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

Invalid Date

# 1. Introduction

Digital self-efficacy (hereinafter DSE), conceived 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 et al. 2018; Claro et al. 2012). Furthermore, studies consistently demonstrate that DSE also allows individuals to acquire and apply digital skills effectively along the life spam (Rohatgi, Scherer, and Hatlevik 2016; Scherer, Siddiq, and Teo 2015).

The conceptualization and operationalization of DSE vary notably in terms of concepts and their measurement. Some studies conceive DSE as an unidimensional construct, measuring individuals’ overall confidence in using digital technologies without distinguishing between types of tools and/or levels of complexity (Hatlevik, Guðmundsdóttir, and Loi 2015; Rohatgi, Scherer, and Hatlevik 2016). In contrast, other studies adopt a multidimensional approach, 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 datamostly analysis. The distinction between general and specialized DSE is particularly relevant in the context of ILSAs, where the measurement of digital self-efficacy can significantly influence the interpretation of students’ digital competencies. For instance, a unidimensional measure may aggregate all digital task-related confidence into a single score, potentially masking important differences in how students perceive their abilities across different digital contexts. In contrast, a bidimensional measure allows for a more precise understanding of students’ digital competencies and can inform educational policies and practices more effectively.

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 an 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 et al. 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, as it is shown in studies that specialized DSE is a stronger predictor of performance in advanced digital tasks than general DSE (Scherer and Siddiq 2019; **scherer\_measuring\_2021?**). Besides, 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 and Shafer 2006; Cai, Fan, and Du 2017; OECD 2021). Therefore, understanding the proper use of the dimensions of DSE in two of the most important ILSAs including DSE measurement is necessary to refine the scientific use of this construct to understand different populations’ expectations with technologies.

The present study’s aim is (i) To assess the bi-dimensional approach of DSE in International Large-scale Assessment studies, 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. Besides, measurement invariance testing is crucial for ensuring that the same construct is being measured across different groups and contexts, which is essential for making valid comparisons of DSE across countries.

# 2. Research Questions and Hypotheses

The present study aims to address the following research questions:

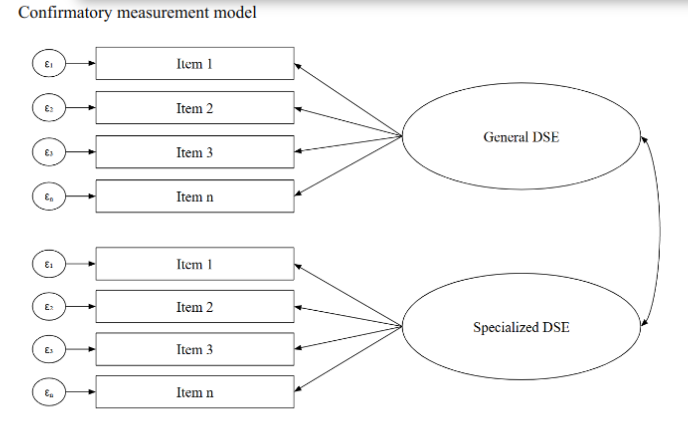
1. Is it possible to identify two latent dimensions of digital self-efficacy (general and specialized) based on related batteries and indicators, in a comparable way between PISA and ICILS?
2. Is the bi-dimensional measurement model of digital self-efficacy equivalent between girls and boys?
3. Is the 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)

**H2:** The bi-dimensional measurement model of digital self-efficacy is invariant (in terms of measurement) girls and boys.

**H3:** The bi-dimensional measurement model of digital self-efficacy is invariant across countries.



Confirmatory measurement model

# 3. Methods

## 3.1 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, problem solving, 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.

The second data source is PISA, implemented by the OECD. It assesses 15-year-olds’ skills in mathematics, science, and reading across multiple cycles (the last three ones 2015, 2018, 2022). The study’s primary objective is to evaluate 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/>.

## 3.2 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 |
| 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) |

## 3.3 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).

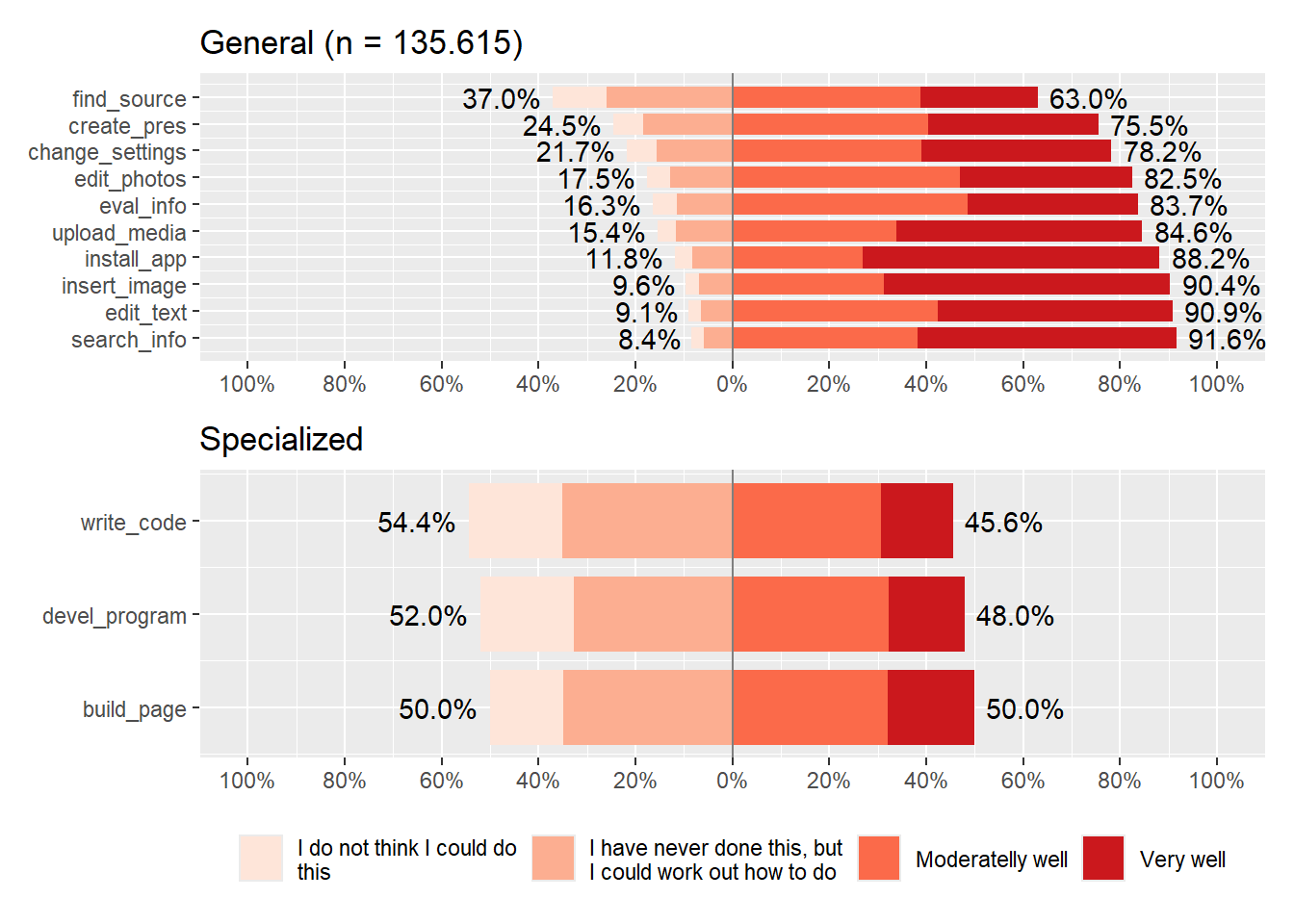
Measurement invariance across gender and countries will be tested using multi-group CFA, progressing through configural, metric, and scalar invariance. Changes in model fit will be interpreted using established thresholds (e.g., ΔCFI ≥ -0.004; ΔRMSEA ≤ .05 for metric, and ΔCFI ≥ -0.004; ΔRMSEA ≤ .01 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, the scripts and data are available at <https://github.com/milenio-nudos/ILSAs_batteries_measurement>.

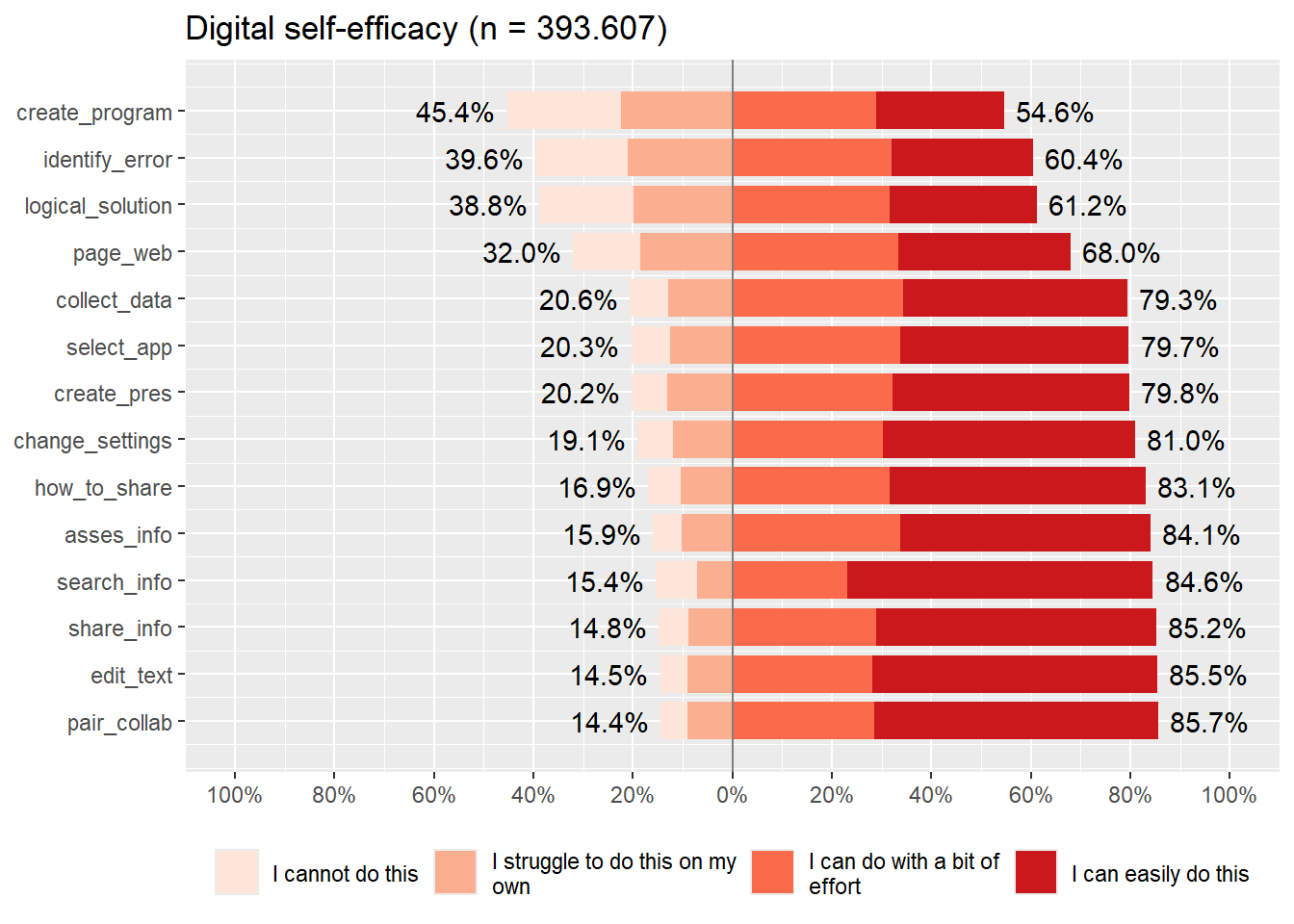
# 4. Preliminary analyses and results

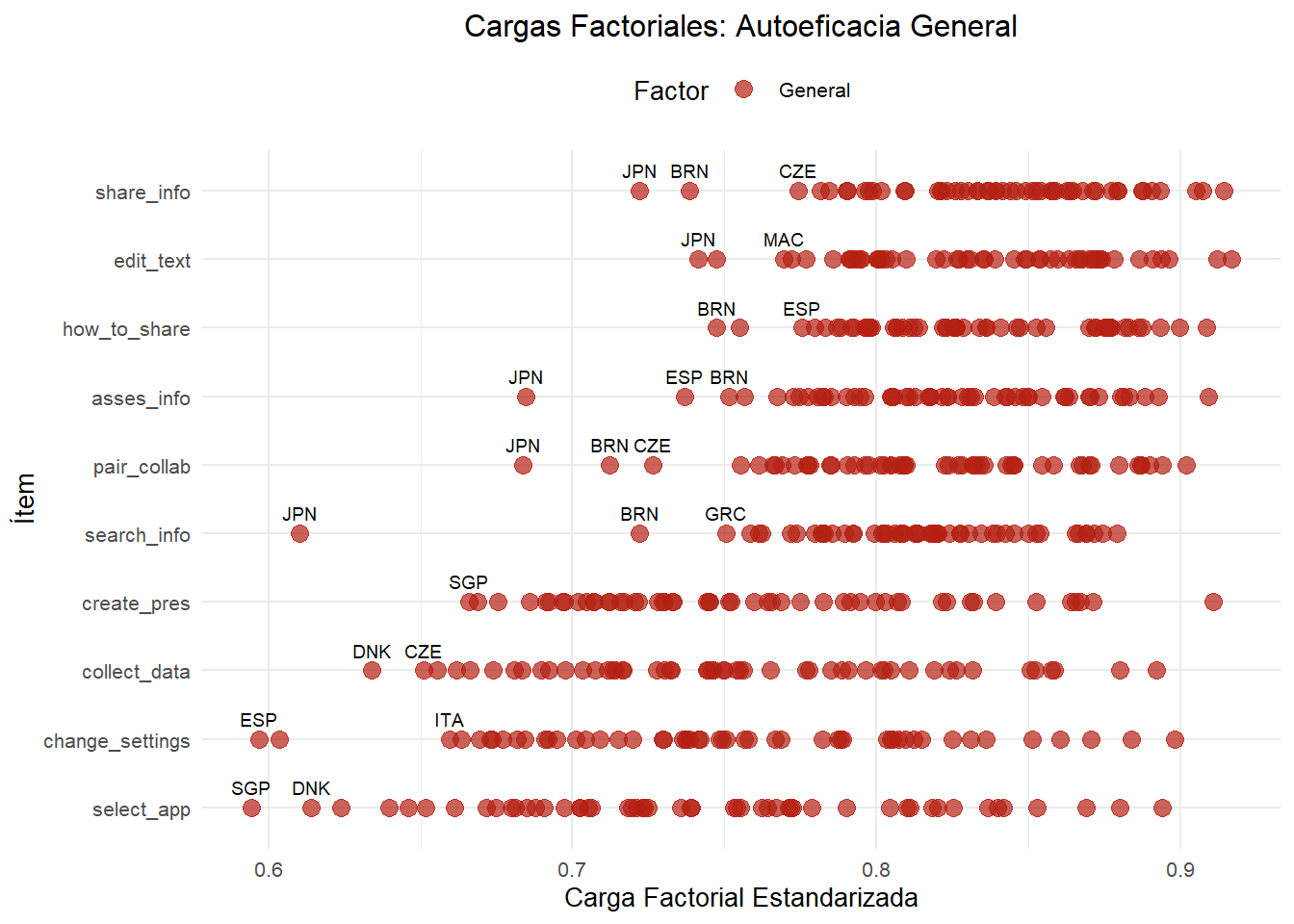
For ICILS 2023, a confirmatory factor analysis (CFA) estimation confirms the bi-dimensional model, with fit indices indicating good model fit (CFI = 0.97, TLI = 0.96, RMSEA = 0.045). In the case of PISA, the CFA results for the two-factor model show acceptable fit (CFI = 0.95, TLI = 0.94, RMSEA = 0.052), supporting the applicability of the bi-dimensional DSE model in both assessments, though with stronger evidence in ICILS.

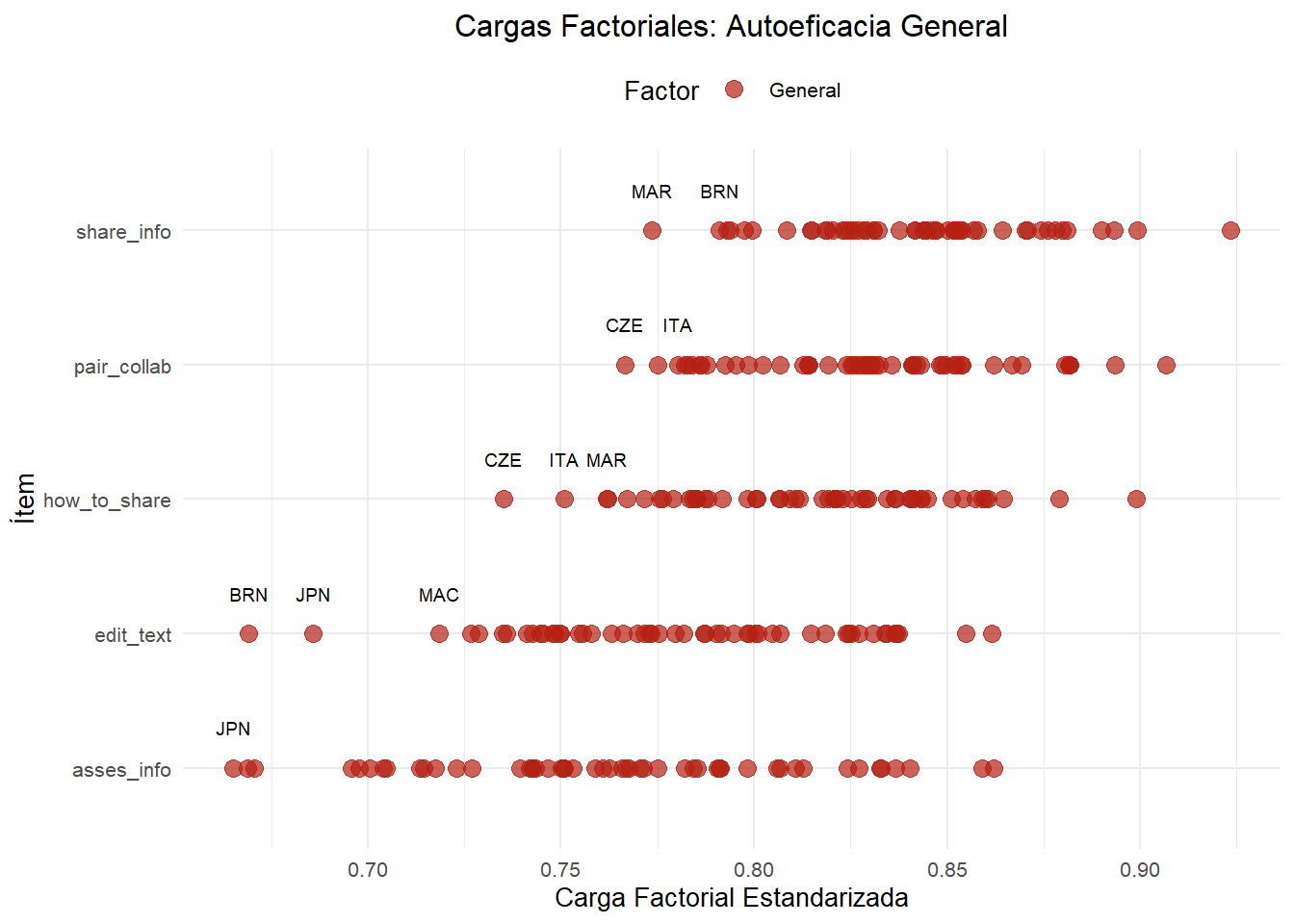
Regarding measurement invariance, the analyses indicate that for gender, both ICILS and PISA achieve configural and metric invariance, suggesting that the bi-dimensional model of digital self-efficacy is structurally similar and factor loadings are equivalent across girls and boys. However, scalar invariance is only partially supported, as some items exhibit differential item functioning by gender. In terms of country-level invariance, configural invariance is supported in both datasets, and metric invariance is achieved in most countries. Nonetheless, scalar invariance is more limited, particularly in PISA, indicating that some item intercepts vary across countries and may affect the comparability of mean scores.

Analysis of gender differences reveals that general digital self-efficacy (DSE) shows minimal disparities between girls and boys in both ICILS and PISA, while specialized DSE consistently favors boys, with effect sizes (Cohen’s d) ranging from 0.20 to 0.35 across countries. Cross-country comparisons indicate that Nordic and East Asian countries tend to have higher average DSE scores in ICILS, and similar patterns are observed in PISA, although specialized DSE exhibits greater variability across countries.









# 5. Summary

* The bi-dimensional model of DSE is empirically supported in both ICILS and PISA, though with stronger evidence in ICILS.
* Measurement invariance is generally supported at the configural and metric levels, but scalar invariance is more challenging, especially across countries.
* Gender gaps are more pronounced in specialized DSE.

Full tables, code, and additional figures are available in the <Avances.html> file and the project repository.

Brown, Timothy A. 2015. *Confirmatory Factor Analysis for Applied Research*. 2nd ed. New York: The Guilford Press.

Cai, Zhenzhen, Xitao Fan, and Jianxia Du. 2017. “Gender and Attitudes Toward Technology Use: A Meta-Analysis.” *Computers & Education* 105: 1–13. <https://doi.org/10.1016/j.compedu.2016.11.003>.

Claro, Magdalena, David D. Preiss, Ernesto San Martín, Ignacio Jara, J. Enrique Hinostroza, Sebastián Valenzuela, Francisco Cortes, and Miguel Nussbaum. 2012. “Assessment of 21st-Century ICT Skills in Chile: Test Design and Results from High School Level Students.” *Computers & Education* 59 (3): 1042–53. <https://doi.org/10.1016/j.compedu.2012.04.004>.

Enders, Craig K., and Deborah L. Bandalos. 2001. “The Relative Performance of Full Information Maximum Likelihood Estimation for Missing Data in Structural Equation Models.” *Structural Equation Modeling* 8 (3): 430–57. <https://doi.org/10.1207/S15328007SEM0803_5>.

Fraillon, Julian, John Ainley, Wolfram Schulz, Tim Friedman, and Daniel Duckworth. 2020. *Preparing for Life in a Digital World: IEA International Computer and Information Literacy Study 2018 International Report*. Springer Nature. <https://doi.org/10.1007/978-3-030-38781-5>.

Hargittai, Eszter, and Steven Shafer. 2006. “Differences in Actual and Perceived Online Skills: The Role of Gender.” *Social Science Quarterly* 87 (2): 432–48. <https://doi.org/10.1111/j.1540-6237.2006.00389.x>.

Hatlevik, Ove E., Greta B. Guðmundsdóttir, and Massimo Loi. 2015. “Digital Diversity Among Upper Secondary Students: A Multilevel Analysis of the Relationship Between Cultural Capital, Self-Efficacy, Strategic Use of Information and Digital Competence.” *Computers & Education* 81: 345–53. <https://doi.org/10.1016/j.compedu.2014.10.019>.

Hatlevik, Ove E., Inger Throndsen, Massimo Loi, and Greta B. Gudmundsdottir. 2018. “Students’ ICT Self-Efficacy and Computer and Information Literacy: Determinants and Relationships.” *Computers & Education* 118: 107–19. <https://doi.org/10.1016/j.compedu.2017.11.011>.

OECD. 2021. “PISA 2022 Assessment and Analytical Framework.” OECD Publishing. <https://doi.org/10.1787/19963777>.

Rohatgi, Anubha, Ronny Scherer, and Ove E. Hatlevik. 2016. “The Role of ICT Self-Efficacy for Students’ ICT Use and Their Achievement in a Computer and Information Literacy Test.” *Computers & Education* 102: 103–16. <https://doi.org/10.1016/j.compedu.2016.08.001>.

Rutkowski, Leslie, and Dubravka Svetina. 2017. “Multidimensional Measurement Invariance in an International Context: Fit Measure Performance with Many Groups.” *Journal of Cross-Cultural Psychology* 48 (7): 991–1008. <https://doi.org/10.1177/0022022117717028>.

Scherer, Ronny, and Fazilat Siddiq. 2019. “The Relation Between Students’ Socioeconomic Status and ICT Literacy: Findings from a Meta-Analysis.” *Computers & Education* 138: 13–32. <https://doi.org/10.1016/j.compedu.2019.04.011>.

Scherer, Ronny, Fazilat Siddiq, and Timothy Teo. 2015. “Becoming More Specific: Measuring and Modeling Teachers’ Perceived Usefulness of ICT in Teaching and Learning.” *Computers & Education* 88: 202–14. <https://doi.org/10.1016/j.compedu.2015.05.005>.

Ulfert-Blank, Ann-Sophie, and Ingo Schmidt. 2022. “Digital Self-Efficacy: A Systematic Literature Review.” *International Journal of Educational Research* 114: 102008. <https://doi.org/10.1016/j.ijer.2022.102008>.