the review. However, with a small sample size, caution must be applied, as the findings might not be transferable to all kinds of assessment instruments. We have tried to capture all studies within the specified selection criteria, but we are not able to claim with complete certainty that all studies are included, for example, because the studies are not published in English or in journals that are not indexed in our selected databases.

Therefore, this review also has its limitations. The main weakness of this study was the paucity of the selected studies, because of the strictly selection criteria. Notwithstanding the relatively limited sample, this work offers valuable insights into the assessment of primary school students' digital competences. Starting from this limitation, we also recommend further research into the assessment of students' digital competences to expand our knowledge in this research field. An expansion of assessment instruments with attention for every component of digital competence according to the DigComp 2.1. framework, would be an important step forward in providing citizens and young people with a more reliable way of evaluating their readiness to cope with the rapidly changing technology in today's knowledge society. Various researchers including Aesaert & van

Braak (2015), and Siddiq et al. (2016) recommend not to measure the entire domain of digital competence with an assessment instrument, but to choose a specific, clearly outlined, and defined area of this competence. Therefore, providing conceptual clarity is an important first step and starts with translating the selected (curriculum) objectives, i.e. the digital competences, into measurable subcompetences. This is done by means of a content analysis of the curriculum objectives. The content of the curriculum objectives should be classified in a test matrix to which the subcompetences are linked. The test matrix then forms the blueprint for developing the assignments. Finally, further research could also focus on the development of assessment tools of the underexposed sub-competences such as e.g. digital citizenship or computational thinking. Both are also relevant competence areas for education, but the results of this review study suggest that the knowledge that can support the design of assessment instruments for these areas is insufficiently developed at an international level.

For follow-up research a few actions could be considered to guarantee the quality of the assessment instrument. A set of design steps introduced by Gyll & Ragland (2019) can be used as a guideline. In the process of test development, the researcher should choose between paper-and-pencil assessment, or computer-assisted assessment. Most studies in this systematic review are based on a computer-assisted environment, both for taking authentic test assignments and for traditional item formats so answers are automatically registered and scored. Computer-assisted assessments generally use the computer to provide richer stimulating material (e.g. video) and/or collect data on a large scale cost-effectively (Fraillon, 2018; Siddiq et al., 2016). In addition, the use of digital technology enables the use of staged/contextualized commands in combination with traditional item types. In this review, most of the studies used self-reporting assessment, whether in combination with traditional or objective and/or performance-based assessment. Performance-based assessment is only cost-effective in computer-assisted assessment because answers are automatically logged. The demand for authenticity, however, is challenged with the demand for standardization (Fraillon, 2018). Although the digital assessment environment should be a realistic representation of the digital tools that students currently use and problems they are faced with, it should be avoided that assessment results are confounded by uncontrollable processes, such as the accessibility of online information outside the assessment environment or dependence on the speed of the internet connection (Aesaert & van Braak, 2015; Heitink, 2018). Therefore, the development of a strictly closed environment at the back-end with an authentic frontend for the users is recommended, offering functional and realistic tasks. This imposes certain demands on the design of the closed environment, such as the inclusion of (platform-independent) software (e.g. office applications, search engine, e-mail, chat) and the integration of a large number of websites that students can consult via search queries.

To conclude, this systematic review is relevant as it provides insight into frameworks that give a starting point for conceptually and methodologically assessing primary school students' digital competences in a valid and reliable way. Continued efforts are needed to develop assessment instruments and assessment environments based on well-defined conceptual frameworks and systematic approaches of design steps.

Table 3: Overview of the criteria of assessment instruments based on the RCEC-framework

Criteria \ Authors	Aesaert et al., 2014	Asil et al., 2014	Jun et al., 2014	Kim et al., 2014	Lee et al., 2015	Zhang & Zhu, 2016	Fernandez-Montalvo et al., 2017	ACARA, 2018	Heitink, 2018	Kiili et al., 2018	Kim et al., 2019	Kong et al., 2019	Chou & Chiu, 2020	Lazonder et al., 2020
1) Purpose and use Is it clear what we are measure and why we do this?														
Is the target population specified?	Age: 11-12 Grade: 6	Age: 10-13 Grade: 6	Age: 6-12 Grade: 1-6	Age: 9-11 Grade: 4-6	Age: 10-17 Grade: 4-11	Age: 9-13 Grade: 5-6	Age: 11 Grade: 6 + extra info (participants must first have completed program)	Age: 11-12 Grade: 6	Age: 10-13 Grade: 5-8	Age: 12-13 Grade: 6	Age: 10-15 Grade: 4-9	Age: NM Grade: 4-6	Age: 11-12 Grade: 6	Age: 8-11 Grade: 5-6
Is the measurement purpose specified?	Mastery of a competence	To validate an instrument Measurement of attitudes Influence of affecting factors	Mastery of a competence	Mastery of a competence Influence of affecting factors	Mastery of a competence	Mastery of a competence Influence of affecting factors	Effectiveness of a program	Test outcomes achievement national objectives	Mastery of a competence	Mastery of a competence	Mastery of a competence Create a new instrument	Mastery of a competence	Mastery of a competence	The development of a competence
Overview of measured constructs	Test matrix/overview or description of items	Test matrix/overview or description of items	NM	NM	IR	Test matrix/overview or description of items	Test matrix/overview or description of items	Test matrix/overview or description of items	Test matrix/overview or description of items	Test matrix/overview or description of items	NM	Test matrix/overview or description of items	Test matrix/overview or description of items	Test matrix/overview or description of items
Is the measurement use specified?	Formative	Formative	Formative	Formative	Formative	Formative	Summative	Formative	Formative	Formative	Formative	Formative	Formative	Formative
2) Quality of the test and examination material Is the instruction and data-collection standardized?														
Are the questions standardized?	Closed online environment Fixed sequence of items Time limit to have information for each part of the construct	Fixed sequence of questions	IR	IR	IR	IR	NM	Different modules (no fixed sequence)	Assigned sequence of items	IR	IR	Two formats: online or paper-pencil test (on request of the school)	IR	Three versions
Is an objective scoring system being used? If the scoring is done by assessors, the scoring requirements need to be complete and clear	Dichotomous (1= correct; 0 = incorrect) Manually scored if it needed intelligible judgement (team of test raters) Test raters: experts, scoring guide and training	5-point Likert Scale 1= "Strongly disagree" to 5 = "strongly agree" Automatically	Dichotomous (1= correct; 0 = incorrect) Automatically	Dichotomous (1= correct; 0 = incorrect)	5-Point Likert Scale IR	5-Point Likert Scale IR	NM	0 (incorrect/insufficient) 1 (correct/sufficient) 9 (no attempt made) ; 1(partially attempt made); 2 (fully attempt made ; Scored by criteria Automatically scored by the system or scored later by trained scorers and prescribed scoring guide	Correct (1), partly correct (0.5) or incorrect (0) → a total score per domain by accumulating the scores 5-point Likert scale Four tasks were scored by hand by trained scorers and prescribed scoring guide	Scoring rubric (0-2 points; 0-5 points)	Scoring → cut-off scores for achievement levels IR	5-point-Likert scale IR	7-point-Likert scale IR	Correct/successful = 1 Incorrect/unsuccessful = 0 (rubric)
Is incorrect use of the test prevented? Are questions, answer scales and/or the answer form designed in such a way as to prevent errors when filling in the form? *Are the items correctly formulated?*	Pilot study: Experts and students During the entire development: items administered to pupils	Pilot study: think aloud students (about readability)	NM	NM	Pilot: think-aloud (readability, and if it sounds familiar) e.g. LOL and OMG (not using acronyms) and ;-) → emoticons	Pilot test students + reviewers (experts)	NM	Pilot/field trial: testing the methodologies, systems, documents, and items	Small pilot testing the online environment	Pilot Study students: some modifications in words were made (instructions and prompts) to reflect the language and culture	IR	Pilot study (experts) Researchers and teachers	Different steps to avoid incorrect use: literature review, expert interviews, focus groups, pilot testing	Pilot study students (comments and performance helped improve the assessment instrument)
Are the instructions for the participant complete and clear?	A picture of the general interface was provided (description of the task)	NM	NM	NM	IR	NM	NM	Test-taking info and navigation facilities + content info	Guided with instructions	IR	IR	NM	NM	NM
3) Representativeness To which degree the instrument measures the construct it purports to measure?														

Is there a blueprint, test program, competency profile, or the operationalization of the construct?	Yes	Yes	No, only a few examples	NM (also no examples)	Operational definitions and measurement scales	Yes	Yes	Yes	Yes	Yes	NM (also no examples)	Yes	Yes	IR
Is the blueprint, test program, competency profile, or the operationalization of the construct an adequate representation of the measurement purpose?	Domain analysis to clearly define and operationalize the concept to be measured Test matrix 1 or more items/ICT competence or skill Review of the test matrix by experts	Based on existing models and theories to become a comprehensive measure of attitudes towards computers Study contains of three dimensions. The constructs and items are provided	Aims to measure the construct 'Computational literacy' and contains of three components. It is based on the national program of Korea No test matrix is provided (only a few examples of items)	Based on national program of Korea The ICT literacy test was developed by considering the ability and content elements based on previous studies No test matrix is provided (they refer to a previous study)	Operationalizing the NML indicators and choice of respective measuring scales. Once operationalized, questions were developed. Afterwards a pilot study was conducted to ensure the choice of words is pitched at the students' comfort level	Items cover the different dimensions of the construct Content is based on previous studies A measurement tool related to digital media literacy is developed and tested; than the digital media literacy is measured; finally, the factors associated with DML are examined	The items are provided (three components) IR	Based on national program and everything is explained in an extra technical report	Construct and domains defined and operationalized by comparing well-known frameworks Expert review Development of tasks in collaboration with experts	They refer to previous studies IR	The aim is to measure the construct ICT literacy. To become an accurate measurement and interpretation of learners' ICT literacy: Experts' advice Achievement standards with reference to the Korean informatics curriculum and ICILS standards. Second expert's review	Literature review Test matrix ok Review by experts	Previous literature is used to describe the construct and its dimensions Blueprint is available	A description of the tasks is provided Based on a literature review
Steps the author(s) made to make sure the instrument is valid	1) Domain analysis 2) Pilot study 3) Classical Test Theory 4) Item Response Theory IRT	1) Item generation 2) Pilot study 3) Statistical and psychometric analysis (factorial validity and criterion validity was measured) - MG-CFA 4) Revision 5) Administration and scoring	Brief explanation of the steps in test validation 1) Angoff method 2) PCA in 3 domains 3) Item discrimination and difficulties	IR	Readability interviews 1) item reduction (low variability, potential redundancy, potential violation of unidimensionality) 2) removal of the 'creation' scale 3) Redistribution of items (see step 1) Content validity: expert panel Construct validity: CFA	Quantitative (pilot study students) and qualitative (by experts) review Data in equal halves: one half EFA with Varimax rotation; another half CFA	NM	Quantitative (pilot study students) and qualitative (by experts) review Scaling procedure using IRT scaling methodology. DIF, Item calibration, horizontal equating, item-rest correlation (as an index of item discrimination), CFA and EFA	Quantitative (pilot study students) and qualitative (by experts) review (1) defining assessment criteria, (2) conceptualization and operationalization of online information literacy, (3) design of authentic assessment tasks and the assessment environment, (4) determining assessment output and scoring, (5) small pilot, (6) development of final prototype.	Large pilot in previous study. Only a translation was made so the word choice sounds more Finnish. Validity was established with a framework document approved by experts, 2 years of cognitive lab testing, and modifications based on a large-scale pilot study	Expert pilot study Goodness of fit Choosing the most appropriate items after using flex-MIRT (version 3.5) to analyze the discrimination and difficulty of items, CFA	Expert pilot study EFA (initial understanding of the factor structure of the scale) Maximum likelihood estimation with promax rotation and Kaiser normalization was used to measure the goodness of fit of the model. CFA CFI, TLI, RMSEA were used as the fit indices for the measurement model of digital empowerment	Quantitative (pilot study students) and qualitative (by experts) review EFA CFA literature review, conceptualized constructs, developed questions, and interviewed experts to develop items of the measurement. item analysis, factor analysis, and inter-item correlations	Pilot test IR
Is the difficulty of the items adjusted to the target group?	Quantitative (pilot study students) and qualitative (by experts) review CTT (item difficulty) was used	Pilot testing with statistical and psychometric analysis which provides evidence for factorial validity and criterion validity	Test was verified and analyzed in 2007 (KERIS, 2007) IR	IR	Only readability interviews	Quantitative (pilot study students) and qualitative (by experts) review	NM	Quantitative (pilot study students) and qualitative (by experts) review Item performance data: fit statistics, scaled difficulties, differential item functioning, scale reliability	Quantitative (pilot study students) and qualitative (by experts) review The difficulty, usability, their appreciation and the technical soundness of the assessment were tested	Large pilot in previous study Only a translation was made so the word choice sounds more Finnish	Expert review: Most appropriate items	Reviewed by experts	Quantitative (pilot study students) and qualitative (by experts) review	IR
4) Reliability To which degree is the measurement instrument free from measurement error?														
Is information on the reliability of the test provided?	TIF (test information function) and IIF (item information	Reliability estimates of each construct were computed.	Cronbach's alpha was calculated and found to be reliable	NM	The reliability of the instrument is established by internal	The internal consistency is calculated	The internal consistency is calculated	Cronbach's alpha (scale reliabilities & their respective correlation with	Yes, Cohen's kappa	The Kuder–Richardson Formula 20 (KR-	NM	The internal consistency is calculated	The reliability of each item of the subscales is represented by	The internal consistency is calculated

	function) as reliability-like statistics and reliability-index (r) is calculated				consistency (Cronbach's alpha).	(Cronbach's alpha more than .70)	(Cronbach's alpha more than .70) And correlation between items is also investigated	ICT scores) are provided		20) estimates of reliability for this assessment		(Cronbach's alpha more than .70)	squared standardized factor loading and the composite reliability (CR) refers to the overall reliability of the whole scale.	(Cronbach's alpha more than .70)
Is the reliability of the test correctly calculated?	Procedure is explained Sample > 200	Cronbach's alpha is calculated. Procedure not explained. Sample > 200	Procedure not explained. Sample > 200	NM Sample > 200	Cronbach's alpha is calculated. Procedure not explained. Sample > 200	Cronbach's alpha is calculated. Procedure not explained. Sample > 200	IR Sample > 200	Procedure is explained Sample > 200	Procedure is explained Sample > 200	The authors refer to other sources for information about the reliability Sample > 200	NM Sample > 200	Cronbach's alpha is calculated. Procedure not explained. Sample > 200	The reliability is given. Procedure not explained. Sample > 200	Cronbach's alpha is calculated. Procedure not explained. Sample < 200
Is the reliability sufficient, considering the decisions that have to be based on the test?	**TIF and IFF (>.70)** r= .86 (good internal consistency)	**Composite reliability (>.70)** Perceived ease of use: .59 Affect toward computer: .74 Perceived usefulness: .72	**Cronbach's alpha (>.70)**: reliable IR	NM	**Cronbach's alpha** for the ten scales, the four constructs and the full survey. The alphas ranged from .72 to .96 showing the evidence of reliability. 4.3.	**Cronbach's alpha** >.70 for all scales (IR)	**Chronbach's alpha** .73 (IR)	**Cronbach's alpha** .91 for year 6 .86 for year 10	**Cohen's kappa** = .80 - .84 (substantial to good)	**The Kuder–Richardson Formula 20** (KR-20) estimates of reliability for this assessment ranged from .70 to .82	NM	**The Cronbach's alpha** coefficients for the four factors were all above .70, indicating that the subscales had good internal consistency	As such, the values of **composite reliability** for the four dimensions were acceptable (.85; .88; .93; .93)	**Cronbach's alpha**: .73
5) Standard setting, norms, and equating														
a. In the case of absolute norms														
Is a (performance) standard provided?	There are no norms/standards provided	There are no norms/standards provided	ANGOFF-method	BOOKMARK-method	There are no norms/standards provided	There are no norms/standards provided	There are no absolute norms provided	Yes, according to a minimum acceptable level of mastery. (the national standards)	There are no norms/standards provided	There are no norms/standards provided	BOOKMARK-method	There are no norms/standards provided	There are no norms/standards provided	There are no norms/standards provided
Is the standard-setting procedure correctly performed?	-	-	Yes	Yes	-	-	-	The standards represent a challenging a reasonable expectation of what typical YEAR 6 students should achieve by the end of the respective years.	-	-	Yes	-	-	-
Are the standard-setting specialists properly selected and trained?	-	-	The experts consisted of two professors and eight elementary school teacher who had majored in computer education	The panels who participated in standard setting were composed of twelve persons: ten elementary school teachers and two content specialists.	-	-	-	The standards-setting groups included currently practicing teachers with specific ICT expertise, ICT curriculum experts and educational assessment experts.	-	-	15 experts with expertise in ICT literacy or experience with teaching elementary and middle school students	-	-	-
Is there sufficient agreement among the specialists?	-	-	Following the rules of the bookmark-method	The cut-off score was determined by the panel after three rounds of round-the-table discussion.	-	-	-	Yes (procedure is explained in technical report)	-	-	Following the rules of the bookmark-method	-	-	-
b. In the case of relative norms														
Is the quality of the norms sufficient? Is the norm group large enough? Is the norm group representative?	-	-	-	-	-	-	Based on an experimental design. The experimental group reached a significantly higher degree of digital literacy	Differences between groups: Between states and territories; student sub-groups; this assessment cycle and previous ones	-	-	-	-	-	-
Are the meaning and the limitations of the	-	-	-	-	-	-	NM	Yes	-	-	-	-	-	-

norm scale made clear to the user and is the norm scale in accordance with the purpose of the test?														
Is the mean and standard deviation of the score distribution provided?	-	-	-	-	-	-	A descriptive analysis has been carried out to determine the characteristics of the sample	A descriptive analysis has been carried out to determine the characteristics of the sample	-	-	-	-	-	-
Is information on the accuracy of the test and the corresponding intervals (standard error of measurement, standard error of estimation, test information) provided?	-	-	-	-	-	-	No	Yes	-	-	-	-	-	-
c. In the case of both relative and absolute norms														
Are standards or norms maintained?	-	-	-	-	-	-	-	Yes	-	-	-	-	-	-
Is the method for maintaining standards or norms correctly applied?	-	-	-	-	-	-	-	Yes	-	-	-	-	-	-
6) Administration and security														
Is sufficient information on the administration of the test available for the participant?	Each task starts with a pop-up window, containing the instruction for the task The instruction remains visible	NM	NM	NM	NM	NM	NM	Test administrators were responsible for running the student tutorial, supervising student participation and monitoring student progression between each section/module	Prior to the administration, participating schools, students, and parents of participating students were informed about the study, about data collection procedures and about the way data would be stored	Virtual online environment, guided by an avatar which gives instruction	IR	NM	Informed consents to inform the parents the researchers went to the participating schools to instruct the students to work on the self-reported survey	Following a short, standardized instruction, children completed the test
Is the information for the proctor complete and clear?	Yes	NM	NM	NM	Yes	NM	NM	Yes	Yes	Yes	IR	IR	NM	NM
Is information on the degree of expertise required to administer the test available?	Yes	NM	NM	NM	NM	Yes	NM	Yes	Yes	Yes	IR	IR	NM	NM
Is the test sufficiently secured?	Closed computer-based environment	NM	NM	NM	Survey hosted by Qualtrics	NM	NM	Closed computer-based environment	Closed computer-based environment	Closed computer-based environment	IR	IR	NM	NM
If computer-based assessment														
Is information on the installation of the computer software provided?	In another article (Aesaert & van Braak, 2015) the students received a personal code connected to his/her name to log in to the test environment	NM	NM	NM	Survey hosted by Qualtrics	NM	NM	Closed computer-based environment supported by a test administrator	Closed computer-based environment supported by a test administrator	Closed computer-based environment supported by a test administrator	IR	IR	NM	NM
Is information on the operation and the possibilities of the software provided?	Six general software applications were designed for this test: a web browser, e-mail software, presentation software, a word	NM	NM	NM	NM	NM	NM	The student screen had three main sections: a surrounding border of test-taking information and navigation facilities; a central information	Closed online environment with chat function; website tool	IR	IR	IR	NM	NM

	processor, a file management system and spreadsheet software							section that could house stimulus materials for students to read or (simulated or live) software applications; and a lower section containing the instructional and interrogative text of the assessment items and the response areas for multiple-choice and constructed response items.							
Are there sufficient possibilities for technical support?	In each school a test assistant is present to solve technical problems	Under the supervision of their class teachers.	NM	NM	Each participating school has appointed one teacher to be the survey administrator.	NM	/	An IT coordinator was responsible for ensuring that the school's computer system was 'test ready' by the scheduled assessment date.	Test leaders were trained to make sure data collection and scoring would be similar for every student.	the researchers provided technical assistance with the test application when needed	IR	IR	NM	Under supervision of the test publisher & teacher

Note.
NM = not mentioned
IR = information restricted
- = not applicable

6. References

ACARA. (2018). *National assessment program - ICT literacy years 6 & 10 report 2018: Australian Curriculum, Assessment and Reporting Authority*. Retrieved from https://www.nap.edu.au/docs/default-source/default-document-library/2017napictlreport_final.pdf?sfvrsn=2

Ainley J. (2018) Students and Their Computer Literacy: Evidence and Curriculum Implications. In J. Voogt, G. Knezek, R. Christensen, & K.-W. Lai (Eds.). *Second Handbook of Information Technology in Primary and Secondary Education* (pp. 69-88). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-53803-7

Ainley, J., Fraillon, J., Schulz, W., & Gebhardt, E. (2016). Conceptualizing and Measuring Computer and Information Literacy in Cross-National Contexts. *Applied Measurement in Education*, *29*(4), 291–309. https://doi.org/10.1080/08957347.2016.1209205

Ala-Mutka, K. (2011). *Mapping digital competence: Towards a conceptual understanding.* Retrieved from ftp://jrc.es/pub/EURdoc/JRC67075_TN.pdf

Alderson, J. C., & Wall, D. (1993). Does washback exist? *Applied Linguistics*, *14,* 115–129. Retrieved from https://www.semanticscholar.org/paper/Does-Washback-Exist-Alderson-Wall/6539569d3df3b740e56ee593d322bd7e53f506ad

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

Aesaert, K., Vanderlinde, R., Tondeur, J., & van Braak, J. (2013). The content of educational technology curricula: A cross-curricular state of the art. *Educational Technology Research and Development*, *61*(1), 131–151. https://doi.org/10.1007/s11423-012-9279-9

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 and Education*, 76, 168–181. https://doi.org/10.1016/j.compedu.2014.03.013

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

Aesaert, K. & van Braak, J. (2018). Information and Communication Competences for Students. In J. Voogt, G. Knezek, R. Christensen, & K-W. Lai (Eds.) *Second Handbook of Information Technology in Primary and Secondary Education*. (pp. 255-269). Amsterdam: Springer International Publishing AG. https://doi.org/10.1007/978-3-319-53803-7

Baartman, L. K. J., Bastiaens, T. J., Kirshner, P. A., & van der Vleuten, C. P. M. (2007). Evaluating assessment quality in competence-based education: A qualitative comparison of two frameworks, *Educational Research Review*, *2*(2), 114–129. https://doi.org/10.1016/j.edurev.2007.06.001

Ballantine, J. A., McCourt Larres, P., & Oyelere, P. (2007). Computer usage and the validity of self-assessed computer competence among first-year business students. *Computers and Education, 49*(4), 976–990. https://doi.org/10.1016/j.compedu.2005.12.001

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

Blömeke, S., Gustafsson, J. E., & Shavelson, R. J. (2015). Beyond dichotomies: Competence viewed as a continuum. *Zeitschrift Fur Psychologie / Journal of Psychology*, *223*(1), 3–13. https://doi.org/10.1027/2151-2604/a000194

Calvani, A., Fini, A., Ranieri, M., & Picci, P. (2012). Are young generations in secondary school digitally competent? A study on Italian teenagers. *Computers and Education*, *58*(2), 797–807, https://doi.org/10.1016/j.compedu.2011.10.004

Carretero, S.; Vuorikari, R., & Punie, Y. (2017). *DigComp 2.1: The Digital Competence Framework for Citizens with eight proficiency levels and examples of use*. Retrieved from https://op.europa.eu/s/oNl8

Casey, L., & Bruce, B. C. (2011). The practice profile of inquiry: Connecting digital literacy and pedagogy. *E-Learning and Digital Media, 8*(1), 76–85. https://doi.org/10.2304/elea.2011.8.1.76

Chaudron S., Di Gioia R., & Gemo M. (2018). *Young children (0-8) and digital technology, a qualitative study across Europe*; https://op.europa.eu/s/oNl9

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

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

Duncan, G. J., Engel, M., Claessens, A., & Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology, 50*(11), 2417-2425. http://dx.doi.org/10.1037/a0037996

Evers, A., Lucassen, W., Meijer, R., & Sijtsma, S. (2010). COTAN Beoordelingssysteem voor de kwaliteit van tests. Amsterdam: NIP. [www.psynip.nl: https://www.psynip.nl/wp-content/uploads/2016/07/COTAN-Beoordelingssysteem- 2010.pdf].

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

Fraillon, J. (2018). International Large-Scale Computer-Based Studies on Information Technology Literacy in Education. In In J. Voogt, G. Knezek, R. Christensen, & K-W. Lai (Eds.) *Second Handbook of Information Technology in Primary and Secondary Education*. (pp. 1161–1179). Springer International Publishing AG. https://doi.org/10.1007/978-3-319-71054-9_80

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Gebhardt, E. (2014). Preparing for life in a digital age. The IEA international computer and information literacy study international report. Springer International

Publishing AG. Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?article=1009&context=ict_literacy

Fraillon, J., Ainley, J., Schulz, W., Friedman, T., & Duckworth, D. (2018). Preparing for life in a digital age. The IEA international computer and information literacy study international report. Springer International Publishing AG. Retrieved from https://www.iea.nl/sites/default/files/2020-04/IEA%20International%20Computer%20and%20Information%20Literacy%20Study%202018%20International%20Report.pdf

Fraillon, J., Gebhardt, E., Friedman, T., Duckworth, D., O'malley, K., & Vernon, K. (2018). *NAP Sample Assessment ICT Literacy.* Retrieved from http://www.nap.edu.au/docs/default-source/default-document-library/2017napictlreport_final.pdf?sfvrsn=2

Fraillon, J., Schulz, W., & Ainley, J. (2013*). International Computer and Information Literacy Study assessment framework.* Retrieved from https://research.acer.edu.au/cgi/viewcontent.cgi?article=1010&context=ict_literacy

Gyll, S. & Ragland, S. (2018). Improving the validity of objective assessment in higher education: Steps for building a best-in-class competency-based assessment program. *The Journal of Competency-Based Education, 3*(1). https://doi.org/10.1002/cbe2.1058

Hatlevik, O. E., & Christophersen, K. (2013). Digital competence at the beginning of upper secondary school: Identifying factors explaining digital inclusion. *Computers and Education*, *63*, 240-247. http://dx.doi.org/10.1016/j.compedu.2012.11.015

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

Hemker, Sluijter, & Sanders (2019). Assessing Computer-Based Assessments. In B. P. Veldkamp and C. Sluijter (eds.), *Theoretical and Practical Advances in Computer-based Educational Measurement*, Methodology of Educational Measurement and Assessment (pp. 73-89). https://doi.org/10.1007/978-3-030-18480-3_4

Huggins, A. C., Ritzhaupt, A. D., & Dawson, K. (2014). Measuring Information and Communication Technology Literacy using a performance assessment: Validation of the Student Tool for Technology Literacy (ST 2 L). *Computers and Education*, *77*, 1–12. https://doi.org/10.1016/j.compedu.2014.04.005

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

Juhaňák, L., Zounek, J., Záleská, K., Bárta, O., Vlčková, K. (2019). The relationship between the age at first computer use and students' perceived competence and autonomy in ICT usage: A mediation analysis. *Computers & Education, 141*, 1–14. https://doi.org/10.1016/j.compedu.2019.103614

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

Kaarakainen, M. T., Kivinen, O., & Vainio, T. (2018). Performance-based testing for ICT skills assessing: a case study of students and teachers' ICT skills in Finnish schools. *Universal Access in the Information Society, 17*(2), 349–360. https://doi.org/10.1007/s10209-017-0553-9

Kahn, S. & Zeidler, D. L. (2016). A case for the use of conceptual analysis in science education research. *Journal Of Research in Science Teaching*, 54(4), 538-551. https://doi.org/10.1002/tea.21376

Kamp, van de, M.T.A. (2012). Herziene taxonomie van de leerdoelen van Bloom volgens Anderson en Krathwohl, 2001. Retrieved from https://snro-instituut.nl/wp-content/uploads/2017/09/bloom.pdf

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. (2019). 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

Kirshner, P. & De Bruyckere, P. (2017). The myths of the digital native and the multitasker. *Teaching and Teacher Education, 67*, 135-142. https://doi.org/10.1016/j.tate.2017.06.001

Klieme, E., Hartig, J., & Rauch, D. The concept of competence in educational contexts. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 3–22). Hogrefe & Huber Publishers.

Knezek, G. A., & Christensen, R. (2008). The importance of information technology attitudes and competencies in primary and secondary education. In J. Voogt & G. Knezek (Eds.), *International handbook of information technology in primary and secondary education* (pp. 321–331). Springer International Publishing AG. https://doi.org/10.1007/978-0-387-73315-9_19

Koeppen, K., Hartig, J., Klieme, E., & Leutner, D. (2008). Current Issues in Competence Modeling and Assessment. *Zeitschrift Für Psychologie / Journal of Psychology*, *216*(2), 61–73. https://doi.org/10.1027/0044-3409.216.2.61

Kong, S. C., Wang, Y. Q., & Lai, M. (2019). Development and Validation of an Instrument for Measuring Digital Empowerment of Primary School Students. *CompEd 2019 - Proceedings of the ACM Conference on Global Computing Education*, 172–177. https://doi.org/10.1145/3300115.3309523

Law, N., Lee, Y., & Yuen, H. K. (2009). *The impact of ICT in education policies on teacher practices and student outcomes in Hong Kong*. In F. Scheuermann & F. Pedro (Eds.), Assessing the effects of ICT in education–Indicators, criteria, and benchmarks for international comparisons (pp. 143–164). OECD. Retrieved from https://www.academia.edu/2579082/The_impact_of_ICT_in_education_policies_on_teacher_practices_and_student_outcomes_in_Hong_Kong

Lazonder, A. W., Walraven, A., Gijlers, H., & Janssen, N. (2020). Longitudinal assessment of digital literacy in children: Findings from a large Dutch single-school study. *Computers and Education*, *143*, 103681. https://doi.org/10.1016/j.compedu.2019.103681

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

Le Deist, F. D., & Winterton, J. (2005). What is competence? *Human Resource Development International*, 8(1), 27–46.

Leutner, D., Hartig, J., & Jude, N. (2008). Measuring competencies: Introduction to concepts and questions of assessment in education. In J. Hartig, E. Klieme, & D. Leutner (Eds.), *Assessment of competencies in educational contexts* (pp. 177–192). Hogrefe & Huber Publishers.

Li, Y., & Ranieri, M. (2010). Are 'digital natives' really digitally competent? A study on Chinese teenagers. *British Journal of Educational Technology, 41*(6), 1029–1042. https://doi.org/10.1111/j.1467-8535.2009.01053.x

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

Markauskaite, L. (2007). Exploring the structure of trainee teachers' ICT literacy: the main components and relationships between general cognitive and technical capabilities. *Educational Technology Research & Development, 55*(6), 547-572. https://doi.org/10.1007/s11423-007-9043-8

Martin, A., & Madigan, D. (2006). *Digital literacies for learning*. Facet Publishing

Merritt, K., Smith, K.D., & Di Renzo, J.C. (2005). An Investigation of Self-Reported Computer Literacy: Is It Reliable? *Issues in Information Systems, 6*(1), 289 – 295. https://doi.org/10.48009/1_iis_2005_289-295

Moher, D. A., Liberati, J., Tetzlaff, & Altman, D. G. (2010). Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *International Journal of Surgery, 89*, 336–341. https://doi.org/10.1371/journal.pmed.1000097

Morgan, N. (2020). Children as digital citizens: Closing the gap on parental engagement. Information *Technology Education and Society*, *17*(1), 41-54. https://doi.org/10.7459/ites/17.1.04

NCES. (2014). *The TEL Assessment Framework*. Retrieved from https://www.nagb.gov/content/nagb/assets/documents/publications/frameworks/technology/2018-technology-framework.pdf

North Central Regional Educational Laboratory (NCREL). (2003). *enGauge 21st century skills: Literacy in the digital age.* Naperville, IL: North Central Regional Educational Laboratory. Retrieved from http://files.eric.ed.gov/fulltext/ED463753.pdf

Ornstein, A.C., & Hunkins, F.P. (2017). *Curriculum: Foundations, Principles, and Issues.* (7th ed.). Pearson Education Limited.

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

Pöntinen, S., & Räty-Záborszky, S. (2020). Pedagogical aspects to support students' evolving digital competence at school. *European Early Childhood Education Research Journal*, *28*(2), 182–196. https://doi.org/10.1080/1350293X.2020.1735736

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

Sanders, P.F., Dijk, P. van, Eggen, T., Otter, D. den, & Veldkamp, B. (2016). *RCEC Beoordelingssysteem voor de kwaliteit van studietoetsen en examens.* Enschede: RCEC. Retrieved from www.rcec.nl/Beoordelingssysteem

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

Spisak, J.R. (2018). *Secondary student information literacy efficacy vs. performance.* (Doctoral dissertation, VA: Virginia Commonwealth University, Richmond, Virginia). Retrieved from: https://doi.org/10.25772/4AP6-F755

Straetmans, G., & Sanders, P. (2001). *Beoordelen van competenties van docenten.* Retrieved from http://www.rcec.nl/publicaties/overige publicaties/EPSbrochure5_Beoordelen.pdf

van Deursen, A., & van Dijk, J. (2011). Internet skills and the digital divide. *New Media Society, 13*(6), 893 – 911. https://doi.org/10.1177/1461444810386774

van Deursen, A., & van Dijk, J. (2016). Modeling traditional literacy, internet skills and internet usage: An empirical study. *Interacting with Computers, 28*, 13–26. https://doi.org/10.1093/iwc/iwu027

Weinert, F. E. (2001). Concept of competence: A conceptual clarification. In D. S. Rychen & L. H. Salganik (Eds.), *Defining and selecting key competencies* (pp. 45–65). Hogrefe & Huber Publishers.

Wild, S., & Schulze Heuling, L. (2021). Re-evaluation of the D21-digital-index assessment instrument for measuring higher-level digital competences. *Studies in Educational Evaluation*, *68*(February), 100981. https://doi.org/10.1016/j.stueduc.2021.100981

Wilson, M., & Sloane, K. (2000). From principles to practice: an embedded assessment system. Applied *Measurement in Education, 13*(2), 181–208. https://doi.org/10.1207/S15324818AME1302_4

Zhang, H., & Zhu, C. (2016). A Study of Digital Media Literacy of the 5th and 6th Grade Primary Students in Beijing. *Asia-Pacific Education Researcher, 25*(4), 579–592. https://doi.org/10.1007/s40299-016-0285-2

Appendices

Appendix A

Search engine

TS=("Digital competenc" OR "ICT competenc*" OR "digital literacy" OR "ICT literacy" OR "Digital capacities" OR "ICT capacities" OR "digital knowledge" OR "digital skills" OR "digital attitudes" OR "ICT knowledge" OR "ICT skills" OR "ICT attitudes" OR "ICT fluency" OR "information and computer literacy" OR "technological competen*")*

#2 (TS=("Assessment" OR "Measurement" OR "Instrument development" OR "Test development" OR "Assessing" OR "Measuring" OR "testing")

#3 TS=("Primary education" OR "Elementary education" OR "Primary school" OR "elementary school" OR "K-6" OR "students" OR "pupils" OR "1st-grade" OR "first-grade*" OR "grade 1" OR "grade one" OR "2nd-grade*" OR "second-grade*" OR "grade 2" OR "grade two" OR "3rd-grade*" OR "third-grade*" OR "grade 3" OR "grade three" OR "4th-grade*" OR "fourth-grade*" OR "grade 4" OR "grade four" OR "5th-grade*" OR "fifth-grade*" OR*
"grade 5" OR "grade five" OR "6th-grade" OR "sixth-grade*" OR "grade 6" OR "grade six")*

#3 AND #2 AND #1

WEBOFSCIENCE = 161

TITLE-ABS-KEY (= "Digital competenc" OR "ICT competenc*" OR "digital literacy" OR "ICT literacy" OR "Digital capacities" OR "ICT capacities" OR "digital knowledge" OR "digital skills" OR "digital attitudes" OR "ICT knowledge" OR "ICT skills" OR "ICT attitudes" OR "ICT fluency" OR "information and computer literacy" OR "technological competenc*" AND "Assessment" OR "Measurement" OR "Instrument development" OR "Test development" OR "Assessing" OR "Measuring" OR "testing" AND "Primary education" OR "Elementary education" OR "Primary school" OR "elementary school" OR "K-6" OR "students" OR "pupils" OR "1st-grade*" OR "first-grade*" OR "grade 1" OR "grade one" OR "2nd-grade*" OR "second-grade*" OR "grade 2" OR "grade two" OR "3rd-grade*" OR "third-grade*" OR "grade 3" OR "grade three" OR "4th-grade*" OR "fourth-grade*" OR "grade 4" OR "grade four" OR "5th-grade*" OR "fifth-grade*" OR "grade 5" OR "grade five" OR "6th-grade*" OR "sixth-grade*" OR "grade 6" OR "grade six") AND (LIMIT-TO (PUBYEAR , 2020) OR LIMIT-TO (PUBYEAR , 2019) OR LIMIT-TO (PUBYEAR , 2018) OR LIMIT-TO (PUBYEAR , 2017) OR LIMIT-TO (PUBYEAR , 2016) OR LIMIT-TO (PUBYEAR , 2015) OR LIMIT-TO (PUBYEAR , 2014) OR LIMIT-TO (PUBYEAR , 2013) OR LIMIT-TO (PUBYEAR , 2012)) AND (LIMIT-TO (DOCTYPE , "ar")) AND (LIMIT-TO (LANGUAGE , "English"))*

Scopus = 222

("assessment" OR "measurement") AND ("digital competence" OR "ICT competence") AND ("elementary school students" OR "primary school students")

Google scholar = 619

Appendix B

Data extraction form
- *Name of the instrument* - *Authors* - *Publication date* - *Country* - *Year of data collection* - *Sample size* - *Grade* - *Age* - *Type test (1= self-report; 2 = assessment based on traditional types of questions; 3 = performance based assessment)* - *Definition of digital competence* - *Explicit content* - *Test matrix or test questions* - *Measured competences according to DigComp 2.1. Framework* o *1. Information and data literacy (1.1. Browsing, searching and filtering data, information and digital content; 1.2. Evaluating data, information and digital content; 1.3. Managing data, information and digital content)* o *2. Communication and collaboration (2.1. Interacting through digital technologies; 2.2. Sharing through digital technologies; 2.3. Engaging in citizenship through digital technologies; 2.4. Collaborating through digital technologies; 2.5. Netiquette; 2.6. Managing digital identity)* o *3. Digital content creation (3.1. Developing digital content; 3.2. Integrating and re-elaborating digital content; 3.3. Copyright and licenses; 3.4. Programming)* o *4. Safety (4.1. Protecting devices; 4.2. Protecting personal data and privacy; 4.3. Protecting health and well-being; 4.4. Protecting the environment)* o *5. Problem Solving (5.1. Solving technical problems; 5.2. Identifying needs and technological responses; 5.3. Creatively using digital technologies; 5.4. Identifying digital competence gaps)* - *Duration of the test* - *Paper-pencil or computer-based assessment* - *Scoring* - *Aspect of competence* - *Amount of items* - *Criteria of assessment instruments based on RCEC-framework (see also guiding questions in Appendix C)* o *Purpose and use* o *Quality of the test and examination material* o *Representativeness* o *Reliability* o *Standard setting, norms and equating* o *Administration and security*

Appendix C

Overview first two dimensions DigComp 2.1.

DIMENSION 1	DIMENSION 2
AREA OF COMPETENCE	COMPETENCES
Information and data literacy	**1.1** Browsing, searching and filtering data, information and digital content
	1.2 Evaluating data, information and digital content
	1.3 Managing data, information and digital content
Communication and collaboration	**2.1** Interacting through digital technologies
	2.2 Sharing through digital technologies
	2.3 Engaging in citizenship through digital technologies
	2.4 Collaborating through digital technologies
	2.5 Netiquette
	2.6 Managing digital identity
Digital content creation	**3.1** Developing digital content
	3.2 Integrating and re-elaborating digital content
	3.3 Copyright and licenses
	3.4 Programming
Safety	**4.1** Protecting devices
	4.2 Protecting personal data and privacy
	4.3 Protecting health and well-being
	4.4 Protecting the environment
Problem solving	**5.1** Solving technical problems
	5.2 Identifying needs and technological responses
	5.3 Creatively using digital technologies
	5.4 Identifying digital competence gaps

Appendix D

Guiding questions RCEC

Component	Guiding questions
Purpose and use	• *Is the target population specified?* • *Is the measurement purpose specified?* • *Is the measurement use specified?*
Quality of the test and examination material	The following questions for paper-and-pencil tests are provided: • *Are the questions standardized?* • *Is an objective scoring system being used? If the scoring is done by assessors, the scoring requirements need to be complete and clear* • *Is incorrect use of the test prevented? Are questions, answer scales and/or the answer form designed in such a way as to prevent errors when filling in the form?* • *Are the instructions for the participant complete and clear?* • *Are the items correctly formulated?* • *What is the quality of the design?* The same questions are provided for computer-based tests except for question 3 that will be replaced by: "Is the software designed to avoid errors due to improper use?".
Representativeness	• *Is the blueprint, test program, competency profile, or the operationalization of the construct an adequate representation of the measurement purpose?* • *Is the difficulty of the items adjusted to the target group?*
Reliability	• *Is information on the reliability of the test provided?* • *Is the reliability of the test correctly calculated?* • *Is the reliability sufficient, considering the decisions that have to be based on the test.*
Standard setting, norms, and equating	*Questions for tests with absolute norms:* • *Is a (performance) standard provided?* • *Is the standard-setting procedure correctly performed? Are the standard-setting specialists properly selected and trained?* • *Is there sufficient agreement among the specialists?* *Questions for tests with relative norms:* • *Is the quality of the norms sufficient? Is the norm group large enough? Is the norm group representative?* • *Are the meaning and the limitations of the norm scale made clear to the user and is the norm scale in accordance with the purpose of the test?* • *Is the mean and standard deviation of the score distribution provided?* • *Is information on the accuracy of the test and the corresponding intervals (standard error of measurement, standard error of estimation, test information) provided?* *Questions for both absolute and relative norms:* • *Are standards or norms maintained?* • *Is the method for maintaining standards or norms correctly applied?*

Administration and security	• *Is sufficient information on the administration of the test available for the participant?*
	    o *Is the information for the proctor complete and clear?*
	    o *Is information on the degree of expertise required to administer the test available?*
	• *Is the test sufficiently secured?*
	
	Extra questions for computer-based tests:
	
	• *Is information on the installation of the computer software provided?*
	• *Is information on the operation and the possibilities of the software provided?*
	• *Are there sufficient possibilities for technical support?*

Appendix E

Summary of the studies included in the systematic review

Name of the instrument	Author(s)	Publication date	Country	Data collection	Sample size	Grade	Age	Type test *	Measured competences according to digcomp 2.1. Framework**	Duration of the test	Paper-and-pencil or computer-based assessment	Scoring	Aspect of competence	# items
Performance based test ICT competences	Aesaert, Van Nijlen, Vanderlinde, van Braak	2014 2015	Belgium (Flanders)	2012	560	6	11 – 12	1 en 3	1.1., 1.2., 1.3. 2.1., 2.2., 3.1., 3.2.	100 min	CBA (test environment)	Dichotomous (1= correct; 0 = incorrect)	Skills	27
Computer Attitude Measure for Young Students (CAMYS) scale	Asil, Teo, Noyes	2014	Singapore	2014	503	6	10-13	1	6.2	10 min	CBA (online questionnaire)	5-point Likert Scale 1= "Strongly disagree" to 5 = "strongly agree"	Attitudes	12
KERIS2014 (Korea Education Research and Information Service, K-12test)	Jun, Han, Kim, Lee	2014	Korea	2012	40072	1 - 6	6-12	2	1.1, 2.1., 3.1.,3.4., 5.1., 6.1.	Not mentioned	CBA (online questionnaire)	Dichotomous (1= correct; 0 = incorrect)	Skills and knowledge	46
ICT literacy test	Kim, Kil, Shin	2014	Korea	2012	11767	4-6	9-11	insufficiently reported	1.1., 1.2., 1.3., 2.1., 3.1., 5.2.	Not mentioned	CBA (online questionnaire)	Dichotomous (1= correct; 0 = incorrect)	Skills	36
New media literacy instrument	Lee, Chen, Li, Lin	2015	Singapore	2014	574	4 - 11	10-17	1	1.2, 1.3, 2.1, 2.2, 2.4, 3.1, 3.2, 4.1, 6.1	40 min	CBA(online questionnaire)	5-Point Likert Scale	Skills	86
Digital Media Literacy	Zhang & Zhu	2016	China	Not mentioned	796	5-6	9-13	1	1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 2.4, 2.5, 2.6, 3.1, 3.2, 4.1, 4.2	Not mentioned	Paper-and-pencil based assessment	5-Point Likert Scale	Skills Knowledge	42
Digital literacy programme	Fernández-Montalvo, Peñalva, Irazabal, López-Goñi	2017	Spain	2011	309	6	11	1 en 2	2.6., 4.2., 4.3.	Not mentioned	Not mentioned	Not mentioned	Skills Knowledge Attitudes	30 (10 MC; 20 Yes/No)

NAP- ICT Literacy 2017 Assessment Instrument	ACARA	2018	Australia	2017	5,439	6	11-12	1, 2 en 3	1.1., 1.2., 1.3., 2.1., 2.2., 2.4., 3.1., 3.2., 3.3., 4.1., 4.2., 4.3., 4.4.	120 min	CBA (online delivery system)	Different scoring methods 0 (incorrect/insufficient) 1 (correct/sufficient) 9 (no attempt made) ; 1(partially attempt made); 2 (fully attempt made ; Scored by criteria	Skills Knowledge	111
Authentic assessment of online information literacy	Heitink	2018	The Netherlands	2017	1036 (407 primary school 629 secondary school)	5-8	10-13	1 en 3	1.1., 1.2., 1.3. 3.1. 3.2, 5.3	55 min	CBA (digital assessment environment)	correct (1), partly correct (0.5) or incorrect (0) → a total score per domain by accumulating the scores 5-point Likert scale	Skills Attitudes	12 assessment tasks 44 items in background questionnaire
Internet Reading Assessment: Online research comprehension Assessment ("Internetlukemisen arviointitesti (ILA))	Kiili et al.	2018	Finland	2014-2015	426	6	12-13	3	1.1., 1.2, 1.3., 2.1, 3.2	50 min	CBA (closed Internet environment)	Scoring rubric (0-2 points; 0-5 points)	Skills	16
A new ICT literacy Test for primary and middle school students	Kim, Ahn, Kim	2019	Korea	2016	6383 primary school 9183 Middle school	4-9	10-15	Not mentioned	1.1., 1.2., 1.3., 2.1., 3.1., 5.1.	40 min	Online	Scoring → cutt-off scores for achievement levels	Skills (IR)	35
New digital empowerment scale	Kong, Wang, Lai	2019	China	Not mentioned	328	4-6	Not mentioned	1	3.1, 4.4, 5.2, 6.2	Not mentioned	Paper-and-pencil questionnare Online questionnaire	5-point-Likert scale	Attitudes Skills	16
Digital Fluency Scale	Chou & Chiu	2020	Taiwan	Not mentioned	666	6	11-12	1	1.1, 1.2, 1.3, 2.1, 2.2, 2.3, 3.1, 3.2, 4.2, 4.4, 5.2, 5.3	8 min	IR	7-point-Likert scale	Skills	25
Longitudinal assessment of digital literacy	Lazonder Walraven Gijlers Janssen	2020	The Netherlands	2016-2017-2018 (longitudal study)	151	5-6	8-11	2 & 3	1.1, 1.2, 3.1, 3.2	50 min	a) Paper-and-pencil (internet test) b) CBA Word assignment (transforming information) c) CBA PowerPoint (creating information)	Correct/successful = 1 Incorrect/unsuccessful = 0 (rubric)	Skills	16 + 2 performance-based tasks +reading comprehension test + math test

Note.
*Type test: 1= self-report; 2 = assessment based on traditional types of questions; 3 = performance based assessment;
**Measured competences: codes are based on *Table 2 and 3.3.*