

GATHERING CONTENT VALIDITY EVIDENCE FOR A MATHEMATICS AND STATISTICS DIAGNOSTIC TOOL USING QUALITATIVE METHODOLOGY

Carlos Fresneda-Portillo¹, Juan F. Luesia², & Carlos Benjamín Córdova-Portillo³

¹Department of Mathematics and Computing, University of Burgos (Spain)

²Department of Psychology, Universidad Loyola Andalucía (Spain)

³Department of Quantitative Methods, Universidad Loyola Andalucía (Spain)

Abstract

Competence in mathematics represents an ongoing challenge for higher education institutions, particularly as universities face evolving educational policies that espouse principles of competence-based education. This means it is crucial to ensure incoming students have adequate mathematical and statistical foundations for academic success across a wide range of disciplines. Unfortunately, very few standardized diagnostic instruments currently exist to measure the mathematical readiness of entering students or to identify learning gaps that could lead to failure at an early stage of their university trajectory. The present study was developed within the framework of the DiagnosticaMATH Innovation Project, whose main aim is to create a standardized diagnostic instrument to measure the mathematical and statistical cross-cutting competence of Spanish students at the moment of entering university, and to evaluate their evolution during the years of study. The present work describes the first phase of this project, which consists of obtaining validity evidence based on test content through qualitative methodology. To this end, three discussion groups were formed with experts in mathematics education and evaluation, until saturation of the proposed themes was reached. The analysis leads to three principal results: a) a single test can hardly evaluate specific mathematical skills, but a diagnostic task could be constructed through problem solving situations involving metacognitive processes; b) comprehension is a key transversal competence, which critically affects mathematical performance; c) while the pre-university education of the students offers a sufficient base of knowledge, it is generally procedural and superficial, depending little on flexible capacity for problem solving.

Keywords: *Mathematics diagnostic test, competencies, pre-university mathematics, standard test.*

1. Introduction

The development of mathematical reasoning constitutes an essential competence for effective learning in mathematics and for constructing critical knowledge among higher education students (Büchele & Feudel, 2023). The ability to reason, validate results, formulate conjectures, and develop a deep understanding of concepts is crucial for addressing complex problems and advancing in technical and scientific fields of study (Zhao, 2020).

However, a growing mismatch has been observed between the mathematical skills students bring to European universities and the level of knowledge those institutions expect. This trend is documented in recent reports from the Programme for International Student Assessment (PISA; European Commission, 2024), which show that the proportion of 15-year-old students failing to reach adequate mathematical proficiency (i.e., interpret ratios or solve linear equations) is approximately twice the expected level.

This deficit persists into first-year calculus and statistics courses (European Commission, 2025). Moreover, the gap is not fully remedied upon entry to higher education, leading universities to report increasing enrollment in remedial mathematics courses within certain STEM programs, in some cases nearly doubling since 2010 (Büchele & Feudel, 2023). In the Spanish context, a similar pattern is evident: secondary-school mathematics does not consistently prepare students for the analytical demands of university-level programs, and as a result, more than 30% of incoming undergraduates require tailored academic support in foundational topics (Ayuga-Téllez et al., 2023). This competency gap reveals a fracture in the educational transition, where assessment and access mechanisms fail to both capture and consequently foster the reasoning skills that university studies demand.

Specifically, curricula and teaching practices frequently prioritize content memorization, procedural application, and problem-solving, while argumentation, proof, and critical reasoning receive less attention (Bhattacharya, 2022). As a result, domain-specific knowledge (Stemler, 2012) is overemphasized, crowding out the cross-cutting competencies that matter for learning mathematics (López-Agudo et al., 2024). This gap between stated educational goals and actual classroom practices results in students entering university with insufficient competencies to cope with higher-level mathematics, where learning requires a formal, structured approach grounded in logical reasoning (Deeken et al., 2020; Faúndez et al., 2017; Hed et al., 2025; Tenorio et al., 2015; Zhao, 2020).

The lack of a psychometrically valid diagnostic test of pre-university mathematics, as well as the lack of a rigorous theoretical model of this construct, leaves HE institutions without reliable and comparable information about the actual mathematical competencies of their first-year students. Consequently, any diagnostic effort tends to be fragmented and non-standardized, preventing a systematic assessment of the problem at the program, faculty or university level, and thus hindering the justification of resources for providing the appropriate support to students.

For this reason, it is essential to develop diagnostic instruments that help (1) identify the potential mathematical competency needs of students entering university studies, and (2) promote effective learning in mathematics throughout the university cycle. This development requires obtaining validity evidence to justify the intended use of test scores (AERA et al., 2014). A first, fundamental, and frequently overlooked step is to gather validity evidence based on test content, which involves examining whether the test adequately represents the intended domain or construct through expert judgment (Sireci & Faulkner-Bond, 2014). Concretely, focus groups are a technique for exploring complex concepts, building consensus among experts, and clearly defining evaluation dimensions, competency indicators, and appropriate item types (Molano et al., 2021).

2. Method

The main objective of this study is to identify the key elements to include in a diagnostic tool measuring the mathematical readiness of entering students from the lecturers' perspective. Fulfilling this objective is the first step towards building a diagnostic test of pre-university mathematics. To attain this target, we used a qualitative focus group methodology. Participants for the focus groups were selected following two main criteria. First, a quota system was implemented to ensure each group included professionals with varying levels of experience, from junior to senior practitioners. Second, all participants were required to be currently engaged in mathematics teaching at the secondary or higher education level and to demonstrate recognized expertise in their field. In total, 21 experts took part, distributed into groups of four to ten members to promote productive discussion and interaction (Ritchie et al., 2017). Regarding gender distribution, the sample comprised 17 men and 3 women. Regarding professional experience, 13 participants were classified as junior professionals and 8 as senior practitioners. Participation was voluntary, and no incentives were provided.

An interview protocol was designed to explore participants' perspectives on the current state of mathematical competence among first-year university students. Specifically, the protocol aimed to identify emerging competencies and indicators to help diagnose potential learning needs across students' academic trajectories in higher education. Ethical approval for the study was granted by the Ethics Committee of Universidad Loyola Andalucía (May 6th, 2025), Ref. No.250325-CE 25588.

Once the evaluation protocol was approved by the Ethics Committee, the focus group sessions were conducted in quiet classrooms and facilitated by two moderators who maintained a neutral and unbiased stance throughout the discussions. All sessions were audio-recorded for subsequent transcription. Recruitment of experts concluded after the fourth session, once thematic saturation was reached (Charmaz & Belgrave, 2002).

For data analysis, a qualitative framework for narrative transcription was applied, following the procedures outlined by Creswell and Poth (2016). The process comprised several sequential stages: (1) data management (transcription); (2) reading, coding of emerging tool elements (e.g., [PrInt] for Interpretation Problems); (3) describing, adding indicators, and narrative information to each element; (4) interpreting, developing an analytical template; and (5) representing, through inter-group agreement (IGA) for each indicator). All stages were monitored through peer review by three researchers to ensure methodological rigor and consistency.

3. Results

The experts' responses enabled the identification of 12 emerging tool elements. Detailed information regarding these two, including their definitions, codes, and illustrative narratives, is presented in Table 1.

Table 1. Emerging Tool Elements Identified in the Focus Groups.

Tool Element	Illustrative Narrative	IGA
Interpretation Problems: connect various topics, globally involve students (PrInt)	Conceptual-level questions. That unit vector...what is it doing? No need to calculate it, but at least understand because there's a vector, you know where it's going, what it means. They should know the concept of vector, the concept of a function that symbolizes this function.	1
Interpret the results (Interpret)	They must understand the concept of a vector and the concept of a function, including what that function represents.	1
Understanding of statements (CompLec)	They should have understanding of the language. Know what they're being asked, what data they have, and which data are relevant.	1
Infer, reason, deduce (Inferir)	Typical questions such as those involving multiple similar diagrams where participants identify which diagram does not fit or what is missing, among similar tasks.	1
Mathematics (Mat)	Calculus, algebra, including operations with fractions, as well as very basic first-degree and second-degree equations, to check if they can solve them.	1
Direct Problems: Problem solving with explicit procedure (PrDir)	Present a series of data and have them apply the formula.	.75
Application Problems: what do I need to do, what should I do (PrApl)	A modeling exercise, involving extracting a mathematical model from a problem statement and selecting the most appropriate strategy, which could be graphical.	.75
Analysis: critical graphic-spatial interpretation (Graf)	Perhaps framed as data interpretation or graph interpretation, something slightly more abstract that reveals critical reasoning.	.75
Verify (Comprobar)	Examine the result and reason whether it is correct or not, because by simply observing how nature works, you can intuit whether it is accurate.	.75
Inverse Problems (PrInv)	You can also set it up the other way around. You can write on a sheet how the three integrals are done and then include the three integrals. Let's see if they're able to understand that mathematically, that you've written it there to solve it later.	.50
Transversality: relate and connect topics (RelCont)	If I change the rules there, the student must examine all the possible ways of combining them, even if they have never encountered them before.	.50
Written Expression (ExpEsc)	I would conduct an interview with them and have them write it out on the spot while drafting it.	.50

The tool elements are displayed in descending order of inter-group agreement (IGA), with perfect consensus (IGA = 1.00) on the top five: (1) Interpretation Problems (PrInt), (2) Interpret the results (Interpretar), (3) Understanding of statements (CompLex), (4) Infer/reason/deduce (Inferir), and (5) Mathematics (Mat).

4. Discussion

From the discussion across the different focus groups, it was evident that there was consensus on the current decline in comprehensive reading competencies. For mathematics lecturers, it is clear that exercises with a statement that involves reading more words are generally perceived as more complicated, regardless of the task's difficulty. These results are in line with the PISA report 2022 (European Commission, 2024).

Secondly, there was consensus that students do not have the appropriate level of competency in the foundations of algebra and arithmetic, as reflected in the learning outcomes from secondary education. Examples of these include: lack of fluency in manipulating and operating with fractions and indices, difficulty in solving linear and quadratic equations, and understanding the fundamental concepts of elementary functions, such as the equation of a line and its applications.

Regarding the syllabus of pre-university mathematics, lecturers considered that the content taught is sufficient, but the problem stems from the teaching and learning method, which seems reduced to a set of procedures and methods, rather than understanding the fundamental concepts of mathematics, which unable students to acquire more sophisticated skills such as abstraction, inference, logical reasoning, analysis, and critical thinking.

In terms of what to include in a standard diagnostic test, questions regarding the fundamentals of algebra and arithmetic are definitely one of the core parts of this diagnostic. Secondly, competency in matrix algebra and precalculus is sought; thus, questions on these topics are generally preferred but should be tailored to different levels to differentiate the level and range of skills applied (as mentioned in the focus groups). Lastly, their comprehensive reading and writing competencies should be tested within this diagnostic test.

Acknowledgments

The researchers acknowledge the funding received from the Teaching and Learning Innovation Grants of the Universidad Loyola Andalucía (Spain), Ref:DiagnosticMath, 2025-2026.

References

- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Ayuga-Téllez, E., González, C., & Martín-Fernández, S. (2023). Statistics in Forest Engineering Degree Programs: What Should be Taught and Who Should Teach It. *International Journal of Engineering Pedagogy (iJEP)*, 13(4), 37–49. <https://doi.org/10.3991/ijep.v13i4.37261>
- Bhattacharya, U. (2022). “I Am a Parrot”: Literacy Ideologies and Rote Learning. *Journal of Literacy Research*, 54(2), 113–136. <https://doi.org/10.1177/1086296x221098065>
- Büchle, S., & Feudel, F. (2023). Changes in students’ mathematical competencies at the beginning of higher education within the last decade at a German university. *International Journal of Science and Mathematics Education*, 21(9), 2325–2347. <https://doi.org/10.1007/s10763-022-10350-x>
- Charmaz, K., & Belgrave, L. (2002). Qualitative interviewing and grounded theory analysis. In J. F. Gubrium & J. A. Holstein (Eds.), *Handbook of interviewing research: Context & Method* (pp. 347-365). Sage Publications.
- Creswell, J. W., & Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches* (4th ed.). SAGE.
- Deeken, C., Neumann, I., & Heinze, A. (2020). Mathematical prerequisites for STEM programs: What do university instructors expect from new STEM undergraduates? *International Journal of Research in Undergraduate Mathematics Education*, 6(1), 23–41. <https://doi.org/10.1007/s40753-019-00098-1>
- European Commission, Directorate-General for Education, Youth, Sport and Culture. (2024). *The twin challenge of equity and excellence in basic skills in the EU: An EU comparative analysis of the PISA 2022 results*. Publications Office of the European Union. <https://data.europa.eu/doi/10.2766/881521>
- European Commission. (2025). *Education and training monitor 2023: Comparative report*. Chapter 2. <https://op.europa.eu/webpub/eac/education-and-training-monitor/en/comparative-report/chapter-2.html>
- Hed, L., Åhag, P., & Norqvist, M. (2025). Changes in mathematical skills among freshman engineering students. *International Journal of Mathematical Education in Science and Technology*, 56(11), 2236–2250. <https://doi.org/10.1080/0020739X.2024.2404427>
- López-Agudo, L. A., De Guevara Rodríguez, M. L., & Marcenaro-Gutierrez, O. D. (2024). The influence of grade retention on students’ competences in Spain. *European Journal of Education*, 59(4). <https://doi.org/10.1111/ejed.12736>
- Molano, M., Valencia, A., & Apraez, M. (2021). Características e importancia de la metodología cualitativa en la investigación científica. *Revista Semillas del Saber*, 1(1), 18–27. Retrieved January 15, 2026, from: <https://revistas.unicatolica.edu.co/revista/index.php/semillas/article/view/314/178>
- Ritchie, J., Lewis, J., Nicholls, C. M. & Ormston, R. (2017). *Qualitative research practice: A guide for social science students and researchers* (2nd ed.). Sage Publications.
- Sireci, S., & Faulkner-Bond, M. (2014). Validity evidence based on test content. *Psicothema*, 1(26), 100–107. <https://doi.org/10.7334/psicothema2013.256>

- Stemler, S. E. (2012). What should university admissions tests predict? *Educational Psychologist*, 47(1), 5–17. <https://doi.org/10.1080/00461520.2011.611444>
- Tenorio, A. F., Martín, A. M., & Bermudo, S. (2015). Alumnos de nuevo ingreso en ingeniería: Un análisis de competencias matemáticas básicas. *Épsilon – Revista de Educación Matemática*, 32(1), 79–89. Retrieved January 15, 2026, from: https://thales.cica.es/epsilon/sites/default/files/2023-04/epsilon89_7.pdf
- Zhao, W. (2020). Predicament and outlook of China's math education. *National Science Review*, 7(9), 1513–1517. Retrieved January 15, 2026, from: <https://doi.org/10.1093/nsr/nwaa070>