Measuring Digital Self-Efficacy in International Large-Scale Assessments: An International Comparison Between ICILS and PISA

Paper prepared for the Invalsi 2025 Conference

Juan Carlos Castillo Daniel Miranda Tomás Urzúa
Nicolás Tobar Ismael Aguayo

2025-05-30

Introduction

Digital self-efficacy (hereinafter DSE), defined as expectations about one's capabilities to learn and accomplish tasks in digital technologies and digital environments, is one of the principal components to promote the formation of digital competences (Ulfert-Blank and Schmidt 2022). DSE is a construct frequently measured in international large-scale assessments (ILSAs), as substantial evidence indicates its critical role as an explanatory variable in the development of digital competences within educational settings (Scherer and Siddiq 2019; **hatlevik_students_2018?**; **claro_assessment_2018?**). Furthermore, studies consistently demonstrate that DSE also allows individuals to acquire and apply digital skills effectively (Rohatgi, Scherer, and Hatlevik 2016; **siddiq_teachers_2017?**).

The conceptualization and operationalization of DSE vary notably in terms of concepts and their measurement. Some studies treat DSE as a unidimensional construct, measuring individuals' overall confidence in using digital technologies without distinguishing between types of tools and/or levels of complexity (**hatlevik_digital_2015?**; Rohatgi, Scherer, and Hatlevik 2016). Such unidimensional approach facilitates modeling and broader comparisons but may obscure important differences in how users perceive their abilities in specific digital contexts. In contrast, other studies adopt a multidimensional approach, mostly distinguishing between general and specialized self-efficacy to account for the nature and complexity of digital tasks (Scherer, Siddiq, and Teo 2015). Whereas *general DSE* encompasses confidence in everyday tasks such as internet navigation or word processing, *specialized DSE* involves more advanced activities such as programming and/or data analysis.

Between the two most relevant ILSAs in the Digital Competence agenda (ICILS and PISA), a critical inconsistency persists in their conceptualization and measurement of DSE. PISA operationalizes DSE as a unidimensional construct, aggregating all digital task-related confidence into a single generalized measure (OECD 2021). In contrast, ICILS adopts a bidimensional framework, distinguishing between general DSE (basic digital tasks) and specialized DSE (advanced tasks) (**fraillon_preparing_2020?**; Scherer and Siddiq 2019). This discrepancy raises essential questions about construct validity and cross-assessment comparability, particularly since the choice of model (unidimensional vs. multidimensional) may influence policy interpretations and pedagogical interventions. For instance, unidimensional models could underestimate the predictive power of DSE for complex digital problem-solving, while multidimensional models offer greater explanatory precision but can introduce challenges such as construct overlap or reduced generalizability across contexts, limiting findings across educational systems and cultural contexts (Scherer and Siddiq 2019; **scherer_measuring_2021?**). Therefore, understanding the proper use of the dimensions of DSE is necessary to refine the scientific use of this construct to understand different populations' expectations with technologies. Actually, the bidimensional differentiation of DSE emerged, in part, from observed gender disparities in self-efficacy patterns: some studies show that while gender gaps in general DSE are minimal or non-existent, women tend to report significantly lower confidence in specialized DSE domains—particularly those involving STEM-related digital tasks (**hargittai_differences_2006?**; Cai, Fan, and Du 2017; OECD 2021).

Aiming to contribute to this research area, the present study's objective is to evaluate the measurement of a two-dimensional model of DSE and its comparability across countries and by gender in different large-scale assessments. Our contribution is twofold: (i) To assess the bi-dimensional approach to DSE (as in ICILS) to PISA data (which assumes unidimensionality), and (ii) To evaluate the comparability of the bi-dimensional measurement of DSE across countries and gender in ICILS and PISA. To achieve this, we will conduct confirmatory factor analysis (CFA) and measurement invariance testing using data from the latest cycles of ICILS (2023) and PISA (2022). This approach will allow us to rigorously evaluate the validity of the two-dimensional model of DSE and its cross-cultural applicability.CFA is particularly well-suited for testing theoretical models where specific latent structures are hypothesized a priori—such as the proposed distinction between general and specialized dimensions of digital self-efficacy. This approach allows for the rigorous evaluation of model fit and the validation of factor structures based on observed indicators from large-scale assessments.

Research Questions and Hypotheses

The present study aims to address the following research questions:

1. Is it possible to identify two latent dimensions of digital self-efficacy (general and specialized) based on related batteries and indicators included in large-scale assessments such as PISA and ICILS?

2. Is the bi-dimensional measurement model of digital self-efficacy equivalent between girls and boys?
3. Is the bi-dimensional measurement model of digital self-efficacy equivalent across countries?

To answer these questions, we will test the following hypotheses:

H1: It is possible to identify two latent dimensions of digital self-efficacy (general and specialized) based on related batteries and indicators included in large-scale assessments such as PISA and ICILS (bi-dimensional hypothesis)

Furthermore, testing for measurement invariance across gender and countries is essential to ensure that the latent constructs are interpreted in a comparable manner across groups (**leitgoeb_measurement_2023?**; **meuleman_why_2023?**). Without such invariance, any observed differences in self-efficacy levels may reflect measurement artifacts rather than substantive differences. From this perspective, the second and third hypotheses are:

H2: The bi-dimensional measurement model of digital self-efficacy is equivalent between girls and boys.
H3: The bi-dimensional measurement model of digital self-efficacy is equivalent across countries.

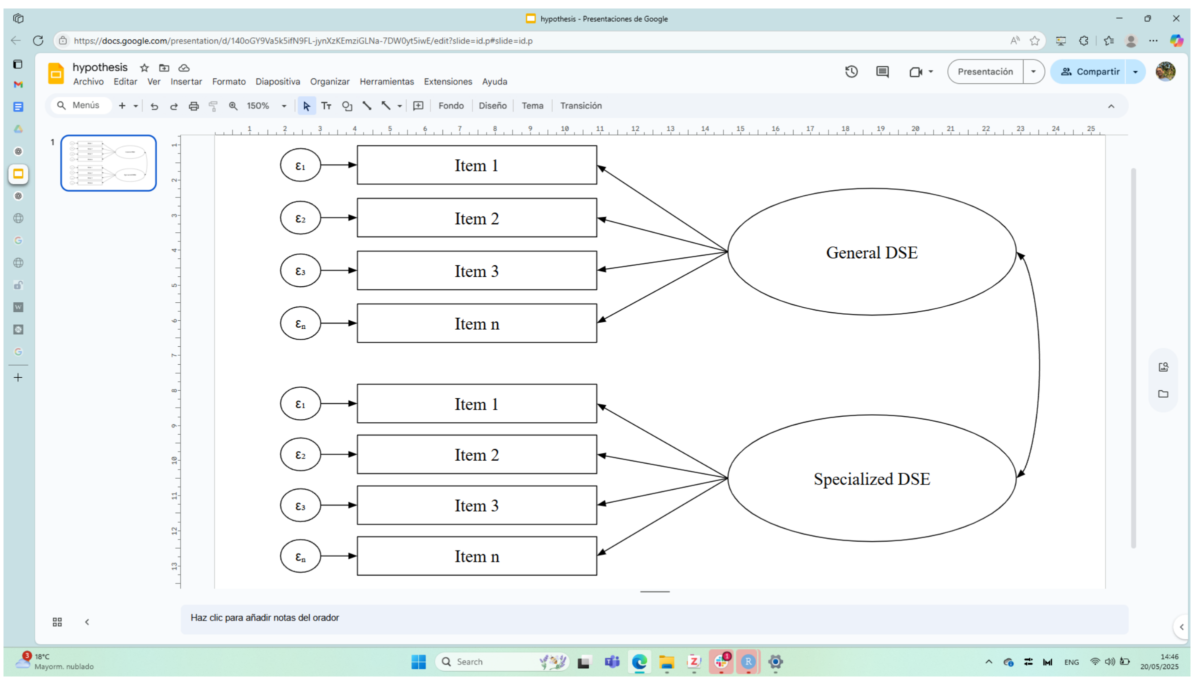

Figure 1: Confirmatory measurement model

Methods

Data

We have two main data sources. The first one is ICILS, developed by the International Association for the Evaluation of Educational Achievement (IEA). We use data from the third wave (2023), which encompasses 35 educational systems, testing 67,682 8th grade students on computer and information literacy (CIL) and computational thinking. The study evaluates students' ability to use digital tools responsibly, solve problems, and collaborate online. Data is collected through performance tests and contextual questionnaires for students, teachers, and schools. A key feature of ICILS is its bidimensional measurement of digital self-efficacy (DSE), distinguishing between general and specialized digital confidence.

The second data source is PISA, organized by the OECD, has assessed 15-year-olds' skills in mathematics, science, and reading across multiple cycles (the last three ones 2015, 2018, 2022). The study's primary objective remains evaluating education systems' effectiveness in preparing students for future challenges, with a growing emphasis on digital readiness. The 2022 assessment covered 81 countries/economies with a sample exceeding 600,000 students. Digital self-efficacy (DSE) was last measured in 2022 as part of the optional ICT familiarity questionnaire, following its absence in the 2018 cycle. This questionnaire was applied in an optional way in 53 countries, which are included in the analysis (N = 279,435).

Both data sources are publicly available and can be accessed through the IEA and OECD websites, respectively. ICILS 2023 data can be found at https://www.iea.nl/studies/iea/icils/2023, whereas PISA 2022 data is available at https://www.oecd.org/pisa/data/pisa-2022-database/.

Variables

The analysis will focus on the digital self-efficacy (DSE) items from both ICILS and PISA. The DSE items in ICILS 2023 are designed to measure students' confidence in performing various digital tasks, while PISA 2022 includes similar items but framed within a unidimensional context.

Table 1 summarizes the measurement batteries for self-efficacy in both studies. These items will be treated as numerical values.

Table 1: ICLS and PISA items comparison

Task Category	ICILS 2023 Item	PISA 2022 Item
Search information	Search for and find relevant info for a school project	Search for and find relevant information online

Task Category	ICILS 2023 Item	PISA 2022 Item
Assess information	Judge whether you can trust information you find	Assess the quality of information you found online
Create multimedia	Create a multi-media presentation	Create a multimedia presentation
Edit documents	Insert an image / edit text for assignment	Write or edit text for a school assignment
Edit images	Edit digital photographs	—
Upload/share content	Upload or share content	Share practical information / explain sharing
Collaborate	Collaborate on group assignment	Collaborate with students
Change settings	Change device settings	Change settings to protect data/privacy
Install/select apps	Install programs	Select most efficient app
Programming	Write program in code	Create visual/text-based program
Build a webpage	Build/edit webpage	Create/maintain webpage or blog
Identify software errors	Identify software error	Partial match (same intent)

Estimation methods

Statistical analyses will address missing data using Full Information Maximum Likelihood (FIML) during confirmatory analyses. Before modeling, PISA responses of "I don't know what this is" will be recoded as missing, and the distribution of missing values will be examined by country. For each country, a two-factor confirmatory factor analysis (CFA) will be conducted separately for PISA and ICILS data, distinguishing between general digital self-efficacy (DSE) for basic digital tasks and specialized DSE for advanced digital tasks. Model fit will be evaluated using chi-square, Comparative Fit Index (CFI), Tucker-Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA) (Brown 2015).

Measurement invariance across gender and countries will be tested using multi-group CFA, progressing through configural, metric, and scalar invariance. Changes in model fit will be interpreted using established thresholds (e.g., ΔCFI -0.004; ΔRMSEA .05 for metric, and ΔCFI -0.004; ΔRMSEA .01 for scalar invariance) (Rutkowski and Svetina 2017). All items will be treated as ordered variables, and FIML will be used for missing data throughout (Enders and Bandalos 2001). Analyses will be conducted in R and MPlus, with scripts and data available at https://github.com/milenio-nudos/ILSAs_batteries_measurement.