United Nations
Educational, Scientific and
Cultural Organization

UNESCO INSTITUTE FOR STATISTICS

SUSTAINABLE DEVELOPMENT GOALS

Information Paper No. 56
January 2019
UIS/2019/LO/IP/56

Recommendations on Assessment Tools for Monitoring Digital Literacy within UNESCO's Digital Literacy Global Framework

UNESCO

The constitution of the United Nations Educational, Scientific and Cultural Organization (UNESCO) was adopted by 20 countries at the London Conference in November 1945 and entered into effect on 4 November 1946. The Organization currently has 193 Member States and 11 Associate Members.

The main objective of UNESCO is to contribute to peace and security in the world by promoting collaboration among nations through education, science, culture and communication in order to foster universal respect for justice, the rule of law, and the human rights and fundamental freedoms that are affirmed for the peoples of the world, without distinction of race, sex, language or religion, by the Charter of the United Nations.

To fulfil its mandate, UNESCO performs five principal functions: 1) prospective studies on education, science, culture and communication for tomorrow's world; 2) the advancement, transfer and sharing of knowledge through research, training and teaching activities; 3) standard-setting actions for the preparation and adoption of internal instruments and statutory recommendations; 4) expertise through technical cooperation to Member States for their development policies and projects; and 5) the exchange of specialized information.

UNESCO Institute for Statistics

The UNESCO Institute for Statistics (UIS) is the statistical office of UNESCO and is the UN depository for global statistics in the fields of education, science, technology and innovation, culture and communication.

The UIS was established in 1999. It was created to improve UNESCO's statistical programme and to develop and deliver the timely, accurate and policy-relevant statistics needed in today's increasingly complex and rapidly changing social, political and economic environments.

This paper was written by Mart Laanpere, Senior Researcher, Tallinn University, Centre for Educational Technology.

Published in 2019 by:

UNESCO Institute for Statistics
P.O. Box 6128, Succursale Centre-Ville
Montreal, Quebec H3C 3J7
Canada

Tel: +1 514-343-6880
Email: uis.publications@unesco.org
http://www.uis.unesco.org

Ref: UIS/2019/LO/IP56

Abstract. One of the UN Sustainable Development Goals aims at substantially increasing the percentage of the world's population who have achieved at least a minimum level of proficiency in digital literacy. Currently, there are no standardised instruments for monitoring the SDG digital literacy indicator in accordance with UNESCO's Digital Literacy Global Framework. This report draws on an analysis of alternative methodological approaches and existing instruments for assessing digital literacy skills to recommend a suitable solution for the UNESCO Institute for Statistics.

Table of contents

Introduction

This report summarises a desk research project that aims to advise the UNESCO Institute for Statistics (UIS) in designing an instrument for the assessment of digital literacy skills in the context of collecting data on Sustainable Development Goal (SDG) Indicator 4.4.2. The SDG Target 4.4 aims to "*By 2030, substantially increase the number of youth and adults who have relevant skills, including technical and vocational skills, for employment, decent jobs and entrepreneurship*" and contains three indicators (UIS, 2017):

- *4.4.1 Proportion of youth/adults with information and communications technology (ICT) skills, by type of skill*
- *4.4.2 Percentage of youth/adults who have achieved at least a minimum level of proficiency in digital literacy skills*
- *4.4.3 Youth/adult educational attainment rates by age group, economic activity status, levels of education and programme orientation*

The UIS is responsible for the development and validation of new methodologies for indicators under SDG Target 4.4. While Indicators 4.4.1 and 4.4.3 have been already implemented in reporting for 2017, the status of Indicator 4.4.2 is still "*under development*" (UIS, 2017). Although many countries have been collecting statistical data on the digital skills or ICT literacy of their citizens for various purposes, there is no common agreement on what constitutes a "minimum or basic level" of proficiency in digital literacy that would allow the aggregation of national data on the global level. As a result, there exists a serious knowledge gap about the global state of digital literacy skills of youth and adults even though these skills play an increasingly important role in achieving the SDG target. And thus, it is important to develop assessment tools for monitoring digital literacy within UNESCO's Digital Literacy Global Framework (DLGF).

A relevant category of digital literacy assessments is supra-national frameworks and policy indicators for measuring digital skills. Some supra-national initiatives or services exist in the field of digital competence assessment, but they have focused on international research projects involving several countries (e.g. the International Computer and Information Literacy Study, ICILS) or professional certification in specific occupational fields (e.g. the International Computer Driving Licence, ICDL). Such supra-national initiatives could definitely inform the UIS in designing a global instrument for collecting reliable and valid data for the SDG digital literacy target, but none of them has been specifically designed to inform Indicator 4.4.2.

The closest to the DLGF is the new standard on digital competence framework for European citizens, DigComp, which has already been used for various purposes in several European countries (Carretero et al., 2017). European agencies DG Connect and EuroStat used DigComp to redesign their Digital Skills Indicator (DSI) in 2015. The DSI survey is carried out on a sample that is representative of the general population of a country, asking respondents about digital activities carried out within the previous three months and assuming that "*persons having realised certain activities have the corresponding skills*" (European Commission, 2016). The DSI instrument defines three levels of proficiency: below basic, basic and above basic. However, there is no common European instrument for a knowledge-based or performance-based assessment of

digital competence of citizens based on DigComp. A recent report suggests using DigComp as the foundation for the DLGF, while expanding it by five additional competences and adding two competence areas.

The UIS has commissioned a report, *A Global Framework of Reference on Digital Literacy Skills for Indicator 4.4.2* (Law et al., 2018), to develop the UIS' framework for digital literacy. The report reviewed digital literacy assessment frameworks used in 47 countries and summarised consultations with experts.

The report raises three challenges. The first is the need for mapping existing instruments for digital skills assessment to the DLGF, pointing out that "*there is not a one-size-fits-all assessment of digital competence that can serve all purposes and contexts*". Second, the report calls for cost-effective, cross-national research and development (R&D) programmes to develop and validate "*context-sensitive and fit-for-purpose digital literacy indicators and assessment instruments*". Third, the report points out the discrepancy between the proficiency levels and related measurement scales of the SDG indicator versus DigComp. While SDG Indicator 4.4.2 focuses on a "*minimum level of proficiency*", DigComp discriminates between eight proficiency levels.

These three challenges raised by Law et al. are addressed by this current desk research exercise that has the following three objectives:

- Mapping existing digital literacy assessments to the DLGF.
- Evaluating the advantages and disadvantages of selected assessments that cover a large part of the DLGF, with emphasis on their cost-effectiveness for rollouts on the population scale.
- Recommending the next steps on developing an assessment tool suitable for Indicator 4.4.2.

1. Mapping existing national and cross-national assessments of ICT and digital literacy skills against the competencies of the DLGF

1.1 Definition of the construct

Digital literacy is a relatively new concept that entered to the semantic space that was already partly occupied by competing concepts such as literacy or competence in information and communications technology (ICT), media, information and computer use. Ferrari (2013) was among the first authors who tried to settle the relationship between these existing labels and newcomers (such as digital literacy/competence) in a similar manner with the definition suggested by Law et al. (2018): "*Digital literacy is the ability to access, manage, understand, integrate, communicate, evaluate and create information safely and appropriately through digital technologies for employment, decent jobs and entrepreneurship. It includes competences that are variously referred to as computer literacy, ICT literacy, information literacy and media literacy.*"

This definition builds on previous practices by incorporating vocabulary from predecessors (e.g. from information, media and ICT literacy frameworks), resulting with a list of 26 competences grouped into 7 areas of competence. As experience with European DigComp has demonstrated, such a competence framework can be used for various pragmatic purposes: redesigning outdated curricula and professional development programmes, developing policy indicators, professional accreditation, recruitment, and, to a lesser extent,

research. Such "pragmatic" competence models are usually created by a panel of experts using a top–down design approach, resulting in relatively large sets of competences and assessment rubrics that involve five to eight proficiency levels. These models suit well the purpose of curriculum design or developing diagnostic self-assessment tools. However, it is quite challenging to come up with a scientifically reliable and valid performance-based test that measures the 22 competences of DigComp on 8 proficiency levels (where each competence is a separate construct to be measured independently). Those who value pragmatically designed competence assessments tend to be more interested and proficient in this domain, making the overall results biased towards assessment over other approaches such as will be described below.

As an alternative to above described pragmatic approach, recent **psychometric** approaches to measuring digital literacy have been guided by Multidimensional Item Response Theory (MIRT) that understands 'computer and information literacy' (ICILS, 2013) or 'digital information literacy' (Sparks et al., 2016) as a single latent trait that cannot be directly observed in test situations and thus should be inferred indirectly through the statistical analysis of the test results. Like any mathematical model, MIRT makes particular assumptions to make valid inferences on the basis of the test results. For instance, the monotonicity assumption requires that the instrument does not make knowledgeable persons more likely to participate in the test (Chenery et al., 1988). The assumption of local independence means that performance in one item in a test does not influence performance in other items. While such assumptions are relatively easier to guarantee in case of knowledge-based multiple-choice tests, the same might be quite difficult in case of authentic performance-based assessment. Psychometric competence assessment frameworks are often created in a bottom–up manner, by experimenting with a large set of test items and grouping these on the basis of factor analysis performed on pilot test results. Such assessment instruments (for instance, ICILS and iSkills) are usually focused on a single digital literacy construct, which might still be multidimensional (having several aspects). For instance, in the ICILS framework, the Computer and Information Literacy construct (CIL) is broken down into two strands that each have three to four aspects:

- Strand 1: Collecting and managing information
    - Aspect 1.1: Knowing about and understanding computer use
    - Aspect 1.2: Accessing and evaluating information
    - Aspect 1.3: Managing information
- Strand 2: Producing and exchanging information
    - Aspect 2.1: Transforming information
    - Aspect 2.2: Creating information
    - Aspect 2.3: Sharing information
    - Aspect 2.4: Using information safely and securely

Although it is possible to map all of the CIL aspects to DLGF competences, the completeness and coherence of this mapping would be quite difficult to validate. Furthermore, the DLGF seems to interpret each competence as an independent construct that should be measured separately.

The two approaches to digital literacy assessment that were described above (pragmatic vs psychometric) illustrate a struggle between external and internal validity in the context of educational assessment. Validity in general is understood as the degree to which test results can be interpreted and used according to the stated purposes of the assessment (AERA et al., 2014). Internal validity refers to the methodological correctness and coherence of a research instrument, while external validity can be interpreted as its reusability through relevance or usefulness for a wider audience.

The pragmatic approach to defining and measuring digital literacy tends to result in weaker internal validity, but better external validity of the assessment instrument, as it is better understood, accepted and adopted by various stakeholders (most of whom may not have backgrounds in mathematical statistics or psychometry). For instance, self-assessment inventories and portfolios tend to prioritise the concerns and expectations of end users by providing them support in self-diagnosing their skill gaps or identifying relevant career and training opportunities. Often such instruments are used for formative assessment through respondents' own initiative, resulting in larger but biased samples (persons who are less competent or uninterested in self-development do not use it).

On the other hand, the psychometric approach guarantees higher internal validity of the assessment, often at the expense of reduced external validity and practical utility for the wider public.

Law et al. (2018) recommend using Pathway Mapping methodology for operationalising the DLGF, focusing on users' perception of digital literacy in various contexts, concerned with the external validity of the assessment. Eventually, the digital literacy assessment based on the DLGF will have to address the challenge of balancing internal and external validity, both through methodological considerations and the design of the digital literacy assessment instrument.

1.2 Framework for analysing existing digital literacy assessments

In order to decide on a methodology for selecting existing digital literacy assessments and mapping these to the DLGF, previous studies that had similar goals were reviewed. Carretero, Vuorikari and Punie (2017) reviewed 22 existing instruments that are used for assessment of digital competence in line with the DigComp framework in various European countries. They grouped these instruments into three major categories based on the data collection approach:

- **Performance assessment**, where individuals are monitored by human observers or software while being engaged in solving authentic, real-life problems by using common software tools (e.g. browser, word processor, spreadsheet) or simulations.
- **Knowledge-based assessment**, where individuals are responding to carefully designed test items that measure both declarative and procedural knowledge.
- **Self-assessment**, where individuals are asked to evaluate their knowledge and skills with questionnaires that range from structured scales to free-form reflection.

These approaches can be strengthened by combining them and also by conducting secondary approaches, for example, by providing an electronic portfolio (e-portfolio) that contains creative works, certificates and other authentic documentary evidences. However, such combinations would definitely increase the cost and duration of assessment, while decreasing its scalability. Carretero et al. did not map the selected instruments exactly to the DigComp model on a detailed level of granularity to specific competences and proficiency levels, but their work did provide guidance to some relevant assessments that have not yet been published.

An alternative set of digital literacy assessments and their analysis frameworks is suggested by Sparks et al. (2016), who consider both the focus of assessment and the design of the item:

1. Assessments having an **information literacy** focus with multiple-choice and constructed response items (such as ILT, SAILS, RSA, ISS, ILAAP and ICILS).
2. Assessments having a **technology literacy** focus with multiple-choice and constructed response items (such as IC3 and ICDL/ECDL).
3. Assessments having a **digital information literacy** focus with performance-based tasks (such as CLA+, PIAAC PS-TRE and iSkills).

Sparks et al. (2016) also discuss variations in the design of digital literacy assessment instruments caused by the tests having different **purposes**. Some assessments are designed for research purposes, while others support accreditation or institutional quality assurance and curriculum design. Yet another purpose is self-assessment to support professional development, where the instrument has been designed to meet the needs of the test-taker as the primary user. In this last case, the flexibility and usability from a respondent's perspective are prioritised at the expense of scientific reliability and the internal validity of the instrument.

Another alternative type of digital literacy assessments is presented by van Deursen et al. (2017), who are mainly concerned with differences in item design and scales:

Type 1: Self-reporting surveys with questions that ask for the frequency of technology use, which are assumed to deliver indirect evidence of the command of skills. When an individual uses an application that is considered to be difficult to use (or uses a large variety of applications), this is held to be an indication of a high level of skills. An example of this kind of assessment is EuroStat's Digital Skills Index.

Type 2: Surveys with questions that request self-assessments of skills on a predefined scale. This is the most commonly used method. Some examples are the Digital Skills Wheel by DigitalDannelse and the MDS questionnaire by van Deursen.

Type 3: Performance tests in a laboratory or another controlled environment that provide subjects with particular assignments to observe their command of internet skills, for example, the PIAAC PS-TRE test.

The most thorough and relevant approach for analysing digital literacy assessment tools was applied by Siddiq et al. (2016). They addressed the differences between ICT literacy assessments in the primary and secondary school contexts in reliability, validity, the match to DigComp (on the level of a single competence) and interaction types used in the item design. The analysis criteria used by Siddiq were:

- Country of implementation
- Year of implementation
- Sample size in empirical study
- Target group: primary, secondary or high school (there was no interest in adults)
- Item interaction type: multiple choice, interactive, authentic
- Framework used for competence modelling
- Availability of test items or their descriptions
- Match with the adapted DigComp model
- Data: qualitative, quantitative or mixed
- Duration of the test (in minutes)

This study for the UIS addresses the three categories of instruments for digital literacy assessment, as described by Carretero et al. (2016), to identify existing practices and evaluate their applicability in the context of data collection for Indicator 4.4.2. The applicability analysis is mainly inspired by Siddiq et al. (2016) and is focused on the relevance (purpose, match with the DLGF and cost-effectiveness) of the given instrument, but also considers its reliability and validity, following the discussion above. The existing digital literacy assessment practices and instruments were searched from three types of sources:

1. Scientific research publications
2. Policy documents in the education and employment domains
3. Professional certification frameworks and related technical documents

As a result of this search, 44 assessments were selected for closer review: 30 from Siddiq's sample (leaving out the assessments that were older than 10 years) and 14 additional ones *(see Annex 1).* While all the assessments in Siddiq's original sample targeted students in primary and secondary schools, the additional 14 instruments have been designed for the wider population.

In addition to the indicators used by Siddiq (see the list above), two additional analytical criteria were applied, as inspired by the discussion above:

- The purpose of assessment:
  - **Research**, mostly international comparative studies
  - **Credentialing** the digital skills, mostly for employment and qualification
  - Educational **statistics**, mostly for policy planning/evaluation or quality assurance of the educational system
  - **Diagnostic** self-assessment, mainly for personal use
- The focus of assessment:
  - **Technical skills**: manipulation of common software applications, e.g. word processing, spreadsheets, presentations, internet browsers and e-mail clients

- **Information literacy**: a single construct that has several facets or dimensions related to skills to find, evaluate and make use of information
- **Digital competence**: a set of multiple (20+) competencies that represent various capabilities of a person to solve problems in an authentic context using digital technologies

All 44 digital literacy assessment instruments were mapped to the DLGF and annotated with the categories from Siddiq, as well as the two additional category sets: the purpose and focus of the assessment. The results of this mapping exercise are described in the next section.

1.3 Overview of existing assessments and their mapping to the DLGF

The current analysis involved 44 assessments of digital literacy, 7 of which were designed for and implemented on a global or multi-country level. The oldest instruments date back to 2006, while the most recent ones were developed in 2017. The main body of cases (30) was taken from Siddiq's paper that focused on digital literacy testing only in primary and secondary schools. An additional 14 cases mostly represent the digital literacy tests that target the wider population: two instruments specifically target university students, five are designed for adult employees, and two have been implemented both in schools and with adult respondents. While a few of the assessments in our selection were implemented on relatively small samples (50 to 60 respondents), the majority of the instruments have been tested with thousands of users. The selection of digital literacy assessments is designed to include various alternative designs with different purposes (R – research, S – statistics, C – credentialing, D – diagnostic self-assessment) or focuses (T – technical skills, I – information literacy, D – digital competence). There are also variations in item types (MC – multiple choice items, INT – interactive/dynamic items, AUTH – authentic performance) and the duration of the test, ranging from 20 to 180 minutes. Most of the instruments are designed in accordance with a specific digital/information literacy framework or national curriculum (in case of school-oriented tests). Only in a few cases was this framework explicitly DigComp. In 18 cases out of 44, the set of test items were explicitly available for analysis and mapping with the DLGF, while in the remaining cases, analysis and mapping relied on information about the specific competence within a framework or curriculum and examples of tasks.

The method of mapping the specific assessment to the DLGF involved comparing the test items (or underlying assessment framework) with DLGF competencies and judging the potential match. Proficiency levels were not included in analysis, only the content of the description of a competency.

2. Analysis of advantages and disadvantages of different approaches to assessment

In the context of global assessment of digital literacy and the SDGs, it is likely that performance assessment and additional analysis of secondary (qualitative) data are neither cost-effective nor scalable approaches to digital skills assessment. Self-reporting would be the easiest and most cost-effective to implement, but it could suffer from low reliability and internal validity. However, it should be possible to combine self-assessment with knowledge-based or performance assessment. For instance, Põldoja et al. (2014) have designed and validated an instrument called DigiMina that combines self-assessment of teachers' digital

competence with peer-assessment, knowledge-based tests and an e-portfolio containing teacher's reflections and creative works. Within the DigCompEdu project, the European Commission's Joint Research Centre (JRC) tried to balance internal and external validity of assessment of school's digital capability with the design of the SELFIE tool, so that schools were allowed to expand the scientifically validated core instrument with additional items from a pre-designed, publicly available pool or even design their own additional items that seemed relevant to them (JRC, 2018). The future instrument that will be designed by the UIS for digital literacy assessment might also benefit from a similar balancing of the need for global standardisation (contributing to internal validity) and the local context (contributing to external validity).

The purpose of digital literacy assessment has a major impact on the design of the instrument, its items and testing procedures. For instance, the instruments that are designed for research purposes apply a mostly psychometric approach (as described in Section 1.1). This results in large efforts invested in item design and validation, sampling and data collection procedures, and the analysis of results. The test items are mostly knowledge-based, combining multiple choice and interactive tasks, and in a few occasions are also simulations (PIAAC PS-TRE, OECD PISA). The data are collected in a short time-frame, in a controlled environment (by group in computer labs). This type of digital literacy assessment is costly, often funded directly by governments and repeated infrequently (once every three to five years).

Digital literacy assessments designed for statistical purposes (e.g. EuroStat DSI) are also administered to a carefully composed sample, but the design of the instrument and items (mostly multiple-choice items) is significantly cheaper and allows frequent use, as it is expected that low effort is required from respondents and analysts. Data collection for statistical purposes usually takes place in a wider and more flexible time-frame in loosely controlled settings. Compared with research-focused digital literacy assessments, statistical data gathering might be less reliable and have weaker internal validity.

Digital literacy assessments designed for credentialing purposes (e.g. ICDL/ECDL) mainly address the needs of employers or academic institutions who are increasingly more concerned about the comparability of digital skills of their employees or students. This is why credentialing-focused assessments are designed for frequent and scalable use in highly controlled environments. A high level of standardisation and a focus on universal, practical technical skills (e.g. word processing and the use of spreadsheets) increases the reusability of items and instruments over time and across nations, thus reducing the cost that is left to be covered, mainly by test-takers. Although the credentialing-type of digital literacy assessment is usually relatively reliable, its construct validity and internal validity are likely weaker than those of research-focused instruments.

Diagnostic assessments of digital literacy do not aim for high reliability and internal validity; instead, they prioritise external validity and the perceived usefulness of assessment to the end users (test-takers). Diagnostic self-assessments (e.g. DigiMina, DigitalCompetenceWheel and MDS) are likely to include mainly self-reporting scales and sometimes include authentic tasks. Authentic assessment reduces the possibilities for automatic scoring and requires the involvement of human reviewers. This is why such assessments are costly and are not scaled up, and are mostly used at the level of a single organization. However, diagnostic

self-assessment that uses only multiple-choice items (e.g. DigitalCompetenceWheel and MDS) is scalable and low cost, though its drawbacks include low reliability and biased sampling, as persons with low self-esteem and poor digital skills are unlikely to take the test.

Looking from the perspective of the UNESCO DLGF, it is possible to combine two or three types of digital literacy assessment. As the aim of using Indicator 4.4.2 is to establish "*the minimum level of proficiency in digital literacy*" among the population of all countries – and evaluate the percentage of youth/adults who have achieved it – an instrument designed with a statistical purpose seems to be the main priority. However, the uptake and external validity of the future DLGF-based assessment can be increased by adding some elements of a credentialing-focused approach or a diagnostic self-assessment approach.

It can be concluded that the best match to the needs of monitoring Indicator 4.4.2 in line with the DLGF was found in these existing assessments:

- The easiest tool to adapt to meet the minimum needs of the DLGF: the DSI survey by EuroStat.
- The widest coverage of all competences in the DLGF with knowledge-based test items: the DigComp test for 9th and 12th grade in Estonian schools; it covers the whole DigComp framework with automatically scored items and has both self-reporting and knowledge-based test components, but the reliability of items related to competence areas 3 to 5 are weak, and it does not at all cover DLGF areas 0 and 6. The items are available only in Estonian.
- The best existing open-source platform that can be adapted for the DLGF: the PIX test in France (pix.fr) has an advanced and user-friendly technical platform (e.g. allowing timed test-parts) and flexible item design (including support to adaptive testing). The existing item set does not cover competence areas 5 and 6 in the DLGF. The items are available in French.
- The most attractive and user-friendly interface: the Digital Competence Wheel in Denmark covers most of the DLGF competence areas and gives attractive visual feedback to users. Its disadvantages are that there are no knowledge-based items and it is only self-reporting.

3. Recommendation on assessment tools for monitoring Indicator 4.4.2

The specific goal of monitoring Indicator 4.4.2. does not require a complex instrument, as it is enough to conduct a simple survey of digital practices on a carefully designed, representative sample in each country. However, such a survey would not motivate or help respondents (and even less the majority of the population who were left outside of the survey sample) to improve their digital literacy skills. This is why it is recommended to extend the core instrument with additional self-assessment and knowledge-testing tools to follow the DLGF-based digital literacy assessment.

Based on the analysis above, five general recommendations are provided regarding the design and delivery of a DLGF-based digital literacy assessment:

- The main assessment of Indicator 4.4.2 should be based on self-reporting on a usage scale that measures a single construct of digital literacy on "*minimum level of proficiency*", similar to EuroStat DSI; the survey could be carried out on a large representative sample in a controlled environment online, using a mobile app or on paper and would take no longer than 15 minutes to do.

- The main instrument should be extended with a voluntary self-assessment similar to the Danish Digital Competence Wheel, with self-reported responses on a five-point scale. The test duration should not exceed 10 to 15 minutes. Results would be given with recommendations on how to improve one's digital literacy skills in a local context, such as through courses and using learning resources in a mother tongue.

- The test should be piloted with at least 1,000 respondents in 3 different languages and 3 different countries to validate the items. An international steering group of researchers should be formed to evaluate and improve the quality of items and procedures, including the reliability and validity of the assessment.

- To increase the external validity, the UIS could enhance the self-reporting instrument with a separate, additional knowledge-based online test (potentially as an adaptation of the open-source PIX software); knowledge-based test items should be composed in line with the Evidence-Centered Assessment Design (ECD) approach and be ready for Multidimensational Item Response Theory (MIRT) analysis. The respondent can share the assessment report and related microcredentials (e.g. Open Badges).

- The software architecture for the knowledge-based testing tool should be similar to PIX, with a built-in data upload to the UIS database (in processed, anonymised form). Software and test items should follow annual versioning in Github. The user interface should be responsive; test runs can also be done on smartphones and tablets; items and tests should be available for export in IMS QTI 2.2 format, allowing migration of the test to existing national assessment platforms. Special attention is needed to meeting high national (and EU) standards for privacy and data protection: data is anonymised in an early stage, no private information of respondents is stored. The Application Programming Interface (API) should enable potential extensions or integration of the knowledge-based testing tool with e-portfolios for diagnostic self-assessment.

4. Conclusion

Out of the large variety of existing digital literacy assessment instruments, only a few are suitable for measuring the minimum level of proficiency on a global scale, in line with the needs of the UIS in the context of monitoring Indicator 4.4.2. This report analysed 44 existing assessments of digital literacy skills, which resulted in a list of five recommendations for developing a suitable instrument. Among these existing instruments, the Digital Skills Indicator by EuroStat provided the closest match to the needs of the UIS regarding the monitoring cost, methodological coherence and coverage of 27 digital literacy skills included in the DLGF. However, the impact of the DLGF would be significantly increased if this core survey instrument would be enhanced with a more advanced online self-assessment tool and also with an optional knowledge-based test.

References

ACARA (2012). National Assessment Program – ICT Literacy Technical Report 2011. Sydney: Australian Curriculum, Assessment and Reporting Authority. http://www.nap.edu.au/verve/_resources/NAP_ICTL_2011_Technical_Report_Final.pdf.

ACARA (2012). National Assessment Program – ICT Literacy Years 6 and 10 Report 2011. Sydney: Australian Curriculum, Assessment and Reporting Authority. http://www.nap.edu.au/verve/_resources/nap_ictl_2011_public_report_final.pdf.

Aesaert, K., van Nijlen, D., Vanderlinde, R., and van Braak, J. (2014). Direct measures of ICT validation of an ICT competence scale. *Computers and Education, 76,* 168–181.

Ahonen, A. K., and Kinnunen, P. (2015). How do students value the importance of twenty-first century skills? *Scandinavian Journal of Educational Research, 59,* 395–412.

Ala-Mutka, K. (2011). Mapping Digital Competence: Towards a conceptual understanding. Seville: European Commission, Joint Research Centre. Institute for Prospective Technological Studies. http://ipts.jrc.ec.europa.eu/publications/pub.cfm?id.4699.

Beile, P. (2008). Information literacy assessment: A review of objective and interpretive measures. In K. McFerrin, R. Weber, R. Carlsen, and D. A. Willis (Eds.), Proceedings of society for information technology and teacher education international conference (SITE) 2008, 1860–1867. Chesapeake, VA: AACE.

Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., et al. (2012). Defining twenty-first century skills. In P. Griffin, B. McGaw, and E. Care (Eds.), Assessment and teaching of 21st century skills, 17–66. Netherlands: Springer.

Calvani, A., Fini, A., Ranieri, M., and Picci, P. (2012). Are young generations in secondary school digitally competent? A study on Italian teenagers. *Computers and Education, 58,* 797–807.

Cartelli, A. (2010). Frameworks for digital competence assessment: Proposals, instruments, and evaluation. Paper presented at the Proceedings of Informing Science and IT Education Conference (InSITE). http://proceedings.informingscience.org/InSITE2010/InSITE10p561-574Cartelli861.pdf.

Castillo, M. (2010). Technological literacy: Design and testing an instrument to measure eight-grade achievement in the technology education (Paper presented at the American Society for Engineering Education, Louisville, KY).

Cha, S. E., Jun, S. J., Kwon, D. Y., Kim, H. S., Kim, S. B., Kim, J. M., et al. (2011). Measuring achievement of ICT competency for students in Korea. *Computers and Education, 56,* 990–1002.

Chu, S. K. W. (2012). Assessing information literacy: A case study of primary 5 students in Hong Kong (Vol. 15). American Association of School Librarians. http://www.ala.org/aasl/slr/volume15/chu.

Claro, M., Preiss, D. D., San Martín, E., Jara, I., Hinostroza, J. E., Valenzuela, S., et al. (2012). Assessment of 21st century ICT skills in Chile: Test design and results from high school level students. *Computers and Education, 59,* 1042–1053.

Dagiene, V., and Stupuriene, G. (2014). Informatics education based on solving attractive tasks through a contest. http://www.bebras.org/sites/default/files/documents/publications/Dagiene%202014.pdf (Accessed 1 Feb 2015).

Darling-Hammond, L., and Adamson, F. (2010). Beyond basic skills: The role of performance assessment in achieving 21st century standards of learning. Stanford, CA: Stanford University, Stanford Center for Opportunity Policy in Education. https://scale.stanford.edu/system/files/beyondbasic-skills-role-performance-assessment-achieving-21st-century-standards-learning.pdf.

Davis, C. H. F., Deil-Amen, R., Rios-Aguilar, C., and Canche, M. S. G. (2012). Social media in higher education. A literature review and research directions. University of Arizona and Claremont Graduate University. https://works.bepress.com/hfdavis/2/.

Dede, C. (2009). Comparing frameworks for "21st century skills". Boston: Harvard Graduate School of Education. http://www.watertown.k12.ma.us/dept/ed_tech/research/pdf/ChrisDede.pdf.

Duncan, G. J., Engel, M., Claessens, A., and Dowsett, C. J. (2014). Replication and robustness in developmental research. *Developmental Psychology, 50(11),* 2417–2425.

Erstad, O. (2010). Educating the digital generation. *Nordic Journal of Digital Literacy, 5,* 5671. http://www.idunn.no/ts/dk/2010/01/art05.

Eshet-Alkalai, Y., and Chajut, E. (2010). You can teach old dogs new tricks: The factors that affect changes over time in digital literacy. *Journal of Information Technology Education, 9,* 176–181.

European Commission. (2014). The international computer and information literacy study (ICILS). Main findings and implications for education policies in Europe. European commission, Education and Training. http://ec.europa.eu/education/library/study/2014/ec-icils_en.pdf (Accessed 6 Dec 2015.)

Ferrari, A. (2013). DIGCOMP: A framework for developing and understanding digital competence in Europe. In Y. Punie, and B. N. Brečko (Eds.), JRC scientific and policy reports. Luxembourg: Publications Office of the European Union.

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

Fraillon, J., Schulz, W., and Ainley, J. (2013). International computer and information literacy study: Assessment framework. Amsterdam: IEA.

Futschek, G., and Dagiene, V. (2009). A contest on informatics and computer fluency attracts school students to learn basic technology concepts. http://www.bebras.org/sites/default/files/documents/publications/Futschek-2009.pdf.

Gallardo-Echenique, E. E., de Oliveira, M. J., Marquès-Molias, L., and Esteve-Mon, F. (2015). Digital competence in the knowledge society. *MERLOT Journal of Online Learning and Teaching, 11(1).* jolt.merlot.org/vol11no1/Gallardo-Echenique_0315.pdf.

Goldhammer, F., Naumann, J., and Keßel, Y. (2012). Assessing individual differences in basic computer skills: Psychometric characteristics of an interactive performance measure. *European Journal of Psychological Assessment, 29,* 263–275.

Griffin, P., and Care, E. (2015). Assessment and teaching of 21st century skills. Methods and approach. In Series education assessment in an information age. Melbourne, Australia: Springer. http://dx.doi.org/10.1007/978-94-017-9395-7.

Griffin, P., McGaw, B., and Care, E. (2012). Assessment and teaching of 21st century skills. Melbourne, Australia: Springer. http://dx.doi.org/10.1007/978-94-007-2324-5.

Gudmundsdottir, G., and Hatlevik, O. E. (2013). Digital competence and students' productive use of computers in school: Observing the role of motivation and family background. In E. Bjørnestad, and J. Heldal Stray (Eds.), New voices in Norwegian educational research, 69–81. The Netherlands: Sense Publishers.

Hakkarainen, K., Ilomäki, L., Lipponen, L., Muukkonen, H., Rahikainen, M., Tuominen, T., et al. (2000). Students' skills and practices of using ICT: Results of anational assessment in Finland. *Computers and Education, 34,* 103–117.

Hatlevik, O. E., Ottestad, G., and Throndsen, I. (2015). Predictors of digital competence in 7th grade: A multilevel analysis. *Journal of Computer Assisted Learning, 31,* 220–231.

Hesse, F., Care, E., Buder, J., Sassenberg, K., and Griffin, P. (2015). A framework for teachable collaborative problem solving skills. In P. Griffin, and E. Care (Eds.), Assessment and teaching of 21st century skills, educational assessment in an information age.

Hohlfeld, T. N., Ritzhaupt, A. D., and Barron, A. E. (2010). Development and validation of the student tool for technology literacy (ST2L). *Journal of Research on Technology in Education, 42,* 361–389.

Hohlfeld, T. N., Ritzhaupt, A. D., and Barron, A. E. (2013). Are gender differences in perceived and demonstrated technology literacy significant? It depends on the model. *Education Technology Research and Development, 61,* 639–663.

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

Kaasbøll, J. (2012). Developing digital competence – Learning, teaching and supporting use of information technology. Oslo: University of Oslo. http://electronics.wesrch.com/pdfEL11TZFGRQBHU (Accessed 13 Oct 2014.)

Kane, M. T. (2006). Validation. In R. L. Brennan (Ed.), Educational measurement, 4th ed., 17–64. Westport, CT: American Council on Education/Praeger Publishers.

Kane, M. T., Crooks, T. J., and Cohen, A. S. (1999). Validating measures of performance. *Educational Measurement: Issues and Practice, 18,* 5–17.

Katz, I. R. (2007). Testing information literacy in digital environments: The ETS iSkills™ assessment. Information Technology and Libraries, 26, 3–12.

KERIS (Korea Education and Research Information Service) (2013). White paper on ICT in education Korea. The Ministry of Education and Korea Education and Research Information Service. http://english.keris.or.kr/whitepaper/WhitePaper_eng_2013.pdf.

Kim, J., and Lee, W. (2013). Meanings of criteria and norms: Analyses and comparisons of ICT literacy competencies of middle school students. *Computers and Education, 64,* 81–94.

Kiss, G. (2012). Measuring Hungarian and Slovakian students' IT skills and programming knowledge. *Acta Polytechnica Hungarica, 9,* 195–210.

Klieme, E. (2004). Assessment of cross-curricular problem-solving competencies. In H. Moskowitz, and M. Stephens (Eds.), Comparing learning outcomes. International assessments and education policy, 81–107. London: Routledge Falmer.

Kuhlemeier, H., and Hemker, B. (2007). The impact of computer use at home on students' Internet skills. *Computers and Education, 49,* 460–480.

Larres, P. M., Ballantine, J., and Whittington, M. (2003). Evaluating the validity of self-assessment: Measuring computer literacy among entry-level undergraduates within accounting degree programmes at two UK universities. *Accounting Education: An International Journal, 12,* 97–112.

Law, N., Lee, Y., and Yuen, H. K. (2009). The impact of ICT in education policies on teacher practices and student outcomes in Hong Kong. In F. Scheuermann, and F. Pedro (Eds.), Assessing the effects of ICT in education – Indicators, criteria and benchmarks for international comparisons, 143–164. France: OECD.

Lee, L., Chen, D. T., Li, J. Y., and Lin, T. B. (2015). Understanding new media literacy: The development of a measuring instrument. *Computers and Education, 85,* 84–93.

Lee, Y., Law, N., and Yuen, A. H. K. (2010). Online performance assessment of students' information literacy skills in science, mathematics and mother tongue. Paper presented at The Annual Meeting of the American Educational Research Association (AERA), Denver, Colorado. http://hkir.ust.hk/hkir/Record/9999-712468.

Lennon, M., Kirsch, I., von Davier, M., Wagner, M., and Yamamoto, K. (2003). Feasibility study for the PISA ICT literacy assessment. Report to Network A. Paris: OECD.

Li, Y., and Ranieri, M. (2010). Are “digital natives” really digitally competent? A study on Chinese teenagers. *British Journal of Educational Technology, 41,* 1029–1042.

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

MCEECDYA (2010). National assessment program – ICT literacy years 6 and 10, report 2008. Carlton South: Ministerial Council for Education, Early Childhood Development and Youth Affairs. Melbourne, Australia. http://www.nap.edu.au/verve/_resources/2008_nap_ictl_public_report.pdf (Accessed 20 Feb 2014.).

Mislevy, R. J., and Haertel, G. D. (2006). Implications of evidence-centered design for educational testing. *Educational Measurement: Issues and Practice, 25,* 6–20.

Newrly, P., and Veugelers, M. (2009). How to strengthen digital literacy? Practical example of a European initiative "SPreaD". https://joinup.ec.europa.eu/community/epractice/document/eu-how-strengthen-digital-literacy-practical-example-european-initiativ (Accessed 20 Feb 2014.)

Newton, P. E., and Shaw, S. D. (2013). Standards for talking and thinking about validity. *Psychological Methods, 18,* 301–319.

Newton, P. E., and Shaw, S. D. (2014). Validity in educational and psychological assessment. London/Cambridge: Sage Publications Ltd and Cambridge Assessment.

OECD (Organization for Economic Co-Operation and Development). (2010). PISA 2012 field trial, Problem solving framework. http://www.oecd.org/pisa/pisaproducts/46962005.pdf (Accessed 4 Feb 2014.)

Parshall, C. G., Harmes, J. C., Davey, T., and Pashley, P. J. (2010). Innovative items for computerized testing. In W. J. van der Linden, and C. A. W. Glas (Eds.), Elements of adaptive testing, 215–230. USA: Springer Science. Business Media.

Ritzhaupt, A. D., Liu, F., Dawson, K., and Barron, A. E. (2013). Differences in student information and communication technology literacy based on socioeconomic status, ethnicity, and gender: Evidence of a digital divide in Florida schools. *Journal of Research on Technology in Education, 45,* 291–307.

Sefton-Green, J., Nixon, H., and Erstad, O. (2009). Reviewing approaches and perspectives on "digital literacy". *Pedagogies: an International Journal, 4,* 107–125.

Senkbeil, M., Ihme, J. M., and Wittwer, J. (2013). The test of technological and information literacy (TILT) in the national educational panel study: Development, empirical testing, and evidence for validity. *Journal for Educational Research Online, 5,* 139–161.

Shaw, S., and Hughes, S. (2015). Issues around how best to provide evidence for assessment validity: The challenge of validation. In Pre-conference workshop at the association for educational assessment – Europe, Glasgow. http://www.aea-europe.net/images/1_Conferences/Glasgow_2015/workshop_template_2015_SDS_SH.pdf.

Short, B. (2012). 21st century skills development: Learning in digital communities – Technology and collaboration (Doctoral dissertation). Department of Educational Methodology, Policy, and Leadership and the Graduate School of the University of Oregon.

Siddiq, F. et al. (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.

Sieber, V. (2009). Diagnostic online assessment of basic IT skills in 1st-year undergraduates in the Medical Sciences Division, University of Oxford. *British Journal of Educational Technology*, *40*(2), 215–226.

Silva, E. (2009). Measuring skills for 21st-century learning. *Phi Delta Kappan, 90,* 630–634.

Silvernail, D. L., Small, D., Walker, L., Wilson, R. L., and Wintle, S. E. (2008). Using technology in helping students achieve 21st century skills: A pilot study. Center for Education Policy, Applied Research, and Evaluation University of Southern Maine and Maine School Administrative District #54.

Sink, C., Sink, M., Stob, J., and Taniguchi, K. (2008). Further evidence of gender differences in high school-level computer literacy. *Chance, 21,* 49–53. http://dx.doi.org/10.1080/09332480.2008.10722886.

Søby, M. (2013). Learning to Be: Developing and understanding digital competence. *Nordic Journal of Digital Literacy, 8,* 134–138. https://www.idunn.no/file/pdf/62414156/#page.3 (Accessed 26 Oct 2014.)

Somerville, M., Smith, G. W., and Macklin, A. S. (2008). The ETS iSkillsTM assessment: A digital age tool. *The Electronic Library, 26,* 158–171. http://dx.doi.org/10.1108/02640470810864064.

Tongori, A. (2013). Innovative assessment tools and instruments for ICT literacy. In I. Csisz_ar (Ed.), Proceedings of Spring Wind conference, Budapest, 591–598.

Tongori, A., and Pluhar, Z. (2014). An instrument to assess the basic technological actions of ICT literacy. *TEE SZEMLE, IV,* 7–20.

Van Deursen, A. J. A. M., and Van Diepen, S. (2013). Information and strategic internet skills of secondary students: A performance test. *Computers and Education, 63,* 218–226.

Van Vliet, P. J. A., Kletke, M. G., and Chakraborty, G. (1994). The measurement of computer literacy: A comparison of self-appraisal and objective tests. *International Journal of Humane Computer Studies, 40,* 835–857.

Wilson, M. (2005). Constructing Measures: An item response modeling approach. Mahwah, NJ: Lawrence Erlbaum Associates, Inc.

Wilson, M., and Scalise, K. (2015). Assessment of learning in digital networks. In P. Griffin, and E. Care (Eds.), Assessment and teaching of 21st century skills, volume 2 – Methods and approaches, 57–81. Dordrecht: Springer.

Wu, H., Kuo, C., Jen, T., and Hsy, Y. (2015). What makes an item more difficult? Effects of modality and type of visual information in a computer-based assessment of scientific inquiry abilities. *Computers and Education, 85,* 35–48.

Zelman, M., Avdeeva, S., Shmis, T., Vasiliev, K., and Froumin, I. (2011). International comparison of information literacy in digital environments. A paper presented at the International Association for Educational Assessment (IAEA), Manila. http://www.iaea.info/documents/paper_30e43f54.pdf (Accessed 3 Jan 2014.)

Zenisky, A. L., and Sireci, S. G. (2002). Technological innovations in large-scale assessment. *Applied Measurement in Education, 15,* 337–362.

Annex 1. Overview of 44 assessments of digital literacy

Country	Year	Sample	Target	Item type	Duration	Focus	Purpose
Australia	2011	11,023	school	MC/INT	120	D	C
Australia	2008	10,926	school	MC/INT	120	D	C
Belgium	2012	650	school	INT	100	D	R
Canada	2016	N/A	university	MC	N/A	I	C
Chile	2009	1,185	school	MC/INT		D	R
China, Hong Kong SAR		199	school	MC		D	R
China, Hong Kong SAR		1,622	school	INT	90	I	R
Denmark	2017	N/A	any	MC	N/A	D	D
Estonia	2017	1,450	school	MC/INT	45	D	C
Finland	2014	3,159	school	MC	60	T	C/D
France – PIX	2017	N/A	any	MC/INT	N/A	D	D
Germany	2010	855	school	MC		I	R
Germany	2009	315	school	MC/INT	120	I	R
Hungary		1,470	school	MC	60	T	R
Hungary	2013	60	school	INT	45	I	R
Israel	2007	58	school	AUTH		I	R
Italy	2010	1,056	school	MC	60	D	R
Lithuania	2017	35,600	school	MC/INT	60	D	C
Netherlands	2011	54	school	AUTH		I	R
Norway	2012	3,335	school	MC		D	R
Norway	2013	4,216	school	MC		I	R
Norway		1,793	school	MC		I	R
Norway	2010	4,087	school	MC/INT	70	I	R
Rep. of Korea		17,547	school	MC/INT	50	D	R
Rep. of Korea		143	school	MC/INT	50	T	R
Rep. of Korea		40,072	school		40	I	C
Rep. of Korea	2011	15,558	school	MC/INT	45	T	C
Rep. of Korea	2011	11,767	school			T	C
Russian Fed.	2011	398,100	school	INT	120	I	R
United Kingdom	2006	172,225	school	MC/INT	100	D	C
United States	2011	100	school	INT	45	D	R
United States	2010	5,884	school	MC/INT	30	D	R
United States	2008	350	school	MC/INT		I	R
United States	2005	1,016	school	INT	75	I	C
Global	2013	60,000	school	INT	60	D	R
NL and UK	2013	1,017	any	MC	25	D	R
EU – DSI	2015	N/A	adults	MC	20	D	S

ECDL – global		N/A	adults	MC		T	C
IC3 – global		N/A	adults	MC/INT	45	T	C
USA – ILT	2014	N/A	adults	MC	60	I	C
USA – SAILS		N/A	adults	MC	N/A	I	C
USA – RRSA		N/A	university	MC	N/A	I	R
USA – CLA+		N/A	any	AUTH	N/A	D	C
PIAAC	2013	N/A	adults	INT	N/A	D	R