Measuring Digital Literacy: Evidence for the Validity of a Conceptual Framework

Abstract

The report in focused on measuring digital literacy among secondary school students (14 year-olds). In order to measure digital literacy, we have developed a new conceptual framework and a special instrument in computer-based form based on the Evidence-Centered Design methodology (Mislevy, Levy 2006). The conceptual framework includes the cognitive, technical, and ethical aspects of digital literacy. The instrument consists of scenario-based tasks that simulate features of students' real-life situations. The assessment is considered as low-stakes and aimed to inform students, teachers, and policy-makers about digital literacy development. The validity evidence and the improvements of the instrument will be discussed.

Introduction

Digital tools are becoming increasingly important to modern society. Digital literacy (DL) is an umbrella term that encompasses an interdisciplinary combination of skills that are essential to interacting with digital tools.

We have developed a new conceptual framework for DL, based on a comprehensive literature review of DL (World Economic Forum, 2015; Digital Economy Outlook, 2018; G20, 2017; etc.) and similar constructs (Fraillon, J., Schulz, W., & Ainley, J. 2013; Katz, IR, 2013; Abbot, 2014). Our DL framework takes into account cognitive, technical and ethical factors (Chetty K., Wenwei L., Josie J., Shenglin B., 2017) and includes 5 components: (1) Access, (2) Creation, (3) Security, (4) Ethics and (5) Interfacing.

Access includes two subcomponents: Search and Analysis. Search describes effective strategies that people use for collecting information in a digital environment. The analysis is the proficiency to estimate the quality of collected information, its relevance, reliability, fullness, or redundancy. Creation refers to the skills needed to create a digital information product (such as an article, infographics, email, etc.) and includes usage of different sources of information. Security encompasses the knowledge and strategies used to stay safe in a digital environment, such as avoiding computer viruses and scams. Ethics includes knowledge of the norms of digital communication, such as. Finally, Interfacing is the capacity to understand and manipulate elements of mainstream digital tools, including shortcuts, hotkeys, special characters, etc. We conducted several expert panels, eventually forming a list of observable behaviors related to each of these components.

DL is important at all ages, but currently, we are focused on school students in the 9th grade. At that stage in their education, children start thinking about a future profession, so measuring DL creates an opportunity to intervene where necessary and help those who are falling too far behind. However, no tools are available to measure DL of Russian-speaking students, so we developed our own instrument following the evidence-centered design approach (ECD) (Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R., 2012).

The aim of this paper is to provide evidence supporting the validity of our new DL conceptual framework.

Method

Instrument

Our instrument consists of 5 scenarios that simulate familiar to adolescents' everyday-life in the context of the digital environment: Internet browser, email client, chat, presentation making platform. Each scenario includes a number of hidden indicators each related to one of our 5 targeted constructs. Initially, about 60 indicators were developed. However, during the following analysis, we elected to exclude some of them from further analyses due to poor psychometric properties.

Participants and procedure

The sample consisted of 940 9-grade adolescents (mean age = 14,7; SD = 0,46; 50,1% female) attending to 11 schools located in a big city in the central district of Russia. The average testing time was approximately 45 minutes. The scenarios were presented in a random sequence. A preliminary analysis indicated that some of the scenarios were not to presented to 6.5% of students due to technical issues. The following quantitative analysis was based on a sample of 879 students.

Statistical analysis

We applied item response theory (IRT; Hambleton, Swaminathan, & Rogers, 1991) to investigate the psychometric properties of our instrument. Multidimensional IRT-extension is useful in the context of complex types of assessment (Levy, 2013). In the study, two types of Rasch multidimensional partial credit models were applied. The first model assumed 5 correlated factors (Adams, Wilson, & Wang, 1997); the second model assumed one general factor and four uncorrelated factors (testlet model; Wang & Wilson, 2005). Further, these models were compared in terms of model-fit parameters (AIC; BIC; Burnham & Anderson, 2004). This analysis allowed to assess the internal structure of the test, and by extension, the DL framework.

As the preliminary analysis, the investigation of local independence violations was carried out. For instance, a common context for several items (indicators) might lead to an additional relationship among variables (Baghaei, 2008). There are several cases in the instrument where different actions of students related to one part of the scenario. In order to overcome possible local independence violations, these items were grouped into bundles (Rosenbaum, 1988). Bundles were analyzed as polytomous items (Quellmalz, Timms, Silberglitt, & Buckley, 2012). After this procedure, residual-based dimensionality analysis was carried out for each of four scales separately (Smith, 2002). According to this model checking framework, the appearance of the component with a rather large eigenvalue (>2) indicates a remarkable additional relationship among items. Besides, this part of the analysis allowed to indicate poor functioning indicators based on fit statistics (INFIT MNSQ; Linacre, 2002).

Results

According to the results of preliminary analysis, five item-bundles were created. They united from two to six indicators. Consequent analysis of each of the scales indicated that most of the items fit the models fairly well (INFIT MNSQ varied from 0.8 to 1.2). For the Safe scale 4 out of 6 indicators demonstrated good functioning; for the Work scale-out 13 of 19 indicators demonstrated good functioning; for the Create scale, 6 out of 11 indicators functioned fine. For the Ethic scale, all 4 indicators functioned fairly well. The results of dimensionality analysis indicated each scale might be considered as substantially unidimensional. One important result

of the preliminary analysis is that indicators from the Interfacing component did not form one holistic factor.

The results of internal structure analysis showed the model with four correlated factors (model 1) demonstrated a better model fit than the testlet model (model 2) (AICΔ = 102.93; BICΔ = 79.03). The results indicate that DL should not be interpreted as a holistic construct but only as a general name of several related skills.

The reliability coefficients of model 1 scales were rather small (0.5 for the Ethics scale; 0.56 for the Creation scale; 0.7 for the Security scale; 0.76 for the Access scale). The highest correlation coefficient was between the Security and Manipulation scales (0.82); the lowest correlation coefficient was between Ethics and Creation scales (0.44).

The results demonstrate that the instrument has satisfactory psychometric properties, however, some improvements are needed. Therefore, a qualitative study in the form of a semi-structured interview will be held. The purposes of the qualitative study are to check the content validity of the scenarios and to clarify the reasons for Interfacing factor absence and unexpectedly high correlation between Security and Access scales.

In summary, as the result of the analysis, the structural validity was partially confirmed: we have obtained four correlated factors describing DL, however, there is no general factor of digital literacy. The content validity was investigated through qualitative study and several important recommendations were achieved.

1. Abbot S. (2014) Hidden Curriculum / S. Abbott (ed.) The Glossary of Education Reform.
2. Adams, R. J., Wilson, M., & Wang, W. C. (1997). *The multidimensional random coefficients multinomial logit model. Applied Psychological Measurement, 21(1), 1–23. https://doi.org/10.1177/0146621697211001*
3. Baghaei, P. (2008). *Local dependency and Rasch measures. Rasch Meas Trans, 21(3), 1105–6.*
4. Burnham, K. P., & Anderson, D. R. (2004). *Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research, 33(2), 261–304. https://doi.org/10.1177/0049124104268644*
5. Chetty, K., Qigui, L., Gcora, N., Josie, J., Wenwei, L., & Fang, C. (2018). *Bridging the digital divide: measuring digital literacy. Economics: The Open-Access, Open-Assessment E-Journal, 12(2018-23), 1-20.*
6. Digital Economy Outlook//OECD. (2017).
7. Fraillon, J., Schulz, W., &Ainley, J. (2013) *International Computer and Information Literacy Study: Assessment framework. Amsterdam: IEA.*
8. G20 Digital Economy Ministerial Conference, Dusseldorf. (2017).
9. Hambleton, R. K., Swaminathan, H., & Rogers, H. J. (1991). *Fundamentals of item response theory. Newbury Park, Calif: Sage Publications.*
10. Katz, I. R. (2007). *Testing information literacy in digital environments: ETS's iSkills assessment. Information technology and Libraries, 26(3), 3-12.*
11. Levy, R. (2013). *Psychometric and Evidentiary Advances, Opportunities, and Challenges for Simulation-Based Assessment. Educational Assessment, 18(3), 182–207.* https://doi.org/10.1080/10627197.2013.814517
12. Linacre, J. M. (2002). *What do infit and outfit, mean-square and standardized mean? Rasch Measurement Transactions, 16 (2), 878.*
13. Mislevy, R. J., Behrens, J. T., Dicerbo, K. E., & Levy, R. (2012). *Design and discovery in educational assessment: Evidence-centered design, psychometrics, and educational data mining. JEDM| Journal of Educational Data Mining, 4(1), 11-48.*

14. Quellmalz, E. S., Timms, M. J., Silberglitt, M. D., & Buckley, B. C. (2012). *Science assessments for all: Integrating science simulations into balanced state science assessment systems. Journal of Research in Science Teaching, 49(3), 363–393. https://doi.org/10.1002/tea.21005*

15. Rosenbaum, P. R. (1988). *Items bundles. Psychometrika, 53(3), 349–359. https://doi.org/10.1007/BF02294217*

16. Smith, E. V. (2002). *Detecting and evaluating the impact of multidimensionality using item fit statistics and principal component analysis of residuals. Journal of Applied Measurement, 3(2), 205–231.*

17. Wang, W.-C., & Wilson, M. (2005). *The Rasch Testlet Model. Applied Psychological Measurement, 29(2), 126–149. https://doi.org/10.1177/0146621604271053*

18. World Economic Forum and the Boston Consulting Group, "Redefining the Future of Growth: The New Sustainability Champions, Research Report, 2011" . (2011).