

# revista de **e**EDUCACIÓN

Nº 412 ABRIL-JUNIO 2026



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE EDUCACIÓN, FORMACIÓN PROFESIONAL  
Y DEPORTES

2025  
CALIDAD  
REVISTAS  
CIENTÍFICAS  
ESPAÑOLAS





**revista de**  
**e**EDUCACIÓN



**Nº 412 ABRIL-JUNIO 2026**

# revista de EDUCACIÓN

**Nº 412 Abril-Junio 2026**

**Revista trimestral**

**Fecha de inicio: 1952**

## **MINISTERIO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL**

SECRETARÍA DE ESTADO DE EDUCACIÓN Y FORMACIÓN PROFESIONAL

Instituto Nacional de Evaluación Educativa  
Paseo del Prado, 28, 4ª planta  
28014 Madrid  
España

Edita  
© SECRETARÍA GENERAL TÉCNICA  
Subdirección General de Atención al Ciudadano, Documentación y Publicaciones

Catálogo de publicaciones del Ministerio: [sede.educacion.gob.es](http://sede.educacion.gob.es)  
Catálogo general de publicaciones oficiales: [publicacionesoficiales.boe.es](http://publicacionesoficiales.boe.es)

Edition: 2026  
NIPO pdf: 164-24-065-X  
NIPO ibd: 164-24-064-4  
NIPO html: 164-24-066-5  
ISSN Online 3045-8595  
ISSN Print: 3045-8587  
Legal Deposit: M-26966-2023

Diseño de portada: Iris Bermejo Rodríguez  
Fotografía: Ron Lach (Pexels)  
Fuente: <https://www.pexels.com/es-es/foto/sentado-joven-nino-adolescente-9783350/>  
Maqueta: Ibersaf Industrial S.L.

## BOARD OF DIRECTORS

### CHAIR

Abelardo de la Rosa  
Secretario de Estado de Educación

Esther Monterrubio Ariznabarreta  
Secretaria General de Formación Profesional

### MEMBERS

Santiago Antonio Roura Gómez  
Subsecretario de Educación, Formación Profesional y Deportes

Mónica Domínguez García  
Directora General de Evaluación y Cooperación Territorial

Carmen Tovar Sánchez  
Directora del Instituto Nacional de Evaluación Educativa

Lucio Calleja Bachiller  
Subdirector General de Cooperación Territorial  
e Innovación Educativa

Helena Ramos García  
Subdirectora General de Ordenación Académica

## EDITORIAL BOARD

### DIRECTOR

Carmen Tovar Sánchez

### EDITOR IN CHIEF

José Luis Gaviria Soto

### ASSISTANT EDITOR

David Reyero García

### MEMBERS

Antonio Cabrales Goitia (University College London); Caterina Casalmigla (Universitat Autònoma de Barcelona); Antonio Lafuente García (Consejo Superior de Investigaciones Científicas); Leoncio López-Ocón Cabrera, (Consejo Superior de Investigaciones Científicas); Carlos Marcelo García (Universidad de Sevilla); Miquel Martínez Martín (Universitat de Barcelona); Francisco Michavila Pitarch; (Universidad Politécnica de Madrid); Juan Manuel Moreno Olmedilla (Banco Mundial); Clara Eugenia Núñez (Universidad Nacional de Educación a Distancia); Lucrecia Santibáñez (Claremont Graduate University); Denise Vaillant (Universidad ORT, Uruguay); Pablo Zoido (Banco Interamericano de Desarrollo).

## EDITORIAL TEAM

**Technical editor:** Julio López Sevilla

**Collaborator:** Ruth Martín Escanilla, Alberto Díaz-Ron-cero Canales, Jasmine Blasco Yúrevich y Marta Carrera González

## SCIENTIFIC ADVISERS

### International

Aaron Benavot (State University of New York, SUNY-Al-bany); Abdeljalil Akkari (Universidad de Ginebra); Mark Bray (University of Hong Kong); José Joaquín Brunner (Universidad Diego Portales, Chile); Dirk Hastedt (Executive Director; International Association for the Evaluation of Edu-cational Achievement, IEA); Andreas Schleicher (Director; Directorate for Education and Skills, OCDE)..

### National

Teresa Aguado Odina (Universidad Nacional de Educación a Distancia); Margarita Bartolomé (Universitat de Barcelona); Antonio Bolívar (Universidad de Granada); Josefina Cambra Giné (Colegio de Doctores y Licenciados); Anna Camps i Mundó (Universitat Autònoma de Barcelona); César Coll Salvador (Universitat de Barcelona); Agustín Dosil Maceira (Universidad LiberQuaré); Gerardo Echeita Sarrionandia (Universidad Autónoma de Madrid); Juan Manuel Escu-dero Muñoz (Universidad de Murcia); Mariano Fernández Enguita (Universidad Complutense de Madrid); Joaquín Gairín Sallán (Universitat Autònoma de Barcelona); José Luis García Garrido (Universidad Nacional de Educación a Distancia); Daniel Gil Pérez (Universitat de València); Fuensanta Hernández Pina (Universidad de Murcia); Carmen Labrador Herraiz (Universi-dad Complutense de Madrid); Miguel López Melero (Universidad de Málaga); Elena Martín Ortega (Universidad Autónoma de Madrid); Rosario Martínez Arias (Universidad Complutense de Madrid); Inés Miret (Neturity S.L., Madrid); Juan Ignacio Pozo (Universi-dad Autónoma de Madrid); Joaquim Prats Cuevas (Universi-tat de Barcelona); Manuel de Puelles (Universidad Nacional de Educación a Distancia); Tomás Recio Muñiz (Universidad de Cantabria); Luis Rico Romero (Universidad de Granada); Juana M.ª Sancho Gil (Universitat de Barcelona); Mercedes Vico Monteoliva (Universidad de Málaga); Antonio Viñao Frago (Universidad de Murcia); Jesús García Laborda (Uni-versidad de Alcalá).



## Research

FERNANDO GIL CANTERO: Education is not a transhumanist activity.....	3
LAURA ORTEGA TORRES Y CÉSAR ANTONIO CHÁVEZ ÁLVAREZ: Differential item functioning analysis in large-scale assessments .....	21
ELENA CASTRO-RODRÍGUEZ, JUAN LUIS PIÑEIRO Y ENRIQUE CASTRO: Mathematical problem-solving in primary education teacher training programmes in Spain .....	45
MARIO ALARCÓN BRAVO Y JOSÉ JOAQUÍN BRUNNER RIED: University autonomy: a multilevel analytical model from a Latin American perspective .....	71
MICAELA SÁNCHEZ-MARTÍN, MARTA GUTIÉRREZ-SÁNCHEZ, EVA MARÍA OLMEDO-MORENO Y FERNANDO NAVARRO-MATEU: Adherence to the PRISMA Statement in meta-Analyses reporting experimental interventions published in Education: A Systematic Meta-Review .....	95
MERY ESTEFANÍA BUESTÁN-JÁATIVA, ELENA ORTEGA-CAMPOS, LETICIA DE LA FUENTE-SÁNCHEZ Y JUAN GARCÍA-GARCÍA: Efficacy of Psycho-socio-educational Programs applied in Juvenile Justice in Spain: A Systematic Review .....	131
GARAZI ORMAZABAL-ARIZKORRETA, ANDONI ARGUIÑANO MADRAZO Y INAKI KARRERA XUARROS: Sowing the Seeds of a Lifelong Journey that Begins in Childhood: Sociocultural Learning in <i>Eskola Txikiak</i> .....	159
ANDREA BETTI, PABLO BIDERBOST Y ESTHER VAQUERO: Comparing Visual Thinking and Flipped Classroom in International Relations Teaching..	183
<b>Reviews</b> .....	213



# **RESEARCH**



# Education is not a transhumanist activity

## La educación no es una actividad transhumanista

<https://doi.org/10.4438/1988-592X-RE-2025-412-740>

**Fernando Gil Cantero**

<https://orcid.org/0000-0002-6636-7601>

*Universidad Complutense de Madrid*

### **Abstract**

**Introduction.** The contemporary debate on transhumanism has led to interpretations that connect it with education as a means of human enhancement. This article challenges that equivalence, arguing that education and transhumanism follow radically different types of logic and pursue fundamentally disparate ends. **Method.** Taking a philosophical and pedagogical approach, the article critically examines the premises of transhumanism from the perspectives of anthropology, ethics, and educational theory. It contrasts the transhumanist ideal of the functional enhancement of capacities with the teleological tradition of human formation. **Results.** The article argues that while transhumanism is aimed at the technical optimization of the individual—especially regarding cognitive, physical, or sensory abilities—education is geared toward the ethical maturation of the individual through human relationships, culture, and personal effort. Not every capacity enhancement constitutes an educational improvement: true improvement occurs only when autonomy is ethically oriented toward valuable ends, acknowledging human vulnerability and dependence. The article also rejects the reduction of ethics to biologicistic parameters and advocates a view of humans as moral agents guided by teleology. **Discussion.** In response to the cultural hegemony of efficiency and fascination with technology, the article defends the irreplaceable role of education as a humanizing process. Pedagogy must not be subordinated to the logic of performance or to intellectual trends, but must safeguard essential values such as effort, prudence, and the acceptance of limits. Within this framework, the article offers a critique of the technocratic illusion of unlimited perfectibility, emphasizing that the ultimate goal of education is not to program functional subjects, but to form individuals capable of acting in a meaningful and responsible manner in a shared world.

**Keywords:** transhumanism; posthumanism; pedagogy; human enhancement; human nature; education; autonomy.

## Resumen

**Introducción.** El debate contemporáneo sobre el transhumanismo ha generado interpretaciones que lo acercan a la educación como vía de mejora del ser humano. Este artículo cuestiona esa equivalencia, argumentando que educación y transhumanismo responden a lógicas y fines radicalmente distintos. **Método.** A través de un enfoque filosófico y pedagógico, se analizan críticamente las premisas del transhumanismo desde la antropología, la ética y la teoría educativa. El artículo contrapone el ideal transhumanista de ampliación funcional de capacidades con la tradición teleológica de la formación humana. **Resultados.** Se sostiene que, mientras el transhumanismo apuesta por la optimización técnica del individuo —especialmente en sus capacidades cognitivas, físicas o sensoriales—, la educación se orienta a la maduración ética del sujeto, mediante relaciones humanas, cultura y esfuerzo. No toda ampliación de capacidades constituye una mejora educativa: esta solo ocurre cuando la autonomía se orienta éticamente hacia fines valiosos, reconociendo la vulnerabilidad y dependencia humanas. Asimismo, se rechaza la reducción de la ética a parámetros biologicistas y se reivindica una comprensión del ser humano como agente moral teleológicamente orientado. **Discusión.** Frente a la hegemonía cultural de la eficiencia y la fascinación tecnológica, el artículo defiende el papel insustituible de la educación como proceso humanizador. La pedagogía no debe subordinarse a la lógica del rendimiento ni a modas intelectuales, sino custodiar valores esenciales como el esfuerzo, la prudencia y la aceptación de los límites. En este marco, se propone una crítica a la ilusión tecnocrática de la perfectibilidad ilimitada, subrayando que el fin de la educación no es programar sujetos funcionales, sino formar personas capaces de obrar con sentido y responsabilidad en un mundo compartido.

*Palabras clave:* transhumanismo; posthumanismo; pedagogía; mejoramiento humano; naturaleza humana; educación; autonomía.

## Introduction

Three theses are asserted in this article. The first is that education is not a transhumanist activity because the notion of *educational* enhancement differs from the notion of transhumanist enhancement. Secondly, reality is much broader and more diverse and varied than what we can grasp merely with our senses and capacities, and therefore, to understand the world—and education—we need realist or substantialist metaphysics rather than atomistic or analytical metaphysics. And finally, to truly understand education, we need perspectives that focus more on significance than on meaning, theories aimed at comprehension instead of description, and reasonable, rather than rational or empirical, arguments.

In pedagogical thought, there is a tendency not only to accept contemporary intellectual trends but to completely identify with them. Thus, for example, in the 90s, when the idea of a close relationship between education and technology began to develop, especially in Spanish academic circles, it was even asserted that education *is* technology. And when the importance of competency-based education started to become clear, the conclusion was that education *is* competency. To emphasize the ethical dimension of educational processes, it was ultimately asserted that education *is* ethics. Hence, the parts tend to become confused with the whole, especially in fields like pedagogy, in which, in addition to working with imprecise concepts, their meanings are often changed at will. In this regard, as shown in this article, the same thing is currently occurring with transhumanism.

The title and contents of this article are, in part, a response to a recent assertion by professor José Luis Gaviria, who affirmed that ‘education is essentially a transhumanist activity’ because ‘(...) transhumanist techniques and education have something in common. Both endeavour to do something to humans, to transform humans themselves (2024, p. 14). But following this reasoning, we could also assert that ‘education is an agricultural activity because agricultural techniques and education have something in common. Both endeavour to implement, grow, develop, etc.’. Education does, indeed, share transhumanist logic, but it is not a transhumanist activity. Likewise, education is reflected in the agricultural metaphor, but it is not an agricultural activity. Gaviria mistakes the parts for the whole because his theory is based on the erroneous assumption that our obligation as educators is to think about ‘(...) the role of education in transhumanism’ (*op. cit.*, p. 14), when in fact the aim is to contemplate the role of transhumanism *in* education.

However, professor Gaviria is to be acknowledged for recognizing that the discussion around transhumanism and education inevitably leads to the most radical questions that can be raised. This is seen when he notes, for example: '(i)t seems logical that, in that very dynamic of utter doubt, we would contemplate something even more basic: what it means to be a human being.' (*op. cit.*, p. 2); or when, in the midst of his own line of reasoning, he is forced to stop and ask, precisely: 'When do values, and therefore, moral life, arise?' (*op. cit.*, p. 10). Nonetheless, he begins with a somewhat poetic consideration when he argues that '(p)erhaps our very nature is defined by the determination to break any rules that prevent us from pursuing every possibility within our reach, and that the fear of losing paradise will not stop us.' (*op. cit.*, p. 2).

We could debate here whether or not it is advisable to install a stop-light at an intersection, but what is clear is that pedagogy is currently at a crossroads never before contemplated in education (Arana, 2024; Bostrom, 2024; Goering, Klein, Specker Sullivan, *et al.*, 2021; Kurzweil, 2025; Lacalle, 2022). Years ago, Fukuyama (2009) already considered transhumanism was the most dangerous idea in the world. G. Steiner also had such concerns, despite being unaware of today's disconcerting<sup>1</sup> technological progress:

Should genetic research continue, regardless of the social and human consequences? Are there orders of abstract truth that are irreconcilable with human justice? (...). One might think that the yearning for knowledge, the hunter's cry that infected the Western mind nearly 3000 years ago, is leading us into a trap (Steiner, 2022, p. 39 and p. 43).

What follows is an effort to defend the title of the article, starting with more metaphysical and epistemological matters and moving toward more specific ones. As an essay, especially one in response to another, the line of discussion revolves around certain points of friction with Gaviria's proposals. Therefore, some sections are written in a negative tone to accentuate the objections to the contradicting thesis held by said author. Finally, one must not lose sight of the pedagogical perspective

---

<sup>1</sup> This adjective may sound childish, but it is unsettling to find, for example, that in 2019 the Kunming Institute of Zoology in China injected a human gene related to intelligence into macaques, which showed higher cognitive capacities than their peers; that the prevalence of pre-implantation genetic testing on embryos is gradually moving toward eugenic practices: in countries like Iceland and Denmark, the population of people with Down syndrome is virtually non-existent (Lukács de Perény, 2023, p. 155); or that in 2023 the team of researchers led by Jacob Hanna at the Weizmann Institute of Science in Israel managed to create the first complete model of a human embryo without using an egg, sperm or uterus (Oldak, Wildschutz, Bondarenko, *et al.*, 2023)

underlying this entire approach; therefore, if it is not directly observed in certain points, particularly the initial ones, it is not because it is not present, but rather because a preliminary *condition of possibility* is being addressed, without which, at least in the author's opinion, neither education nor pedagogy would exist.

## Reality exceeds the knowable

This premise is relevant in the present analysis, not so much for its implications in the understanding of physical reality, but because it establishes an essential ontological underpinning: the assertion that reality is not limited to what we can know. It opens up the possibility of recognizing values as an objective dimension of life, accessible, but not reducible, to knowledge. This topic will be discussed in greater detail below.

Authors like Gaviria —and with him, those who conceive of transhumanism as a natural evolutionary step— however, take an acritical stance to a premise with Kantian roots: that what we know is what is real. In other words, they confuse epistemology with ontology, identifying the conditions of possibility of human knowledge with the structure of being itself. Thus, empirical knowledge, constructed by means of our cognitive powers (whether natural or artificial), is erroneously taken as a complete and accurate depiction of the essence of things. This is reflected in assertions such as the following: '(m)oral judgements may be subjected to scientific study (...). Therefore, moral judgements can be supported by machines.' And later:

(...) a hybrid or symbiotic transhuman with artificial intelligence could not only have moral life, but could be morally more effective, given that, while it maintains the conscious drive to endure, its moral judgements could be more appropriate because the artificial component would boost its judgement capabilities and therefore its *assessment of the specific circumstances in which the moral dilemma was raised* (*op. cit.*, p. 11; emphasis added).

This is a fallacious assertion because it appears logical but harbours inaccurate premises. Reality is not limited to the capacity to understand it, even if said capacity is enhanced with artificial technology. There is no direct or automatic fixed link between knowledge (always partial, positioned and revisable) of the world's structure and moral action. In fact, no clinical ethics committee would ever assume that a moral judgement could be generated automatically by cross-checking data and theoretical principles. There is no causal or deductive process that intervenes be-

tween the clearest moral principles and the precise action to be taken in a specific situation. This is not an operation that can be solved with algorithms because moral action requires deliberation, interpretation, responsibility and, above all, *experience*, not just calculation. A certain means may be effective to achieve an end, but that does not guarantee that such end has intrinsic value or is humanly desirable. Technical effectiveness may be a necessary condition for an action but it is never sufficient to justify its ethical legitimacy.

In this regard, two important warnings must be added here. The first is related to the risks involved in delegating too much ethical decision-making to automated systems, which could lead to a progressive ‘moral disabling’ or even ‘human moral impoverishment’ that infantilizes humans and impairs our growth as responsible subjects (Green, 2018, p. 22). Another difficulty is what has been referred to as ‘the problem of value alignment’. As Eberl (2018) noted, while it is true that certain bioenhancement methods may serve to augment the ability to reason prudentially and assist moral agents to align their wills with their *higher-order rational desires*, a decisive question remains unanswered: Who determines those desires? How are they formed and legitimized?

Furthermore—and more importantly, from an ontological perspective—if we accept that reality is limited to what our capacities, whether natural or artificial, are able to transform or dominate, we fall into the trap of thinking that *what we can do defines what we are*. Thus, Gaviria mistakenly concludes that an ability to transform or expand is sufficient proof of human nature and its aims. But malleability does not fully explain the knowledge of human nature, just as we do not comprehend the intrinsic value of nature or of living beings simply based on their usefulness or availability to us. Truth does not always lie in what we can do, and human sense is not limited to the capacity to intervene effectively. When what is possible is confused for what is legitimate, we are not exceeding human limits but rather denying their essential profoundness.

## **The natural from a teleological perspective**

The notion of ‘the natural’ can be approached from two traditional perspectives. On the one hand, from a geneticist’s viewpoint, focusing on biological origin: genesis or the principle of life is what defines what is natural. In this respect, the growth of an oak seed according to the genetic information it contains from the beginning is natural. This perspective places the emphasis on the starting point. On the other, a teleological

understanding of nature is possible: that which fulfils its purpose and deploys its full form is natural. In this second example, an oak tree is completely natural not merely because it emerges as such but because it reaches its complete form, with branches, leaves and fruit, thus realizing its *potential*. It is natural not merely because of its origin but due to its *plenitude*.

Transhumanism, however, muddles these distinctions by eliminating, as Gaviria does, the teleological perspective of human nature. Denial of this perspective renders it impossible to adequately respond to certain fundamental questions. Indeed, if there is no ultimate purpose for human beings, on what grounds can we assert that X is better than Y?

If human nature has no inherent purpose, if we eliminate the possibility of finding a normative order from a teleological perspective—the idea of ‘be all that you can be’—it is radically impossible, even pointless, to discuss which transhumanist proposals are acceptable and which are not. The consequences are forthcoming: a ‘(...) human nature naturalized to the extreme is also an available human nature, prone to technical transformation and alleged enhancement’ (Marcos and Pérez Marcos, 2018, p. 16). If there is no defining purpose of human nature to strive for, then there is also no defining purpose of human nature to shy from. If there is no human form to fulfil, then there also no deformation to avoid. Therefore, anything is possible, and everything is allowed.

## **The objective force of a moral norm is not based on its biological usefulness**

One of the central assumptions of transhumanist thought is the reduction of the human being to an exclusively natural or physico-biological entity, comparable to any other living being, whose principal desire is durability, understood as survival and future projection. From this perspective, transhumanism poses no problem at all, but instead appears as a logical extension of the desire to preserve and amplify our adaptive capacities. If, in addition, it is maintained that values arise simply from the conscious desire of an organism interacting with its environment—as Gaviria affirms when he states that “(...) values arise (...) at the moment when a conscious living being confronts the world with a desire of its own” (op. cit., p. 10)—then values do not refer to an objective reality, but are reduced to functional expressions of biological desire. Within this framework, transhumanism appears not only as compatible with ethics, but as something that *enhances it*.

The decisive philosophical problem arises precisely here. If values have their exclusive origin in desire—whether individual or collectively agreed upon—how can their normative force be explained? On what grounds can we speak of moral truth or falsity, of ethical rightness or error? If values are nothing more than projections of subjective preferences or contingent social constructions, it becomes difficult to justify why we regard concepts such as moral error, lying, misrepresentation, or injustice as morally significant. The very idea of a search for greater moral precision presupposes the existence of a reality that transcends our particular perspectives. Only if we acknowledge an objective horizon does it make sense to speak of right and wrong, of truth or falsity, in the moral domain as well. As Thomas Nagel writes:

The serious attempt to identify what is subjective and particular, or relative and communal, in one's outlook leads inevitably to the objective and universal. That is so whether the object of our scrutiny is ethics, or science, or even logic (Nagel, 2000, p. 28).

This means that reality is not merely physical and chemical, but also evaluative and estimative: it is laced with meanings and purposes that are not limited to their biological usefulness. Are environmental friendliness or animal rights solely a response to a shift in cultural viewpoints, or are they an acknowledgement of objective values that were once veiled and have now been *revealed*? Is women's suffrage merely a concession made by men, or is there *something more*? If, throughout history, numerous cultures converge on certain moral principles, might this not be due to a shared perception of a single moral reality linked to human nature? (Lewis, 1990, pp. 81-96).

This need for an ontological point of reference is not exclusive to classical natural law positions. Even authors who are not particularly sympathetic to that approach acknowledge that moral understanding refers to something more than mere consensus. As Savater puts it:

This human capacity to comprehend that which is of the essence is inseparable from the aspiration to natural law (...). *And the ability to recognize that fundamental truth also means that such truth exists, that there is a natural truth, or if one prefers, a true nature, that is radically distinct from opinions that humans have in common* (Savater, 1988, p. 167; emphasis added).

The nature of things must, once again, be acknowledged, especially the ethical assessments that are revealed in human nature itself. Without

ontology, ethics does not exist. Without ontology, human issues, across the board, are relativized. As professor Ibáñez-Martín argued: ‘(...) if there were no truths whatsoever about human beings, we could do whatever we wanted to them, as long as we were powerful’ (Ibáñez-Martín, 2017, p. 21). Indeed:

(u)ltimately, duty is relative to being. And as duty concerns man, it is relative to human beings, to human nature. (...) How could something that has nothing to do with a being be necessary for that being? It’s absurd. *What is good or bad for a being depends on what that being is, not what it “seems to be” or what “I want it to be”...* (Millán-Puelles, 2014, p. 516; emphasis added).

For this reason, the objective force of a moral norm is not grounded in its biological usefulness for the survival of the individual or the group, but in an intrinsic property of the very object to which that norm refers. This distinction becomes especially clear in extreme situations. Consider the gesture of a police officer who, in the face of a kidnapping, offers to exchange himself for one of the hostages because that hostage has young children. From a strictly adaptive or algorithmic logic, it is difficult to morally justify such an action: what kind of machine would recommend it? What survival-maximizing calculation would make it reasonable? Yet from a human point of view, that moral judgment is entirely intelligible. The police officer acts out of the recognition of a higher value, out of the conviction—lived and embodied—that there is something more important than one’s own survival. His decision is not the result of a utilitarian calculation, but a response to an order of meaning inscribed in reality itself, which imposes itself upon him as an objective moral demand and enables him both to know, and to feel, what he ought to do.

## Education is not a scheduled enhancement

It is true that educating entails perfecting, but this does not happen immediately or mechanically. In part, education involves updating the potential inherent to our capacities, but this occurs in a process that is indirect, measured, slow and, above all, profoundly mediated: by cultural activities, by relationships with others (Bárcena, 2025) and by personal effort (Woo, 2024). It is a refinement that requires the subject’s time, effort and participation. Far from being an external optimization or a guaranteed outcome, education is an internal experience of transformation: non-linear, not fully controllable nor assured in its effects, yet intentional

and oriented toward ends that can be reasonably anticipated, though never produced in a mechanical way<sup>2</sup>.

Therefore, educational reasoning is radically different to transhumanist reasoning. The latter view enhancement as a swift technical process with no cultural mediation or social ties, no effort by the subject, no sense of narrative or personal trajectory. Enhancement, in sum, is conceived as an end product, not a path.

In education, not only the results achieved are crucial, but also the very process through which they were achieved. For example, reaching the top of a mountain after a tough climb is not the same thing as being placed there directly. Likewise, improving a text following a harsh critique is not the same thing as accepting unquestioningly the automatic corrections of artificial intelligence software.

Ultimately, education is not about merely improving functions or capacities with some programmed, guaranteed outcome, but rather, it is a process, a path, an enduring intention entailing personal effort. 'And this is what makes education truly magnificent: its self-structuring power, not only as a perfective end but also as a means for our own human development' (Gil Cantero, 2022, p. 23). This development, moreover, gives rise to a certain way of being, a unique narrative and biographical story about our *learner self*, found in the path marked by the anthropological core of the effort itself.

## Education is enhancement with the world

There is one particularly revealing aspect of the way in which education promotes human enhancement that radically distinguishes it from the transhumanist approach. While the latter, as a closed-circuit, technical project, allows for the establishment of an anticipated link between intervention and outcome—the amount of memory to be gained, the intelligence quotient to be reached, how much vision will be improved, how much attention span will be improved—education does not, and cannot, work that way. There is no, nor can there be any, technical guar-

---

<sup>2</sup> I am grateful to one of the article's reviewers for the suggestion to interpret education not only as transformation, but also -in a stronger sense- as transfiguration, understood as the process through which a person attains and reveals their fullness, or that which they are called to become. Without disregarding the interest and pedagogical fruitfulness of this perspective (Mortari, Valbusa, & Bombieri, 2026; Gil Cantero, 2025; García Gutiérrez, 2023; López Quintás, 2014), this paper has deliberately opted for the concept of transformation, in order to underscore the internal, non-technical, and non-guaranteed character of educational change, as well as its constitutive openness, without introducing from the outset a fully developed teleology that might shift the argumentative focus of the proposed analysis.

antee of results, no cause-effect linearity that determines success. And not just because we cannot precisely predict the outcomes, but —more profoundly— because we do not want to reduce human enhancement to a matter of functional efficiency.

Transhumanism views enhancement as an abstract, enabling improvement of functions unrelated to any content or context. Education, on the other hand, is always a *process of improvement embedded in the world*: it takes place within a framework of cultural mediation, personal ties and experiences. In short, one might say that transhumanist technology develops without life experience whereas education takes place with the world, in the world and for the world. *With the world*, because it takes place through cultural mediation of contents; *in the world*, because it is rooted in specific experiences, in human relationships (Vila and Álvarez, 2025) and in social contexts; and *for the world* because it pursues a shared reality, not just enhancing the individual.<sup>3</sup>

In line with transhumanist thinking, Gaviria argues that the ideal of human enhancement ‘(...) (is) not so different to the optimizing nature of education. What is achieved by technical means in one case takes place through symbolic elements, cultural contents and personal relationships in the other.’ (*op. cit.*, p. 12). But this equivalence loses sight of something essential: it is not possible to liken a symbolic, embodied, relational enhancement with functional, abstract, technical optimization. For example, what effect would transhumanist interventions have on a wild child deprived of the world of education? It would not be significant. The ‘good savage’ cannot realize their humanity without that shared world that only education can provide.

Therefore, transhumanism and education are not merely distinct means to the same end, but rather essentially different, non-interchangeable processes. Education responds to an anthropological structure that transhumanism cannot replicate because human nature has unique cultural, symbolic and relational demands that mere technical enhancements cannot replace. Therefore, it is incorrect to assert that education is a form of transhumanism.

## **Productive activity is not the same as perfective action**

At this point, a key issue must be addressed: if, as argued above, transhumanism and education are not interchangeable enhancement processes, why should we reject the idea that education could begin at a techno-

<sup>3</sup> Argument borrowed from Gomá (2023, p. 25 et seq.), used to develop his idea of a worldly philosophy.

logically enhanced ‘starting point’? If artificial means exist for improving memory, attention, intelligence, physical endurance or sensory acuity, why not implement them before starting to educate?

Some transhumanists with humanistic leanings strive to reach a time-based compromise: first, technological enhancement of certain capabilities and then education as a comprehensive process. In this regard, Gaviria holds that ‘transhumanist artificial enhancement is nothing more than defining a new starting point, one in which education begins at a higher level’ (*op. cit.*, p. 17). This approach has one significant strength: it acknowledges the inevitable need for education. In other words, no enhanced state —whether artificially or otherwise— of cognitive, sensory, or physical capabilities can guarantee a fully educated individual.

However, while this may be true, it should be noted that the artificial enhancements of transhumanism are not a *neutral* starting point from the perspective of education. As Žižek observed:

(...) what if our inner life, inclusive of our highest spiritual achievements, is rooted in our finite bodily existence and its limitations, so that, with the transition into Singularity, we are deprived of the basic features of our inner life? (2023, p. 157).

Indeed, such interventions could impair the individual’s willingness to engage in their own development: Why should I make an effort to develop my ability to concentrate if I already pay attention easily thanks to an external intervention? Why bother to push my physical limits if my capabilities have been modified to perform without fatigue? In fact, this thinking is already being seen with artificial intelligence: the effort I no longer make to write a summary, understand a paragraph in a book or prepare a hiking path myself is a *lost opportunity in education* and personal improvement that no longer stays with me and does not shape or form me, despite the fact that the results of the summary, my understanding of the paragraph or the hiking route are better than what I could have done myself. In Aristotelian terms, I have performed a productive activity but not a perfective action. The outcome has indeed improved, but the process took place externally, without shaping me, leaving no marks, without living or experiencing it internally.

## **Autonomy: a sufficient pedagogical criterion?**

After accepting the proposed time sequencing —transhumanist interventions first, followed by education— advocates of transhumanism assert that the only remaining debate is which enhancements are acceptable

and which are not. Gaviria concludes his article with this consideration, brilliantly, by the way, and fully in harmony with the text by Bostrom (2005, p. 9 et seq.). The proposed criteria will not be discussed individually here, given that, as one can see, they are reasonably supported on the grounds of ethical achievements linked to respect for human rights. What *is* questionable is the guiding principle, or metacriterion, that defines and justifies the rest: individual autonomy as the ultimate educational measure. As Gaviria explained:

(...) it is perfectly valid to establish certain criteria to determine which biotechnical interventions are acceptable and which are not. I believe the main criterion has to do with the individuals' autonomy. It is a matter of distinguishing between what increases our autonomy and what decreases it. Increasing our abilities boosts our autonomy. Implementing contents impervious to criticism or uncontrollable automatic behaviours reduces our autonomy (*op. cit.*, p. 20).

However, this approach falls into a common trap when considering the relationship between education and transhumanism from a strictly scientific and naturalistic viewpoint: it assumes that autonomy is a measurable reality that can be expanded or reduced in a technical, linear fashion, as if it were a function independent of the individual and their lived experience in the world.

To start with, it is unclear what, exactly 'increasing our autonomy' entails. Who is more autonomous in a hyper-consumerist society: someone with unlimited resources to consume, or someone who chooses to live austere? Who is more autonomous when confronting religious tradition: someone who sticks to their faith with conviction, or someone who rejects it in the name of personal autonomy? Who is more autonomous: the child that becomes a supporter of Atlético de Madrid like their parents, or the one who changes teams and decides to follow Real Madrid?

Furthermore, it is also unclear what *specific* kinds of capacities increase our autonomy. Intelligence, willpower, emotions, attention, freedom...? Does the capacity to fall in love increase our autonomy? Does the capacity to raise a child increase our autonomy? Liberals, for example, believe that what makes us more free —autonomous— is ownership, while others believe it enslaves us. What's more, there is no upward trend of increasing autonomy in the human being that guarantees a parallel increase in their happiness.

Upon deeper reflection, it should be noted that 'true education', to quote Gaviria, does not consist in maximizing autonomy as if it were an end in and of itself or a cumulative indicator of improvement. This sug-

gestion not only leads to a functionalist reduction of education but also, once again, confuses the parts with the whole. Educating is not merely about expanding the individual's possibilities for action, but about *shaping their inner willingness to freely base their actions on a greater good through prudent, responsible judgement*. From this perspective, autonomy must not be viewed as a technical capability to be acquired, but as an ethical virtue to be cultivated. It is shaped rather than gained; it is not measured by the number of options available or by efficacy of execution, but by the capacity to discern what is valuable and to act accordingly. A truly educated individual is not the person who has the most autonomy, but the person who knows how to use their autonomy in a humanizing way, for the greater good and serving others. For this reason, humanity's greatest moral role models —heroes, saints, sages— are not admired for the magnitude of their autonomy, but for their noble, just and compassionate use of that autonomy.

## **The task of pedagogy in the face of transhumanism**

Transhumanism has gone from being a futuristic theory to become a cultural and technological process in motion and increasingly accepted by society as part of the realm of possibility. The supply of artificial enhancements —some of which are already available— will grow and become increasingly natural in our lives. Even if international institutions impose ethical or legal limits, parallel initiatives will inevitably arise to bypass them, such as biohackers or 'garage biologists' operating outside the system.

In light of this scenario, it does not seem advisable for the field of pedagogy to become directly involved in the technical, legal or sociological assessment of each new artificial enhancement. While necessary, these tasks belong to other fields. What pedagogy should do, however, if it is ambitious, is to protect a legacy and hold a conviction. The *legacy* of safeguarding the humanizing richness of education: helping to reach a level of maturity that highlights higher-order rational desires in relation to that which truly enriches human life. The *conviction* that education is not limited to functional enhancement of human beings but must aim to shape a unique, inherent biography that strives to be better each day through personal effort, geared toward inner change as an ongoing duty rather than as artificial enhancement.

In an era of fascination with possibility, we must remember what is valuable. It is not our job to keep pace with biotechnological innovation, but to preserve formative values that, even if they end up being

considered marginal and antiquated, represent the humanizing core of all educational experience: effort, self-acceptance, prudence in the use of freedom, relating to others in a shared world, and awareness of our vulnerability and dependence.

Of course, I realize that many would prefer '(...) to settle for the feeling of victory without setting foot on the basketball court or feeling loved without a lover' (Güell, Echarte-Alonso and Murillo, 2019, p. 207), but the response of pedagogy should be clear: it is not the same thing, nor will you be the same if you choose one or the other.

## References

- Arana, J. (Dir.). (2024). *El futuro de la identidad humana a debate. Protagonistas de la polémica sobre el transhumanismo*. Tecnos.
- Bárcena, F. (2025). Una filosofía de la relación educativa: mediación existencial, transmisión y testimonio [A Philosophy of the Educational Relationship: Existential Mediation, Transmission and Testimony]. *Teoría de la Educación. Revista Interuniversitaria*, 37(2), 1-18. <https://doi.org/10.14201/teri.32244>
- Bostrom, N. (2005). Transhumanist values. *Review of Contemporary Philosophy*, 4(1-2), 87-101. <https://nickbostrom.com/ethics/values.pdf>
- Bostrom, N. (2024). *Deep Utopia. Life and Meaning in a Solved World*. Ideapress.
- Eberl, J. T. (2018). Can Prudence Be Enhanced? *Journal of Medicine and Philosophy*, 43(5), 506-526. <https://doi.org/10.1093/jmp/jhy021>
- Fukuyama, F. (2009, 23 de octubre). Transhumanism, *Foreign Policy*. <https://foreignpolicy.com/2009/10/23/transhumanism/>
- García Gutiérrez, J. (2023). Pensar la antropología educativa desde la transfiguración: Apuntes críticos al transhumanismo desde una pedagogía de la vida del espíritu. *Temps d'Educació*, 65, 93-111. <https://doi.org/10.1344/TempsEducacio2024.65.7>
- Gaviria, J. L. (2024). ¿Transhumanismo 'contra' educación? [Transhumanism 'against' Education?]. *Teoría de la Educación. Revista Interuniversitaria*, 36(2), 1-23. <https://doi.org/10.14201/teri.31762>
- Gil Cantero, F. (2022). La Pedagogía ante el desfase prometeico del transhumanismo. *Revista de Educación*, 396, abril-junio, 11-33. <https://doi.org/10.4438/1988-592X-RE-2022-396-528>
- Gil Cantero, F. (2025). Controversias en torno a la educación transformadora. En C. Naval y A. Bernal (Eds.), *La educación transformadora, luces y sombras* (pp. 31-64). Eunsa.

- Goering, S., Klein, E., Specker Sullivan, L. *et al.* (2021) Recommendations for Responsible Development and Application of Neurotechnologies. *Neuroethics*, 14, 365-386. <https://doi.org/10.1007/s12152-021-09468-6>
- Gomá, J. (2023). *Universal concreto. Método, ontología, pragmática y poética de la ejemplaridad*. Taurus.
- Green, B. P. (2018). Ethical Reflections on Artificial Intelligence. *Scientia et Fides* 6(2), 9-31. <http://doi.org/10.12775/SetF.2018.015>
- Güell, F., Echarte-Alonso, L. E. & Murillo, J. I. (2019). Objetividad y bi-omejora humana. La paradoja de lo natural. *Scientia et Fides*, 7(2), 195-214. <https://doi.org/10.12775/SetF.2019.023>
- Ibáñez-Martín, J. A. (2017). *Horizontes por los educadores. Las profesiones educativas y la promoción de la plenitud humana*. Dykinson.
- Kurzweil, R. (2025). *La singularidad está más cerca: cuando nos fusionamos con la IA*. Deusto.
- Lacalle, M. (Dir.).(2022). *Transhumanismo: ¿homo sapiens o ciborg?* Universidad Francisco de Vitoria.
- Lewis, C. S. (1990). *La abolición del hombre*. Encuentro.
- López Quintás, A. (2014). Transfiguraciones que deciden el proceso de crecimiento personal. En Autor, *La ética o es transfiguración o no es nada* (357-396). BAC
- Lukács de Perény, M. (2023). *Neo entes. Tecnología y cambio antropológico en el siglo XXI*. Homo Legens.
- Marcos, A. y Pérez Marcos, M. (2018). *Meditación de la naturaleza humana*. BAC.
- Millán-Puelles, A. (2014). *Obras completas*. Vol. V. *Economía y libertad* (1974). Rialp.
- Mortari, L., Valbusa, F., y Bombieri, R. (2026). Aprendizaje socioemocional y florecimiento humano. *Teoría de la Educación. Revista Interuniversitaria*, 38(1), 123-145. <https://doi.org/10.14201/teri.32149>
- Nagel, Th. (2000). *La última palabra*. Gedisa.
- Oldak, B., Wildschutz, E., Bondarenko, V. et al. (2023). Complete human day 14 post-implantation embryo models from naive ES cells. *Nature* 622, 562–573. <https://doi.org/10.1038/s41586-023-06604-5>
- Savater, F. (1988). *Ética como amor propio*. Mondadori.
- Steiner, G. (2022). *¿Tiene futuro la verdad? El hombre al norte del mañana*. Almuzara.
- Vila, E. S. y Álvarez, V. E. (2025). La relación educativa en un mundo digitalizado: paradojas de la hiperconexión [The Educational Relationship in a Digitalized World: Paradoxes of Hyperconnection]. *Teoría*

- de la Educación. Revista Interuniversitaria*, 37(2), 19-36. <https://doi.org/10.14201/teri.32283>
- Woo, J. G. (2024). Education amid the deluge of enhancement discourses. *Educational Philosophy and Theory*, 56(12), 1198–1209. <https://doi.org/10.1080/00131857.2024.2364036>
- Žižek, S. (2023). *Hegel y el cerebro conectado*. Paidós.



# Differential item functioning analysis in large-scale assessments

## Análisis de funcionamiento diferencial en evaluaciones a gran escala

<https://doi.org/10.4438/1988-592X-RE-2025-412-733>

**Laura Ortega Torres**

<https://orcid.org/0009-0009-5776-0581>

*Centro Nacional de Evaluación para la Educación Superior, CENEVAL (México)*

**César Antonio Chávez Álvarez**

<https://orcid.org/0009-0000-4197-8281>

*Centro Nacional de Evaluación para la Educación Superior, CENEVAL (México)*

### Abstract

Differential Item Functioning (DIF) analyses are essential to support the quality and fairness of an assessment, as they identify items that may be biased. Various methods have been developed to perform these analyses, and their effectiveness has been compared by manipulating different variables, including the number of examinees. Typically, studies focus on groups with some thousands of participants; however large-scale assessments are administered to a significantly larger number of individuals. This study explored the effectiveness of four traditional methods for analyzing DIF in samples similar in size to the groups of examinees taking the Entrance exams (EXANI by its acronym in Spanish) in a single administration in Mexico and that could be applied in other entrance tests administered nationwide or complete administrations of International Large-Scale Assessments (ILSA). Simulations were conducted, in which samples of sizes comparable to those in the EXANI's administrations were generated, and the difficulty parameter values of 10% or 20% of the items of one of their content areas were manipulated to exhibit moderate and high Uniform DIF. The effectiveness of the Mantel-Haenszel, Logistic regression, Lord and Raju methods was verified by comparing the percentage of items correctly detected with their degree of DIF (manipulated items) as well as the percentage of false detections in non-DIF items. The Mantel-Haenszel, Lord and Raju methods were the most effective in detecting uniform DIF in the simulated samples, confirming their potential im-

plementation in large-scale assessments with dichotomous items for samples of up to 200,000 examinees.

*Keywords:* Uniform differential item functioning, Dichotomic items, Large-scale assessments, Mantel-Haenszel Method, Logistic Regression Method, Lord Method, Raju Method

### Resumen

Los análisis de funcionamiento diferencial de los ítems (DIF, por sus siglas en inglés), son esenciales para sustentar la calidad y equidad de una evaluación, ya que señalan ítems que pudieran estar sesgados. Se han desarrollado diferentes métodos para realizar estos análisis y se ha comparado su eficacia por medio de la manipulación de diversas variables, incluido el número de participantes. Generalmente se estudian grupos con hasta algunos miles de integrantes, sin embargo, existen evaluaciones a gran escala que se aplican a un número considerablemente mayor de personas. Este estudio exploró la eficacia de cuatro métodos tradicionales para analizar DIF en muestras de personas con un tamaño similar al de los grupos de evaluados que presentan los Exámenes Nacionales de Ingreso (EXANI) en una misma aplicación en México y que pudieran usarse para analizar otras pruebas de ingreso que se apliquen a nivel nacional o aplicaciones completas de Evaluaciones Internacionales a Gran Escala (ILSA, por sus siglas en inglés). Se realizaron simulaciones en las que se generaron muestras de tamaños similares a las de las aplicaciones de los EXANI y se manipularon los valores de los parámetros de dificultad del 10% o el 20% de los ítems de una de sus áreas de contenido para que presentaran DIF Uniforme moderado y alto. Se verificó la eficacia de los métodos de Mantel-Haenszel, Regresión Logística, Lord y Raju, comparando el porcentaje de ítems detectados con su grado de DIF de forma correcta (ítems manipulados), así como el porcentaje de falsas detecciones en los reactivos sin DIF. Los métodos de Mantel-Haenszel, Lord y Raju fueron los más eficaces para detectar funcionamiento diferencial uniforme en las muestras simuladas, por lo que se confirma su posible implementación en evaluaciones a gran escala con ítems dicotómicos con muestras de hasta 200,000 evaluados.

*Palabras clave:* Análisis de Funcionamiento Diferencial Uniforme, Ítems Dicotómicos, Evaluaciones a Gran Escala, Método Mantel-Haenszel, Método de Regresión Logística, Método de Lord, Método de Raju

## Introduction

A fundamental property of any assessment is that it should not favor any person or group with specific characteristics. Differential Item Functioning (DIF) analyses help to determine that measurements are not biased,

and they reflect whether the same construct is being assessed in all examinees; they are therefore essential in test development and validation processes (Walker, 2011). Specifically, standard 4.10 of the Standards for Educational and Psychological Testing suggests using these analyses to determine a portion of the psychometric properties of items (AERA et al., 2014).

DIF means that when two groups of people with the same ability to be measured have different probabilities of success in answering an item, in other words, after controlling for their ability, it is observed that this probability differs between the two groups. An item may behave statistically differently from the set of items with which it was assembled without presenting a bias in content (Wu et al., 2016). Therefore, DIF flagging is a necessary though not sufficient condition for determining bias. To do so, the cause of the differential functioning must be attributed to a factor irrelevant to the construct, and further content analysis is necessary to include both statistical support and qualitative interpretation indicating that DIF is due to a characteristic of the item that is irrelevant to the purpose of the assessment, rule it as biased, and decide whether to send it for review or remove it from the bank (Sireci & Rios, 2013; Zwick, 2012).

DIF analyses seek to identify items that are particularly difficult for members of specific groups, comparing participants of different genders, cultures, races, or who speak different languages. These variables remain at the core of DIF analyses in different tests or specific instruments or as part of longitudinal studies around the world (Cassani-Miranda et al., 2025; Shamsaldeen et al., 2024; Suk & Han, 2024; Villalonga-Olives et al., 2024) to rule out that the differences observed in these groups are due to intrinsic characteristics of the items.

There are several methods for performing these analyses, which are supported by simulation studies as part of the research for their use. In these studies, different variables are manipulated because some may influence the detection of differential functioning. These manipulations may be common procedures for different DIF methods, such as purification, which refers to the iterative removal of items with differential functioning to improve the accuracy of the results (French & Maller, 2007; Özdemir, 2015; Wang et al., 2009). They can also be based on general assessment conditions to analyze their most common scenarios and include the number of items that make up a test or area, the range of values taken by the psychometric item parameters (such as their difficulty and discrimination), the degree of DIF, the number of participants per group, and their score distributions (Fidalgo et al., 2000; Gierl et al., 2004; Liu et al., 2019; Rogers & Swaminathan, 1993; Wang et al., 2009).

Other manipulations have an empirical basis. For example, 10% of items with DIF is used, as this is the percentage found in standardized achievement tests, or which is related to race and gender. Alternatively, 20% of items with DIF is employed, which is associated with different learning opportunities. (French & Maller, 2007; Narayanan & Swaminathan, 1994).

However, manipulating the number of participants in previous simulation studies cannot be generalized to large-scale tests, such as those used in International Large-Scale Assessments (ILSAs). DIF studies that manipulate group sizes to be compared report numbers between 100 and 2,000 participants (French & Maller, 2007; Gierl et al., 2004; Halpin, 2024; Narayanan & Swaminathan, 1994; Park et al., 2021; Wang et al., 2009). However, instruments that are part of ILSAs are answered by a larger number of test takers. For example, in the Programme for International Student Assessment (PISA), almost 700,000 students from 81 countries can participate (OECD, 2023); around 300,000 students per grade from 50 countries can take the Trends in International Mathematics and Science Study (TIMSS) (von Davier et al., 2024); the Progress in International Reading Literacy Study (PIRLS) is taken by almost 400,000 students from over 50 countries (Fishbein et al., 2024); more than 80,000 students from 24 countries sit the International Civic and Citizenship Education Study (ICCS) (Schulz et al., 2025), and the International Computer and Information Literacy Study (ICILS) can be administered to nearly 133,000 students from 34 countries (Fraillon, 2024).

Around the world, tests are also used as tools in the entrance process to a country's education system (Velázquez Ramírez, 2022) and can be administered to a large number of people. In Madrid, Spain, in 2024, the University Entrance Examination (EvAU) was taken by over 30,000 applicants (Autonomous University of Madrid, 2024). In Mexico, the National Assessment Center for Higher Education develops the National Entrance Exams (EXANI), which are used to enter high school and higher education levels and are taken by hundreds of thousands of applicants.

The results of all these instruments have a high impact on a national or personal level, which makes it paramount to ensure that their items are free of bias. Traditional DIF methods have been used in studies with some ILSAs, though with samples of between approximately 500 and 7,000 students (Abulela & Rios, 2022; Demirus & Pektaş, 2022; Kucam & Gülleroğlu, 2023; Liu & Bradley, 2021; Özdemir, 2015; Uyar, 2020), but not on larger populations. Testing administration modes to enter higher education have also been compared with EXANI-II using DIF analysis and comparable sizes (Vidal González et al., 2022). However, to compare groups with a larger number of participants, it is necessary to prove that

these methods are robust under these conditions, since neither simulation studies nor those using empirical test data have tens or hundreds of thousands of observations. This study addresses this issue by conducting simulations with sample sizes and other conditions that are comparable to those of EXANI assessments, testing the effectiveness of four traditional methods. To perform any DIF analysis, it is necessary to define a comparison variable, which is usually a participants' demographic variable. It is possible to make comparisons among several groups, but typically only one pair is compared. The next step is to identify the comparison groups, where one is named the reference group and the other the focal group. The former is the standard against which the latter is compared, and tends to be a majority group, while the latter is a group of special interest and usually comprises the minority of examinees (Zumbo, 2007). Finally, the conditional variable to link the examinees is usually their test score. If the exam is composed of different areas, it is preferable to use their scores separately (Sireci & Rios, 2013).

There are two types of DIF depending on the existence of an interaction between the ability level and the group to which the examinee belongs, and both stem from additional traits to the main ability measured by the item, for example, verbal abilities in a math item (Mellenbergh, 1982). In this way, members of two groups with the same main ability may differ in a secondary trait, affecting their performance on the item. A constant difference across the primary ability indicates uniform DIF, whereas a difference that varies with the primary ability indicates non-uniform DIF. Since there is no interaction between ability level and group membership in uniform DIF, an item is easier for all members of one group than for members of the other, giving the former a consistent advantage (Walker, 2011). This DIF occurs most frequently in standardized tests (Narayanan & Swaminathan, 1996), and it has been suggested that attention should focus on detecting it, especially when tests consist of dichotomous items, since non-uniform DIF depends on specific interactions that occur more frequently in polytomous items (Jodoin & Gierl, 2001).

There are several methods for analyzing DIF (for reviews, see Magis et al., 2010, and Ohiri et al., 2024), however, four methods, although developed several decades ago, remain widely used and serve as reference points for new approaches. Unlike other methods lacking criteria for classifying different levels of DIF, these methods have criteria for quantifying effect size with large populations (Sireci & Rios, 2013) and can therefore provide specific and standardized information on item functioning in large-scale assessments.

The Mantel-Haenszel method, which is not based on Item Response Theory (IRT), is easy to use and does not require large samples of subjects. It tests whether there is an association between group membership and item response conditionally on the total test score (Magis et al., 2010). It compares the odds ratio of correct responses for the two groups for different ability intervals, based on a  $\chi^2$  test (Dorans & Holland, 1992). The result is the Mantel-Haenszel (MH) alpha, a statistic that takes values from zero to positive infinity, where a value of one indicates that there is no difference in the performance of the groups; if the value is greater than one, the performance of the reference group is better than the focal group's, and the opposite occurs if said value is less than one.

MH alpha is usually converted to a symmetrical scale known as Delta, whose statistic is calculated using the following formula:  $-2.35 \cdot \ln(\text{MH alpha})$  (CENEVAL, 2009). This statistic is interpreted as a difference between the difficulty of the items for the reference and focal groups on the Delta scale (Roussos et al., 1999). Positive values will indicate that the item works to the advantage of the focal group (it is more difficult for the reference group); negative values indicate that it favors the reference group (it is harder for the focal group); and a value of zero indicates that there is no DIF. Its absolute values are also used to classify the level of DIF:

**Table I.** Interpretation of the Mantel-Haenszel Delta statistic

Level of DIF	Delta Value
Negligible or Nonsignificant	$ \delta  < 1$
Slight to Moderate	$1 \leq  \delta  < 1.5$
Moderate to Large	$1.5 \leq  \delta $

Source: Prepared by the authors based on Zwick, 2012

An absolute Delta value lower than one indicates that the difference between the reference and focal groups is not significant at a level of 5%, so its degree of DIF is negligible. An absolute Delta value greater than or equal to 1.5 means that the difference between both groups is significantly greater than one at the level of 5% and is classified as Moderate to Large. Items that do not meet the above criteria are classified as having Slight to Moderate DIF, because a Delta difference of one point, although undesirable, can be tolerated, while a Delta difference of two or more should be avoided, with 1.5 representing the lower limit (Zwick, 2012). The Mantel-Haenszel method is a benchmark for others because

of its easy and quick application, and since it has DIF level classification criteria, satisfactorily detects uniform DIF, and can help detect non-uniform DIF.

Swaminathan and Rogers (1990) indicated that the Mantel-Haenszel method does not have statistical power to detect non-uniform DIF and proposed a method based on logistic regression to detect both types of differential functioning. A logistic model is fitted with the probability of correctly answering an item as a dependent variable, and the total test score, group membership, and the interaction between the two as independent variables. Uniform DIF is tested through the effect of the group, and non-uniform DIF through the interaction of the variables. Zumbo and Thomas proposed criteria for detecting DIF based on the difference in the coefficient of determination for the logistic regressions ( $\Delta R^2$ ) as a measure of effect size (in Magis et al., 2010). Subsequently, Jodoin and Gierl (2001) proposed a less conservative scale (see Table II), reporting that with this new scale, the percentage of items detected with Slight to Moderate DIF increased from 6.8% to 68.2%. In contrast, it has been shown that the percentage of items reported with DIF is lower compared to those detected with the Mantel-Haenszel Delta (Hidalgo & López-Pina, 2004).

**Table II.** Interpretation of the logistic regression  $\Delta R^2$  statistic

Level of DIF	Zumbo and Thomas Criteria	Jodoin and Gierl's Criteria
Negligible	$\Delta R^2 \leq 0.13$	$\Delta R^2 \leq 0.035$
Slight to Moderate	$0.13 \leq \Delta R^2 < 0.26$	$0.035 \leq \Delta R^2 < 0.07$
Moderate to Large	$0.26 \leq \Delta R^2$	$0.07 \leq \Delta R^2$

Source: Prepared by the authors based on Magis et al., 2010

The logistic regression method is considered an extension of the Mantel-Haenszel method. It was designed to detect both types of differential functioning, using a recognized, flexible, and easy-to-use model that is suitable for small samples (Jodoin & Gierl, 2001) and has lower false positive rates than the Mantel-Haenszel method (Abulela & Rios, 2022).

The Lord method is based on the null hypothesis that both groups have the same estimated item parameters based on IRT and a statistic for their difference with a  $\chi^2$  distribution (Lord, 1980 in Magis et al., 2010). Any IRT model can be adjusted to detect uniform and non-uniform DIF, as long as the parameters are scaled to a common metric before performing the statistical test. Its usefulness in detecting DIF has been proven

because it controls the detection of false positives (Uyar, 2020), although it is not always consistent with the Mantel-Haenszel method, as the latter can detect items with DIF that are not detected by the Lord method (Raju et al., 1993).

This method requires moderate to large sample sizes for stable parameter estimates, more advanced knowledge of psychometrics, and more complex programming and calculations. Therefore, if exams are designed under any IRT model, this method is very easy to apply as it is an extension of commonly performed analyses. Criteria have only been established to determine the magnitude of differential functioning for the single-parameter model, which is used to measure uniform DIF, as it only calculates the discrepancy between the difficulty parameters of the focal and reference groups. If the difference is negative, the difficulty parameter of the focal group is lower than that of the reference group, and the former gets favored. If the difference is positive, the difficulty parameter of the focal group is greater than the reference group's, and the latter is favored. Effect size is estimated as 2.35 times the difference between the item difficulties calculated for the groups under analysis (Penfield & Camilli, 2007). Since this statistic is solely based on item difficulties, it is similar to Mantel-Haenszel's, and therefore the same Delta scale criteria are used to interpret it (see Table I).

The Raju method was developed because one of the main assumptions of the Lord method is that the participant's ability is known. However, it has been shown that when both the ability and item parameters are unknown, Type I errors may occur (Cohen & Kim, 1993). To address this, another IRT-based approach was proposed to detect DIF by estimating the area between the Item Characteristic Curves of both groups under study and the Z statistic being the null hypothesis that the true difference between the areas is zero (Raju, 1990). Once again, any IRT model can be applied and therefore used to detect both types of DIF. Given the more complex calculations required by this method, Lord's method is often regarded as more efficient. However, when the sample size is small and there is little DIF, Raju's method is more effective (Hidalgo-Montesinos & Lopez-Pina, 2002).

This method solely has criteria for the single-parameter model and uses the difference between the difficulty parameters and the Penfield and Camilli (2007) procedure to estimate the direction of bias and effect size, as well as the Delta scale criteria (see Table I). Therefore, although the statistics for estimating the differences for the Lord and Raju methods differ (Özdemir, 2015), the estimation of the bias size for the single-parameter model in both methods is identical.

## Method

Simulations were carried out in which groups with numbers of participants similar to those in EXANI populations and other design characteristics of these exams were replicated, such as the number of items or score distribution. Based on the items in one of these tests, some were manipulated to show Slight to Moderate or Moderate to Large uniform DIF. Subsequently, DIF analyses were performed using the Mantel-Haenszel, logistic regression, Lord, and Raju methods, and correct and incorrect detections as well as non-detections of differential functioning were compared.

## Sample

EXANI populations are large, although they differ from one exam to another. In 2019<sup>1</sup>, one version of EXANI-III (entrance exam for graduate education) was taken by 14,650 participants, 1,698 of whom were applying to the same institution; one version of EXANI-I (entrance exam for high school) was taken by 114,757 participants, with 77,449 examinees in a single entrance process, and one version of EXANI-II (entrance exam for undergraduate education) was taken by 202,264 participants, 42,966 of whom were applying to the same institution. Based on these numbers and to give continuity to previous studies, it was determined that each reference and focal group would consist of 400, 1,000, 2,000, 10,000, 50,000, and 100,000 members, with a total of 800, 2,000, 4,000, 20,000, 100,000, and 200,000 simulated participants per study.

Since EXANI exams are designed to produce normally distributed scores with an average of 50% correct answers, the abilities of all comparison groups in this study were simulated with a mean of zero and a standard deviation of one. This approach ensured that any observed differences were attributable solely to the number of examinees in the large-scale assessment, rather than other factors.

## Instruments

EXANI tests are high-impact, high-demand, norm-referenced score tests aimed at populations participating in entrance exams for high school (EXANI-I), undergraduate education (EXANI-II), and graduate education (EXANI-III) in Mexico. Until mid-2021, these exams consisted of an entrance test and a diagnostic test, with the former having a similar number of areas and items: EXANI-I and II consisted of four areas; in EXANI-I,

---

<sup>1</sup> This year was used as a reference to avoid biased data due to the pandemic.

each area had 20 items, and in EXANI-II, there were 25 items per area. EXANI-III had seven areas, each with 20 items. Therefore, 20 was the number of items simulated for this research. All items in this generation of exams were dichotomous and were analyzed using the two-parameter IRT model to determine their statistical quality.

**Procedure**

The two-parameter IRT model was used to estimate item parameters; the difficulty and discrimination values were calculated for the items assessing reading comprehension in a version of the EXANI-I Entrance test taken by 76,772 participants. The minimum discrimination value was 0.520 and the mean difficulty was 0.237. Of the 20 items, four were selected, two with medium difficulty values and two with extreme values (an easy one and a difficult one). These values were manipulated so that the absolute difference between them reflected Slight to Moderate or Moderate to Large DIF, modifying their difficulty parameters so that the absolute differences between them were 0.6 or 1.1. The following table shows the original difficulty values, the modified values, and their absolute differences.

**Table III.** Values of the discrimination parameter (a) and difficulty parameter (b) of the DIF- simulated items

Item	Discrimination	Difficulty			Absolute Difference (bf -br)	DIF Size	Favored Group
		Original	Reference Group (br)	Focal Group (bf)			
2	1.317	-1.255	-0.755	-1.855	1.1	Moderate to Large	Focal
8	1.073	-0.329	-0.029	-0.629	0.6	Slight to Moderate	Focal
13	1.131	0.494	-0.006	1.094	1.1	Moderate to Large	Reference
19	1.112	1.598	1.298	1.898	0.6	Slight to Moderate	Reference

Source: Authors' own work

Table III shows resulting items with Slight to Moderate and Moderate to Large DIF values, with medium, high, and low difficulties, favoring both groups. In terms of discrimination, the four values were close to 1. Ten percent or 20% of the items with DIF were manipulated, using the first two values in Table III or all four.

The simulations were run using R version 3.5.2 (R Core Team, 2018) and the DIF analyses used version 5 of the difR package (Magis et al., 2010).

The abilities and scored answer strings were simulated for each reference and focal group (ranging in size from 400 to 100,000), based on both the parameter values of the items corresponding to the study (10% or 20% with DIF) and the mean and standard deviation of ability specified from the EXANI sample (zero and one, respectively). The absolute differences between the mean abilities of the focal and reference groups were lower than 0.0052, and their standard deviations were below 0.0026.

Once the scored answer strings were obtained, they were labeled with the corresponding group (focal or reference) and integrated to perform DIF analyses using the Mantel-Haenszel, logistic regression, Lord, and Raju methods. In the Lord and Raju methods, only the Rasch model was used (the discrimination parameter had a fixed value of 1). In all simulations, the effect size statistic was calculated for each of the 20 items, together with the DIF classification based on the criteria defined for each method.

One hundred iterations were carried out for each sample of participants and each percentage of items with DIF, beginning with the simulation of abilities and answer strings for each group and ending with the DIF analyses. Based on the 100 iterations, the percentage of items classified as Negligible (A), Moderate (B), or Large (C) DIF was calculated for each method in each item, yielding detection and non-detection rates. When the item was manipulated to have a moderate degree of DIF, the rates of correct detections (classified as B), incorrect detections (classified as C), and non-detections (classifications as A) were calculated. When the item was manipulated to have Large DIF, the rates of correct detections (classified as C), incorrect detections (B ratings), and non-detections (A ratings) were determined.

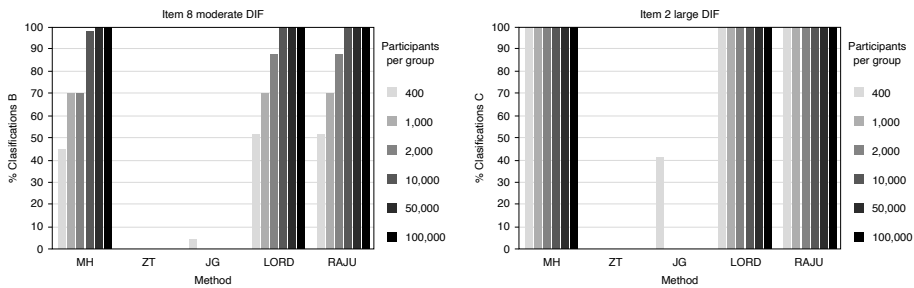
For the remaining items whose parameters were not manipulated and were identical in both groups, the proportions of incorrect DIF detections were calculated together as they were classified as having DIF (Moderate or Large).

## Results

First, method effectiveness was verified in the test on 10% of the items with differential functioning. Graph I shows the percentage of correct detections in the two items with DIF. The Mantel-Haenszel, Lord, and

Raju methods detect both Moderate and Large DIF. In the detection of the former, sample size has a clear effect: as the number of participants increases, the number of correct DIF detections also rises. With 400 participants, approximately 50% of items were correctly detected across all three methods, while with 10,000 participants, this percentage approaches 100%. The Lord and Raju methods are more accurate for detecting Moderate DIF, as their detection rates are higher than those of Mantel-Haenszel. The logistic regression method does not detect Moderate DIF using the Zumbo and Thomas criteria or Jodoin and Gierl criteria. It is also observed that the Mantel-Haenszel, Lord, and Raju methods accurately detect Large DIF with any sample size. The logistic regression method only detected it in 42% of the simulations with the Jodoin and Gierl criteria and 400 participants.

**Graph I.** Percentage of correct DIF detections for each method<sup>2</sup> with 10% DIF

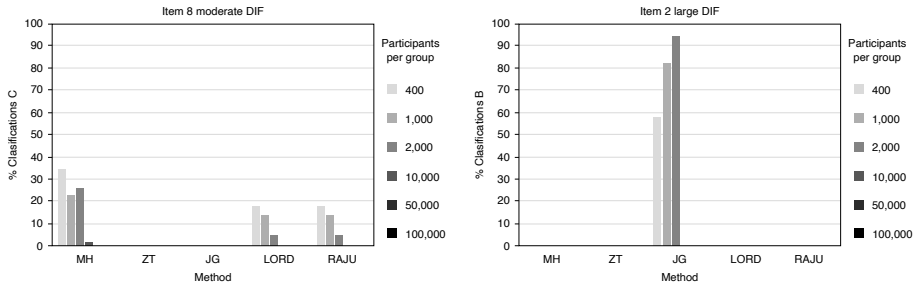


Source: Authors' own work

The following graph shows incorrect DIF detections, defined as cases in which methods classified items with DIF into a category other than the one in which they were generated. The Mantel–Haenszel, Lord, and Raju methods can classify Moderate DIF as Large with 10,000 participants or fewer, with overestimation being more pronounced for the Mantel–Haenszel method than for the IRT-based methods, particularly in samples of 400 examinees. The logistic regression method classifies Large DIF as Moderate when using the Jodoin and Gierl criteria and for sample sizes up to 2,000 participants.

<sup>2</sup> MH: Mantel-Haenszel; ZT: logistic regression, Zumbo and Thomas criteria; JG: logistic regression, Jodoin and Gierl criteria.

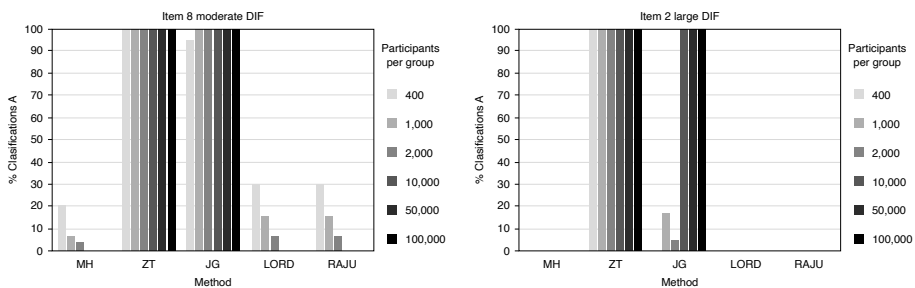
**Graph II.** Percentage of incorrect DIF detections with 10% DIF



Source: Authors' own work

Graph III shows DIF non-detections, where non-detection rate of Moderate DIF using the Mantel-Haenszel method is lower than the Lord and Raju methods', which reach a maximum of 30% in small samples. This percentage decreases as the number of participants increases, becoming zero from 10,000 onwards. The logistic method does not detect such DIF with any criterion. The same graph shows there are no Large DIF non-detections with the Mantel-Haenszel, Lord, and Raju methods and confirms that the logistic regression method never detects it using the Zumbo and Thomas criteria, and under the Jodoin and Gierl criteria, detection does not occur with over 10,000 participants.

**Graph III.** Percentage of DIF non-detections with 10% DIF

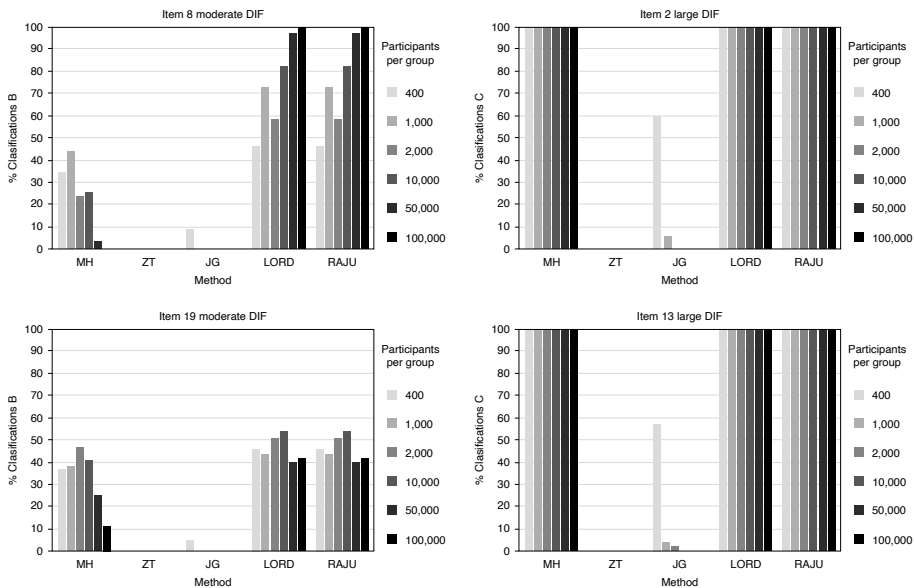


Source: Authors' own work

The percentage of false DIF detections was calculated for the 18 items whose parameters were identical between the comparison groups. There were only wrong detections with the Mantel-Haenszel, Lord, and Raju methods in less than 1.90% of these items in the simulations with 400 participants and in less than 0.07% with the IRT methods in simulations with 1,000 participants.

The effectiveness of the same methods was later verified with 20% of the items with DIF: two with Moderate DIF and two with Large DIF. The following graph shows the percentage of correct detections. Having more items with DIF altered the effectiveness of the methods for detecting Moderate DIF. When comparing the figures on the left of graphs I and IV (first row), which correspond to the same item, it is observed that the rates of correct detections of the Mantel-Haenszel, Lord, and Raju methods decreased, however, the IRT-based methods remain more effective than the Mantel-Haenszel method. For the latter, the effect of population size is reversed compared to the previous case: the number of correct detections decreases as the number of participants increases. With the Lord and Raju methods, the former effect remains: their accuracy increases as the number of participants grows, although both methods are accurate only with 100,000 participants. The logistic regression method does not detect Moderate DIF with any criteria.

**Graph IV.** Percentage of correct DIF detections with 20% DIF



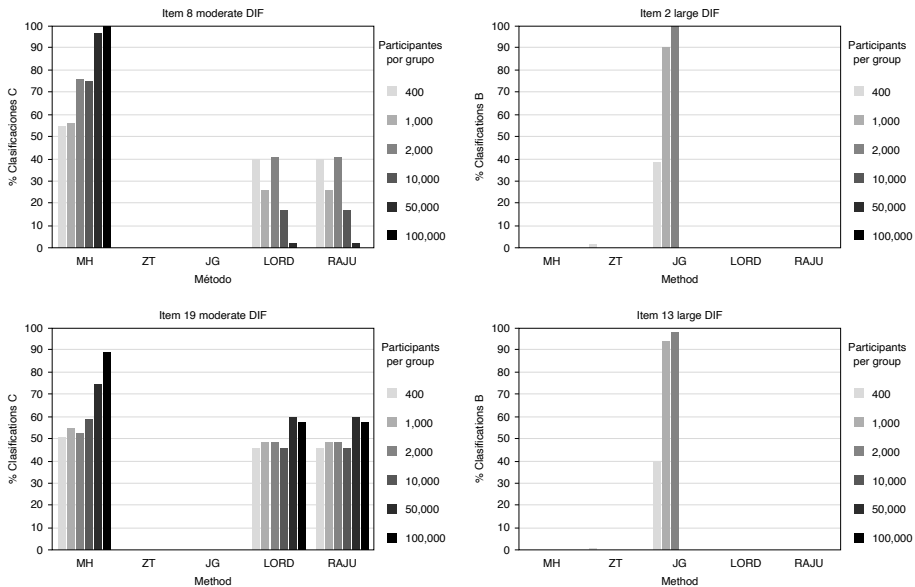
Source: Authors' own work

The figures on the right in graphs I and IV (first row) correspond to the same item and show that the effectiveness of detecting Large DIF was not altered by adding items with DIF: the same three methods continue to detect it in all instances, while the logistic regression method detected it in up to 61% of cases with 400 participants under the Jodoin and Gierl criteria.

The figures in the second row of Graph IV show correct DIF detections in the items that were additionally manipulated for this condition. On the left, the percentage of Moderate DIF detection is less than 50% with the Mantel-Haenszel, Lord, and Raju methods, and none of them show an effect of population size. It should be noted that this item (item 19 in Table III) was originally a difficult one, but with the manipulation to produce DIF, it became the most difficult with DIF and the second most difficult in this version, so its extreme difficulty may have affected DIF detection effectiveness. The graph on the right shows that the previous results are replicated in Large DIF: the Mantel-Haenszel, Lord, and Raju methods have a detection rate of 100%, while the logistic regression method detected it in nearly 60% of the simulations with 400 participants under the Jodoin and Gierl criteria.

Graph V shows incorrect DIF detections, where it can be seen again on the left side that when the Mantel-Haenszel, Lord, and Raju methods fail to correctly detect Moderate DIF, they classify it as Large. For the former, this overestimation increases as the number of participants increases. The logistic regression method never classifies Moderate DIF as Large. The graphs on the right show that this method classifies both items with Large DIF as Moderate only under the Jodoin and Gierl criteria, with populations of 400 to 2,000 participants.

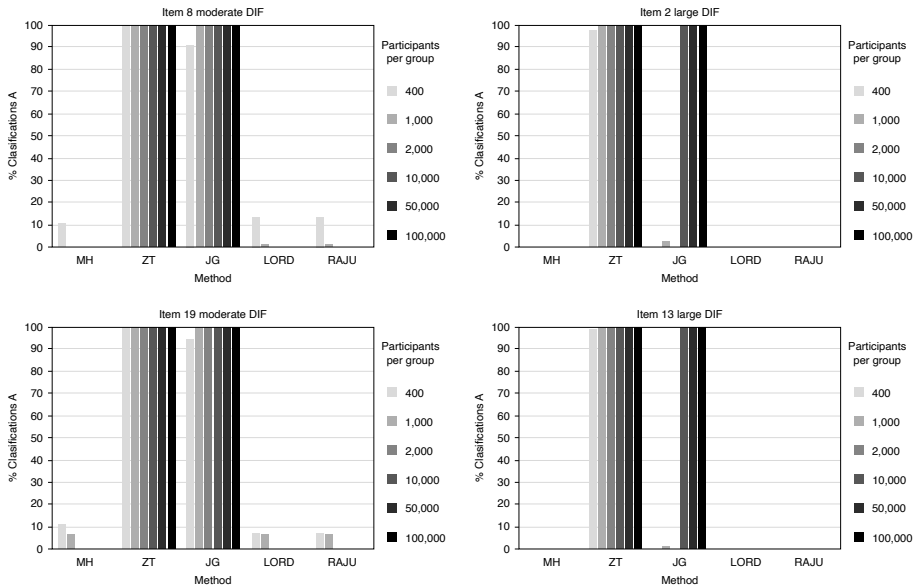
**Graph V.** Percentage of incorrect DIF detections with 20% DIF



Source: Authors' own work

Regarding non-detection rates, the right side of the following graph shows that the Mantel-Haenszel, Lord, and Raju methods fail to detect Moderate DIF with 400 or 1,000 participants, whereas the logistic regression method classifies these items as having no DIF under both criteria. The same pattern occurs for Large DIF items, except in cases where logistic regression classifies them as having Moderate DIF.

**Graph VI.** Percentage of DIF non-detections with 20% DIF



Source: Authors' own work

Finally, the percentage of incorrect DIF detections for items with equal parameters in both comparison groups decreased as the number of DIF items in the test increased. In simulations with 400 participants, this percentage was 1.63% for the Mantel–Haenszel, Lord, and Raju methods, and it dropped to zero under all other conditions using the logistic regression method.

## Conclusions

This study evaluated the effectiveness of four methods for detecting uniform differential functioning in populations five to fifty times larger than those analyzed in previous simulation studies and those that employed empirical data to confirm their applicability to complete populations in some ILSAs or other large-scale assessments.

The results show that the logistic regression method is not appropriate for analyzing DIF in large-scale assessment items, since it only detected it using the Jodoin and Gierl criteria with up to 2,000 participants. These results are consistent with previous reports indicating that this method is effective for groups of up to 1,000 test takers (French & Maller, 2007; Jodoin & Gierl, 2001).

Additionally, it was found that when using 20 items, with 10% exhibiting DIF (one item with Moderate DIF and one with Large DIF), increasing the number of participants per group improved the detection of Moderate DIF with the Mantel–Haenszel, Lord, and Raju methods. The Lord and Raju methods were more accurate with up to 10,000 participants per group, whereas the Mantel–Haenszel method tended to overestimate DIF or classify it as Large. It was also found that these three methods are 100% accurate in detecting Moderate DIF starting at 50,000 examinees per group, although this accuracy decreased considerably when the percentage of items with DIF increased to 20%. It should be noted that none of the three methods matched in their classification in all Moderate DIF simulations. For example, in 70 simulations of groups with 1,000 participants, all three methods detected it correctly, however, in 18 of these simulations, their classification differed.

The Mantel-Haenszel, Lord, and Raju methods were accurate in detecting Large DIF starting at 400 participants per group, regardless of the percentage of DIF in the area, the number of examinees, the item difficulty, or whether it benefited the focal group or the reference group. It is also confirmed that the results obtained with the Lord and Raju methods are identical with large sample sizes, for both Moderate and Large DIF.

Finally, it was observed that across comparison groups ranging from 400 to 100,000 participants, the Mantel–Haenszel method did not produce increased Type I errors; that is, it did not detect DIF when none was present. In contrast, the IRT-based methods showed minimal incorrect detections, which occurred only in groups of 1,000 participants or fewer. Overall, false positives for all three methods in items without DIF are minimal and absent in large participant samples.

Although no method was completely accurate in detecting uniform DIF, the Mantel-Haenszel, Lord, and Raju methods were the most effective. While their classification was not always accurate, differential functioning did not go unnoticed, and when the methods did not classify it correctly, they overestimated DIF. Therefore, for studies on large-scale assessment biases that aim to compare entire populations rather than samples, the Mantel-Haenszel, Lord, or Raju methods should be used for

analyzing DIF in dichotomous items in tests of approximately 20 items, with participant numbers ranging from 800 to 200,000, or comparison groups of 400 to 100,000 participants each. To maximize information, multiple methods can be applied (Özdemir, 2015); however, since the Lord and Raju methods produce the same results with these sample sizes, as they use the same criteria, and considering the latter involves more complex calculations, it may be sufficient to use the Mantel-Haenszel and Lord methods together. Based on the resulting classifications, items flagged as having Large DIF by both methods can be identified for qualitative analysis to determine potential bias and guide appropriate actions.

Using these methods to analyze DIF in such large populations could enable research using all available information, without the need to draw samples from the groups evaluated in ILSAs or to compare participants from countries, geographical regions, or economic areas who took the test in the same language. Similarly, studies could be conducted with entire populations of EXANI administrations or other nationwide assessments. In all these cases, it would be a first step toward ensuring that these high-impact assessments are free of bias.

This study provides insight into the use of three traditional DIF methods in large samples of participants responding large-scale assessment instruments. However, since sample size is only one of the variables considered, further research is needed to examine whether other variables—such as different distributions, varying sizes of the comparison groups, larger numbers of test items, higher percentages of DIF items, and purification methods—affect the effectiveness of these methods. In addition, new methods for detecting differential functioning continue to emerge (Halpin, 2024; Liu et al., 2019; Suk & Han, 2024; Villalonga-Olives et al., 2024), and it is necessary to evaluate their effectiveness in participant samples comparable to those in this study. Finally, the effectiveness of these methods should be substantiated with empirical data, for example, studies in which DIF has been detected in ILSA items could be revisited to perform analyses using the entire population.

## Bibliography

- Abulela, M. A. A., & Rios, J. A. (2022). Comparing the robustness of three nonparametric DIF procedures to differential rapid guessing. *Applied Measurement in Education*, 35(2), 81–94. <https://doi.org/10.1080/08957347.2022.2067542>
- American Educational Research Association, American Psychological Association, & National Council on Measurement in Education. (Eds.).

- (2014). *Standards for educational and psychological testing*. American Educational Research Association.
- Cassiani-Miranda, C. A., Scopetta, O., Barrios-Villadiego, M. A., Tirado-Otálvaro, A. F., & Duran-Bedoya, A. C. (2025). Construct validity of the PHQ-9 in university students in Colombia: A Rasch analysis approach. *Medicina Clínica y Social*, 9(1), e521. <https://doi.org/10.52379/mcs.v9.521>
- Centro Nacional de Evaluación para la Educación Superior. (2009). *Procedimientos Básicos para el Análisis de Reactivos*.
- Cohen, A. S., & Kim, S. H. (1993). A comparison of Lord's  $X^2$  and Raju's area measures in detection of DIF. *Applied Psychological Measurement*, 17(1), 39-52. <https://doi.org/10.1177/014662169301700109>
- Demirus, K. B., & Pektaş, S. (2022). Investigation of TIMSS 2015 science test items in terms of differential item functioning according to language and culture. *International Journal of Education Technology and Scientific Researches*, 7(18), 1166-1178. <http://dx.doi.org/10.35826/ijetsar.499>
- Dorans, N. J., & Holland, P. W. (1992). DIF detection and description: Mantel-Haenszel and standardization. *ETS Research Report Series*, 1992(1), i-40. <https://doi.org/10.1002/j.2333-8504.1992.tb01440.x>
- Fidalgo, A. M., Mellenbergh, G. J., & Muñiz, J. (2000). Effects of amount of DIF, test length, and purification type on robustness and power of Mantel-Haenszel procedures. *Methods of Psychological Research Online*, 5(3), 43-53.
- Fishbein, B., Yin, L., & Foy, P. (2024). *PIRLS 2021 User Guide for the International Database* (2nd ed.). Boston College, TIMSS & PIRLS International Study Center. <https://pirls2021.org/data>
- Fraillon, J. (2024). *An international perspective on digital literacy: Results from ICILS 2023*. International Association for the Evaluation of Educational Achievement. [https://www.iea.nl/sites/default/files/2024-11/ICILS\\_2023\\_International\\_Report\\_0.pdf](https://www.iea.nl/sites/default/files/2024-11/ICILS_2023_International_Report_0.pdf)
- French, B. F., & Maller, S. J. (2007). Iterative purification and effect size use with logistic regression for differential item functioning detection. *Educational and Psychological Measurement*, 67(3), 373-393. <https://doi.org/10.1177/0013164406294781>
- Gierl, M. J., Gotzmann, A., & Boughton, K. A. (2004). Performance of SIBTEST when the percentage of DIF items is large. *Applied Measurement in Education*, 17(3), 241-264. [https://doi.org/10.1207/s15324818ame1703\\_2](https://doi.org/10.1207/s15324818ame1703_2)
- Halpin, P. F. (2024). Differential item functioning via robust scaling. *Psychometrika*, 89(3), 796-821. <https://doi.org/10.1007/s11336-024-09957-6>

- Hidalgo-Montesinos, M. D., & Lopez-Pina, J. A. (2002). Two-stage equating in differential item functioning detection under the graded response model with the Raju area measures and the Lord statistic. *Educational and Psychological Measurement*, *62*(1), 32-44. <https://doi.org/10.1177/0013164402062001003>
- Hidalgo, M. D., & López-Pina, J. A. (2004). Differential item functioning detection and effect size: A comparison between logistic regression and Mantel-Haenszel procedures. *Educational and Psychological Measurement*, *64*(6), 903-915. <https://doi.org/10.1177/0013164403261769>
- Jodoin, M. G., & Gierl, M. J. (2001). Evaluating type I error and power rates using an effect size measure with the logistic regression procedure for DIF detection. *Applied Measurement in Education*, *14*(4), 329-349. [https://doi.org/10.1207/S15324818AME1404\\_2](https://doi.org/10.1207/S15324818AME1404_2)
- Kucam, E., & Gülleroğlu, H. D. (2023). Examination of differential item functioning in PISA 2018 Mathematics Literacy test with different methods. *Journal of Measurement and Evaluation in Education and Psychology*, *14*(2), 128-153. <https://doi.org/10.21031/epod.1122857>
- Liu, R., & Bradley, K. D. (2021). Differential item functioning among English language learners on a large-scale mathematics assessment. *Frontiers in Psychology*, *12*, 657335. <https://doi.org/10.3389/fpsyg.2021.657335>
- Liu, Y., Yin, H., Xin, T., Shao, L., & Yuan, L. (2019). A comparison of differential item functioning detection methods in cognitive diagnostic models. *Frontiers in Psychology*, *10*, 1137. <https://doi.org/10.3389/fpsyg.2019.01137>
- Magis, D., Béland, S., Tuerlinckx, F., & De Boeck, P. (2010). A general framework and an R package for the detection of dichotomous differential item functioning. *Behavior Research Methods*, *42*(3), 847-862. <https://doi.org/10.3758/BRM.42.3.847>
- Mellenbergh, G. J. (1982). Contingency table models for assessing item bias. *Journal of Educational Statistics*, *7*(2), 105-118. <https://doi.org/10.2307/1164960>
- Narayanan, P., & Swaminathan, H. (1994). Performance of the Mantel-Haenszel and simultaneous item bias procedures for detecting differential item functioning. *Applied Psychological Measurement*, *18*(4), 315-328. <https://doi.org/10.1177/014662169401800403>
- Narayanan, P., & Swaminathan, H. (1996). Identification of items that show nonuniform DIF. *Applied Psychological Measurement*, *20*(3), 257-274. <https://doi.org/10.1177/014662169602000306>
- Organisation for Economic Co-operation and Development. (2023). *PISA 2022 results (Volume I): The state of learning and equity in education*. <https://doi.org/10.1787/53f23881-en>

- Ohiri, S. C., Momoh, O. C., & Ikeanumba, C. B. (2024). Differential item functioning detection methods: An overview. *International Journal of Research Publication and Reviews*, 5(2), 1555-1564. <https://doi.org/10.55248/gengpi.5.0224.0505>
- Özdemir, B. (2015). A comparison of IRT-based methods for examining differential item functioning in TIMSS 2011 mathematics subtest. *Procedia-Social and Behavioral Sciences*, 174, 2075-2083. <https://doi.org/10.1016/j.sbspro.2015.02.004>
- Park, S. E., Ahn, S., & Zopluoglu, C. (2021). Differential item functioning effect size from the multigroup confirmatory factor analysis for a meta-analysis: A simulation study. *Educational and Psychological Measurement*, 81(1), 182-199. <https://doi.org/10.1177/0013164420925885>
- Penfield, R. D., & Camilli, G. (2007). *Differential item functioning and item bias*. In C. R. Rao & S. Sinharay (Eds.), *Handbook of Statistics: Vol. 26. Psychometrics* (pp. 125-167). Elsevier. [https://doi.org/10.1016/S0169-7161\(06\)26005-X](https://doi.org/10.1016/S0169-7161(06)26005-X)
- R Core Team. (2018). *R: A language and environment for statistical computing. R foundation for Statistical Computing*.
- Raju, N. S. (1990). Determining the significance of estimated signed and unsigned areas between two item response functions. *Applied Psychological Measurement*, 14(2), 197-207. <https://doi.org/10.1177/014662169001400208>
- Raju, N. S., Drasgow, F., & Slinde, J. A. (1993). An empirical comparison of the area methods, Lord's chi-square test, and the Mantel-Haenszel technique for assessing differential item functioning. *Educational and Psychological Measurement*, 53(2), 301-314. <https://doi.org/10.1177/0013164493053002001>
- Rogers, H. J., & Swaminathan, H. (1993). A comparison of logistic regression and Mantel-Haenszel procedures for detecting differential item functioning. *Applied Psychological Measurement*, 17(2), 105-116. <https://doi.org/10.1177/014662169301700201>
- Roussos, L. A., Schnipke, D. L., & Pashley, P. J. (1999). A generalized formula for the Mantel-Haenszel differential item functioning parameter. *Journal of Educational and Behavioral Statistics*, 24(3), 293-322. <https://doi.org/10.3102/10769986024003293>
- Shamsaldeen, F., Wang, J., & Ahn, S. (2024). Evaluating the fairness of a high-stakes college entrance exam in Kuwait. *Language Testing in Asia*, 14(1), 27. <https://doi.org/10.1186/s40468-024-00301-4>
- Schulz, W., Ainley, J., Fraillon, J., Losito, B., Agrusti, G., Damiani, V., & Friedman, T. (2025). *Education for citizenship in times of global chal-*

- lence: IEA international civic and citizenship education study 2022 international report. Springer. <https://doi.org/10.1007/978-3-031-65603-3>
- Sireci, S. G., & Rios, J. A. (2013). Decisions that make a difference in detecting: differential item functioning. *Educational Research and Evaluation, 19*(2-3), 170-187. <https://doi.org/10.1080/13803611.2013.767621>
- Suk, Y., & Han, K. T. (2024). A psychometric framework for evaluating fairness in algorithmic decision making: Differential algorithmic functioning. *Journal of Educational and Behavioral Statistics, 49*(2), 151-172. <https://doi.org/10.3102/10769986231171711>
- Swaminathan, H., & Rogers, H. J. (1990). Detecting differential item functioning using logistic regression procedures. *Journal of Educational Measurement, 27*(4), 361-370. <https://doi.org/10.1111/j.1745-3984.1990.tb00754.x>
- Universidad Autónoma de Madrid. (2024). *EvAU 2024: Presentación de resultados* [Presentación de PowerPoint]. <https://www.uam.es/uam/media/doc/1606958184663/rueda-de-prensa-uam.-datos-evau-2024-compressed.pdf>
- Uyar, S. (2020). Latent class approach to detect differential item functioning PISA 2015 science sample. *Eurasian Journal of Educational Research, 20*(88), 179-198. <https://ejer.com.tr/latent-class-approach-to-detect-differential-item-functioning-pisa-2015/>
- Velázquez Ramírez, S. (2022). Exámenes que contribuyen a los procesos de admisión. *Revista Ceneval Investiga, 4*, 22-35. <https://online.flippingbook.com/view/556549215/>
- Vidal Gonzáles, F. L., Pacheco Leyva, Y., & Galindo Cortés, G. (2022). Comparabilidad en las aplicaciones remotas y presenciales del EXANI-II. *Revista Ceneval Investiga, 6*, 28-37. <https://online.flippingbook.com/view/917214422/>
- Villalonga-Olives, E., Khademi, A., Pan, Y. Y., & Ransome, Y. (2024). Unveiling disparities: examining differential item functioning's impact on racial health equity among white and black populations. *Public Health, 231*, 80-87. <https://doi.org/10.1016/j.puhe.2024.03.006>
- von Davier, M., Kennedy, A., Reynolds, K., Fishbein, B., Khorramdel, L., Aldrich, C., Bookbinder, A., Bezirhan, U., & Yin, L. (2024). *TIMSS 2023 international results in Mathematics and Science*. Boston College, TIMSS & PIRLS International Study Center. <https://doi.org/10.6017/lse.tpisc.timss.rs6460>
- Walker, C. M. (2011). What's the DIF? Why differential item functioning analyses are an important part of instrument development and valida-

- tion. *Journal of Psychoeducational Assessment*, 29(4), 364-376. <https://doi.org/10.1177/0734282911406666>
- Webster, N. S., & Bowe, A. G. (2021). Examining measurement equivalency within the Latin American module of the International Civic Competence Study 2009. *Journal of Community Psychology*, 49(7), 2972-2982. <https://doi.org/10.1002/jcop.22515>
- Wang, W. C., Shih, C. L., & Yang, C. C. (2009). The MIMIC method with scale purification for detecting differential item functioning. *Educational and Psychological Measurement*, 69(5), 713-731. <https://doi.org/10.1177/0013164409332228>
- Wu, M., Tam, H. P., & Jen, T.-H. (2016). *Educational measurement for applied researchers*. Springer.
- Zieky, M. (2006). Fairness review in assessment. In S. M. Downing & T. M. Haladyna (Eds.), *Handbook of test development* (pp. 359-376). Lawrence Erlbaum Associates.
- Zumbo, B. D. (2007). Three generations of DIF analyses: Considering where it has been, where it is now, and where it is going. *Language Assessment Quarterly*, 4(2), 223-233. <https://doi.org/10.1080/15434300701375832>
- Zwick, R. (2012). A review of ETS differential item functioning assessment procedures: Flagging rules, minimum sample size requirements, and criterion refinement. *ETS Research Report Series*, 2012(1), i-30. <https://doi.org/10.1002/j.2333-8504.2012.tb02290.x>



# Mathematical problem-solving in primary education teacher training programmes in Spain<sup>1</sup>

## Resolución de problemas matemáticos en la formación del profesorado de educación primaria española

<https://doi.org/10.4438/1988-592X-RE-2025-412-734>

**Elena Castro-Rodríguez**

<https://orcid.org/0000-0002-2560-8982>

Universidad de Granada

**Juan Luis Piñeiro**

<https://orcid.org/0000-0002-9616-3925>

Universidad Metropolitana de Ciencias de la Educación, Chile

**Enrique Castro**

<https://orcid.org/0000-0002-2786-500X>

Full professor (retired). Universidad de Granada

### Abstract

**Introduction:** Primary education plans underscore the importance of imparting mathematical problem-solving (PS) skills, for which purpose teachers must receive appropriate training and their study plans made clear. In this paper, we consider the presence of problem-solving and the skills required to teach it in the teaching guides for compulsory and elective subjects within the Didactics of Mathematics content for primary education teacher-training degrees in Spain. **Method:** A content analysis and cluster analysis was made of 151 teaching guides at 39 Spanish public universities. **Results:** The study findings show that greater weight is attached to PS competency than to elements of specialised knowledge. Regarding PS knowledge, the topic most frequently referred to is

---

<sup>1</sup> This study is funded by the research projects PID2020-113601 GB-I00, PID2024-157106NB-I00; financed by MCIN/AEI/10.13039/501100011033, and FEDER, UE, and the research project FOVI240238 financed by Ministry of Science (Chile).

that of problem types, while in PS pedagogical knowledge, posing problems is the area most often mentioned. **Conclusions:** The study results suggest that each university interprets PS teaching and learning according to its own experience and understanding. Accordingly, teacher training in Spain in this field is highly variable. Aspects such as due problem solver's consideration and of non-cognitive factors should be addressed in order to improve teacher training in PS.

**Keywords:** Curriculum research, Initial training, Mathematics, Problem solving, Teacher, University studies.

## Resumen

**Introducción:** La relevancia de la resolución de problemas matemáticos queda manifestada en los documentos curriculares de Educación Primaria. Sin embargo, para que esto pueda llevarse a cabo es necesario formar a los docentes y conocer sus planes de estudio. Por ello, este trabajo describe la presencia de la resolución de problemas y su enseñanza en las guías docentes de las asignaturas obligatorias y optativas del área de Didáctica de la Matemática pertenecientes a los grados de Maestro en Educación Primaria en España. **Método:** Realizamos un análisis descriptivo y un análisis clúster de 151 guías docentes de asignaturas obligatorias y optativas del área de Didáctica de la Matemáticas de 39 universidades públicas. **Resultados:** Los resultados muestran un énfasis en la competencia para resolver problemas sobre elementos de conocimiento especializado. Asimismo, en relación al conocimiento sobre la resolución de problemas, aparece con mayor frecuencia las menciones a tipos de problemas, mientras que, respecto al conocimiento didáctico de la resolución de problemas, los tópicos relativos a la invención de problemas son los más frecuentes. **Conclusiones:** Concluimos que cada universidad ha interpretado la enseñanza y aprendizaje de la resolución de problemas desde su propio entendimiento por lo que la formación entre maestros españoles es muy heterogénea. Además, destacamos la necesidad de incluir aspectos como la consideración del resolutor o factores no cognitivos para mejorar la formación sobre resolución de problemas de los futuros maestros.

**Palabras clave:** estudios universitarios, formación inicial, investigación curricular, matemáticas, profesor, solución de problemas

## Introduction

Problem solving (PS)—i.e., performing a task for which the solver does not have an immediate, direct procedure, but for which a path is assumed to exist—is an important means of developing mathematical knowledge in primary school children (Britz & Richard, 1992; Schoenfeld, 2013). Solving meaningful problems contributes to the development of higher-order

thinking skills and the discovery of a repertoire of strategies that prepare students to solve new problems (Lester & Cai, 2016). Thus, children acquire the meaning of mathematical ideas via their active participation in resolving a variety of mathematical problems (Britz & Richard, 1992).

The importance of PS in early childhood is reflected in the current Spanish curriculum document for primary education (RDL 157/2022, of 1 March), which underscores the fundamental role played by this activity and as a pedagogical principle seeks to promote the integration of competencies by ensuring teaching time is dedicated to PS (RDL 157/2022, of 1 March). In this document, moreover, PS receives special attention, “considering it more as a methodological principle than as content in itself” (Contreras, p. 66, 2022).

However, for this recommendation to be implemented, future teachers must receive adequate training in the regard. The present paper contributes to this goal by focusing on two objectives. First, we describe the presence of subjects in the area of the Didactics of Mathematics in teacher-training degree courses for primary education in Spain; we then consider the presence and teaching of PS skills in the teaching guides published for these degrees.

Knowledge of related study plans is important because when curricular content and experiences are internally coherent, they are usually of more benefit to future teachers (Tatto, 2018). In this context, Hiebert et al. (2019) investigated whether novice teachers retained and made use of the mathematics they had studied during their initial training. Although results differed among the participants, they consistently performed better on the content covered in training than on content that had not been covered. In this respect, too, Morris and Hiebert (2017) examined whether the mathematics content of a primary school teacher-training programme had any subsequent effect on these teachers’ practices. The study results showed that when teachers had assimilated the training content, they tended to incorporate it into their teaching plans. Therefore, the evidence suggests that the content studied in training influences the specialised knowledge later used in teaching. Accordingly, knowledge and understanding of the teacher-training programmes of Spanish universities can be of great use both to the trainers involved and to curriculum designers.

## **Mathematical problem solving in teacher training**

Although researchers of mathematics education have focused on PS for over 50 years, its implementation remains a challenge for teachers (Liljedahl & Cai, 2021). It has been suggested that this may be related to

teachers' beliefs, and may be overcome by focusing on procedural aspects (Son & Lee, 2021). But whatever the cause of the problem, many teachers need support to implement PS in their classrooms, and one source of this support is ongoing professional development.

Felmer and Perdomo-Díaz (2017), Leung (2016) and Wake et al. (2016), among other experts in the field, have focused on professional development related to PS. For example, an analysis conducted in Chile considered an initiative termed *Problem-Solving Activation in the Classroom*, in which PS was introduced into the classrooms of primary and secondary school teachers (Felmer & Perdomo-Díaz, 2017). To this end, three practical workshops were held in various regions of the country, promoting collaborative, non-routine PS. The workshops provided opportunities for teachers to solve mathematics problems and to reflect on their characteristics. The study revealed a statistically significant improvement in problem-solving performance among students whose teachers had participated in the workshops.

On other occasions, professional development has been carried out via projects in which researchers held meetings with teachers to reflect on lesson planning and implementation (Hähkiöniemi & Francisco, 2019; Leung, 2016; Wake et al., 2016). For example, the *Informal Mathematical Learning* project in the United States and the *Inquiry-Based Mathematics Teaching* project in Finland each include a professional development component to help secondary school teachers guide students in PS. Lessons were recorded, enabling teachers and researchers to jointly reflect on the classroom performance observed. A significant conclusion drawn from these studies is that the practical implementation of PS sessions is sometimes hampered by time constraints. In a related project in the United Kingdom (Wake et al., 2016), researchers collaboratively planned PS sessions with secondary school teachers. Immediately after implementation, they discussed the teacher's interventions and considered how the prior activities had affected lesson planning and communication with students regarding their PS performance.

In other projects, problem formulation is included as an essential element. For example, Leung (2016), in a study carried out in Taiwan, focused on training primary school teachers in how to incorporate problem formulation into their classes. These researchers concluded that the learning process is influenced by how teachers organise tasks, how they interact with students and how children engage in group work (Leung, 2016).

Other analyses have focused on the professional development in PS that takes place in initial secondary-school teacher training (Gourdeau, 2019; Mellone et al., 2021). In this respect, Mellone et al. (2021) con-

sidered the changes observed in prospective teachers' conceptions and beliefs about PS after their participation in a Mathematics course. Using instruments such as surveys and interviews, these authors concluded that teachers' outlook on PS is conditioned by prior experiences and by the prevailing cultural context.

In another study conducted in an initial teacher-training setting, Gourdeau (2019) considered a course for future secondary school teachers in Canada. PS plays a central role in this course, as the goal is for future teachers to experience PS at a profound level, acquiring a good understanding of the processes involved. To achieve this, in addition to fostering motivation and the personal ambition to solve the problem presented, the course considers five aspects: understanding or trying to understand processes and objects, extracting and defining characteristics, finding patterns and structures, representing, and communicating our understanding.

In Spain, the design of the initial training plan for primary education teachers is reflected in various official documents at the national, institutional and practitioner levels, applying different levels of specification (Rico et al., 2014). At the national level, a fixed number of credits are awarded for the teacher-training degree, both in total and their distribution among the constituent modules; thus, the didactic-disciplinary training module, which covers the teaching and learning of mathematics, social sciences, experimental sciences, languages, physical education, music, plastic arts and visual arts, is awarded 100 ECTS credits (ECI/3857/2007 Order, of 27 December).

At the institutional level, each university applies the national guidelines in accordance with the design and structure of its curriculum, and does the same regarding the descriptors and credits for each subject, thereby establishing both the number of course subjects and their design.

At the most direct level of application, the university departments develop their subject programmes (termed "teaching guides") in accordance with the provisions of the national and institutional documents. These teaching guides contain general information (number of ECTS credits, semester, department, nature, faculty responsible and bibliography), together with aspects related to the planning of the subject, such as the curricular levels of learning expectations (establishing learning priorities), teaching content (items by which knowledge is organised in terms of concepts, procedures and/or attitudes), teaching methods (strategies, actions, techniques, tasks and/or materials, jointly reflecting how the teaching will be carried out) and evaluation processes (guidelines, systems or techniques to identify what students have learned). These

concepts, procedures and attitudes coincide with the basic curricular levels proposed by Tyler (1986). In the present analysis, we consider this third level of specification.

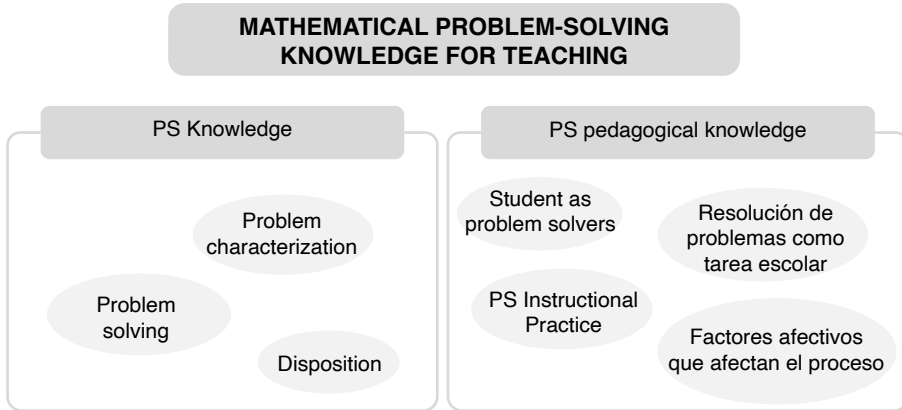
Some universities include courses in initial primary school teacher-training aimed at developing teachers' professional knowledge about PS, including instruction in strategies or heuristics, problem types and metacognitive questions. However, to our knowledge no previous studies have focused on PS teaching in the context of the current plan for initial teacher-training programmes in Spain.

## **Professional knowledge to teach problem solving**

This paper takes a category-based perspective to the question of knowledge, in which two theoretical constructs are used to describe the knowledge needed to teach PS in primary education: PS competence and the didactic triangle. The latter construct is used for its potential to isolate and clarify particular aspects of teaching (Schoenfeld, 2013). In addition, we draw on two perspectives that allow us to further specify the categories. The first is the PS-teaching framework *Mathematical Problem-Solving Knowledge for Teaching* (Chapman, 2015). The second aspect is the curriculum guidelines issued by various countries (Piñeiro et al., 2022), which describe and define teachers' knowledge (Ball et al., 2008). In short, we present a proposal for teaching and learning PS that highlights elements of content knowledge and PS-related pedagogical knowledge, taking into consideration what teachers are required to impart on the subject (Figure 1).

In the model presented, we distinguish between PS competency and professional (or theoretical) knowledge in this respect. PS competency is the ability to formulate, represent and resolve mathematical problems, the existence of productive disposition (including beliefs), and the use of adaptive reasoning, given the capacity for logical thinking and reflection (Chapman, 2015). Professional knowledge, on the other hand, is the expertise necessary to perform the task in question, and in traditional teachers' knowledge models is usually divided into content knowledge and pedagogical content knowledge.

**Figure 1.** Components of professional knowledge for teaching PS



Source: Piñeiro et al. (2019)

The first component of this model, PS knowledge, incorporates three main elements: characterization of the problem, its PS process and the student's disposition.

The problem characterization category refers to the knowledge necessary to recognise whether or not a task constitutes a problem, i.e., whether this understanding depends on the solver (problem solver's consideration); it is assumed that the solver does not have a predetermined or direct procedure to address the task (i.e. task with no known solving procedure); it is also assumed that consideration of the task as a problem depends on its structure, i.e., the task-type is prosed as a problem.

Knowledge related to the PS process category includes the following theoretical aspects: a) PS stages and their characterization, i.e., understanding the problem, establishing a plan, executing it and evaluating the outcome (Polya, 1981); b) Strategies in PS (heuristic strategies); c) knowledge of the ways in which solvers self-regulate, monitor and control heuristics, specific strategies and mathematical knowledge, enabling them to make sound decisions (metacognition in PS); d) knowledge about the influence of the affective domain in PS, since (depending on the challenge posed) emotions may come into play, thus mobilising the intellect (non-cognitive factors in PS).

Regarding the student's disposition, the would-be solver must be willing to accept the PS challenge. Motivation to resolve a problem successfully is not essential, but there must be a commitment to address the task.

The second component, PS pedagogical knowledge, primarily concerns the didactic triangle and its interactions, where four elements can

be distinguished: a) Knowledge about the student as a problem solver, i.e. interaction between the student and the PS task requires the teacher to be aware of what can be done, what might interfere and how far the student's PS skills can be developed; b) PS as a worth-while task, i.e., interaction between the teacher and PS corresponds to knowledge about the task that is identified as a problem and about the process whereby a solution can be obtained. In other words, knowledge about problem selection, PS strategies and models, and the benefits of devising mathematical problems (where and when to do so); c) Knowledge of non-cognitive factors affecting PS, i.e., interaction between teacher and student in terms of each other's non-cognitive characteristics (emotions, attitudes and beliefs) regarding PS. In this regard, we should propose situations that foster beliefs related to the time required to solve a problem, and which offer multiple solution options, or where diverse answers are possible. In addition, we should be aware that anxiety and frustration are normal emotions in the PS process, as are the enjoyment and satisfaction when success is achieved. Finally, d) Knowledge of PS instructional practice, i.e., the threefold relationship between teacher, student and PS. It includes knowledge of teaching approaches to PS, discourse on PS, stuck state in PS, assessment in PS and resources on PS, both tangible and intangible.

## **Method**

In the qualitative approach used in this study, a document analysis (Bowen, 2009) was conducted of all the teaching guides for Didactics of Mathematics courses included in the university undergraduate degrees currently provided at the 39 Spanish public universities that qualify students for the profession of primary education teacher. We identified 151 course subjects, of which 114 were basic or compulsory and 37 were elective. All these guides referred to the 2022-2023 academic year, and, as public documents, were all readily accessible. The units of analysis were the excerpts from the guides mentioning PS, within the sections on objectives, competencies, content, methodology and/or assessment.

## **Categories**

In this study, the categories considered for document analysis concern two broad dimensions: the knowledge needed to teach PS and the curricular levels concerned. These dimensions were obtained from the theoretical framework presented above. In the first dimension, two sub-categories were stipulated: "Theme," which includes text fragments that only contain the term "Problem-solving" without further specification; and "PS

Competency,” coding fragments that mentioned the PS proficiency of future teachers.

The dimension termed Curricular Levels refers to the structure of the teaching guides and includes the categories considered fundamental in course planning and in the curriculum (Rico, 2013; Tyler, 1986), namely *learning expectations* (in the teaching guides, these are expressed through competencies, learning objectives and/or learning outcomes), *content* (described in the syllabus), *methodology* and *assessment*. All of these categories are represented in Table I.

**Table I.** Categories of analysis

<b>Dimension 1: Professional knowledge for teaching PS</b>		
PS competency		
	Problem characterization	Task with no known solving procedure Problem solver's consideration Types of tasks posed as problems
PS knowledge	PS process	PS phases and their characterisation Strategies in PS Metacognition in PS Non-cognitive factors in PS
	Disposition	Acceptance to accept the PS challenge
PS pedagogical knowledge	Student as a problem solver	Characteristics of successful problem solvers Difficulties and mistakes
	PS as a worthwhile task	Problem selection PS models and strategies Problem posing
	Non-cognitive factors	Knowledge of non-cognitive factors affecting PS
	PS instructional practice	Teaching approaches to PS s Discourse on PS Stuck State in PS Assessment in PS Resources on PS
Theme		
<b>Dimension 2: Curricular levels</b>		
Learning expectations		
Content		
Method		
Evaluation		

Source: The authors

These categories, which are deductive in nature and formulated taking into account previous theoretical contributions (Fraenkel et al., 2011), were evaluated by three experts with extensive professional experience in this field of research. All agreed that no modifications were necessary.

### **Data analysis**

The study data were analysed in two phases. First, the documents were coded in accordance with the pre-established categories for analysis, using the MAXQDA tool. For example, the objective “Acquire, understand and apply criteria to devise mathematics problems for primary education” was coded within the *Learning expectations* category in the Curricular Levels dimension, and within the Knowledge for teaching PS dimension in the category *Pedagogical PS Knowledge - PS as a worth-while task - problem posing*. This process was not exclusive because a given text could refer to different knowledge items according to our categories. In this phase of the research, two researchers working independently coded 5% of the documents, producing a kappa index value of 0.944 and a 95% agreement rate. Any disagreements were reviewed by the researchers and a consensus reached.

In the second phase, three calculations were performed: descriptive statistics for the number of compulsory and elective subjects, and for the credits awarded in each case; frequency tables and descriptive statistics for the categories in each area; and concurrency tables and matrices of relationships between the Professional Knowledge and Curricular Levels components. A code relationship matrix is a visual representation of the intersections of two codes in a single statement (Figure II). Finally, a cluster analysis was performed, considering the components of Professional Knowledge as variables, to obtain groups of documents. In this analysis, the similarity method based on the simple coincidence coefficient was used. This method considers the occurrence of codes in the document, taking into account only whether a code has been assigned to a document, and defines the similarity between the attributes considered as the number of coincidences divided by the total number of attributes.

### **Results**

The study results are presented in three sections. First, we describe the general situation of the Didactics of Mathematics subjects in the degree courses considered. We then focus on the presence of professional knowledge components for teaching PS. Finally, we present the results of the cluster analysis and describe the characteristics of the document groups thus generated.

## Subjects within Didactics of Mathematics in primary education degree courses

As can be seen in Table II, the median number of core or compulsory subjects in primary education degree courses is three. Six institutions require only two subjects (e.g., the universities of Murcia and Vigo), and none require more than four (only in the University of Huelva and two others are four subjects obligatory). The median number of total credits is 18, and the minimum, 12 (the latter is the case at the universities of León and Vigo). The maximum is 27 (e.g., at the University of Cádiz).

**Table II.** Descriptive statistics of the number of compulsory and elective subjects and of credits awarded for the university degree of Primary Education Teaching in the area of Didactics of Mathematics

	Subjects						Credits					
	N	Min	Max	Me	$\bar{x}$	$\sigma$	N	Min	Max	Me	$\bar{x}$	$\sigma$
Compulsory	114	2	4	3	2,9	0,6	724	12	27	18	18,5	2,368
Elective	37	0	4	1	0,9	1,17	197	0	24	3	4,9	6,3

Note: N= total; Min=minimum; Max=maximum; Me=median;  $\bar{x}$ =mean;  $\sigma$ =standard deviation.  
Source: The authors

The number of credits for the compulsory subjects ranges from three to twelve, but in most cases (82.5%) six credits are awarded (see Table III).

**Table III.** Frequency of credits in the compulsory and elective subjects for the university degree of Primary Education Teaching in the area of Didactics of Mathematics

Compulsory			Elective		
ECTS (n)	Frequency	%	ECTS (n)	Frequency	%
3	1	0,9	3	7	18,9
5	3	2,6	4	2	5,4
6	94	82,5	4,5	2	5,4
7	2	1,8	6	25	67,6
8	1	0,9	9	1	2,7
9	12	10,5			
12	1	0,9			
<b>TOTAL</b>	<b>114</b>	<b>100,0</b>	<b>TOTAL</b>	<b>37</b>	<b>100</b>

Source: The authors

As can be seen in Table IV, these subjects are taught during the first, second, third and/or fourth years of the course, but mainly in the second and third years.

**Table IV.** Frequency of the course year in which compulsory and elective subjects are taught for the university degree of Primary Education Teaching in the area of Didactics of Mathematics

	Compulsory		Elective	
	Frequency	%	Frequency	%
1	20	17,5	4	10,8
2	41	36	2	5,4
3	39	34,2	9	24,3
4	14	12,3	22	59,5
<b>TOTAL</b>	<b>114</b>	<b>100</b>	<b>37</b>	<b>100</b>

Source: The authors

Twenty of the 39 universities offer one or more elective subjects within the Didactics of Mathematics, with a median number of one such subject (Table II). Three institutions (e.g., the University of Valencia) offer the highest number, four, of elective subjects, while nineteen offer none (e.g., the Universities of Alcalá and Burgos). The total number of ECTS credits per institution varies greatly, ranging from 0 to 24, with a median number of three (Table II). Two thirds of the universities typically award six ECTS credits for each subject (Table III), which is most commonly studied in the fourth year (Table IV).

In general, the subject matter in these courses is organised according to two forms or models. In the first, subjects are primarily focused on developing mathematical content knowledge and on developing pedagogical content knowledge. Such is the case of the University of Sevilla, which offers two compulsory subjects: Specific Mathematics for Teachers and the Didactics of Mathematics for Teachers. The second model combines both types of knowledge in all subjects, which are separated according to the mathematical content covered. For example, the University of Las Palmas de Gran Canaria has three compulsory subjects: Mathematics and its Didactics I (Numbers and Algebra), Mathematics and its Didactics II (Measurement and Geometry) and Mathematics and its Didactics III (Statistics and Probability).

## Teaching guide components of knowledge for teaching PS

Regarding the number of occurrences of each of the knowledge categories (see Table V), most teaching guides make some mention of PS; for only 5% of compulsory subjects and 16% of elective subjects is it not mentioned at all. However, there is only one PS-focused compulsory subject (at the University of Almería) and three elective ones (at the Autonomous University of Madrid and at the Universities of Coruña and Girona), and in the latter case it is no longer offered.

Regarding different types of knowledge, there are no major differences in their presence in the teaching guides for compulsory subjects. There are slightly more references to PS competency, with 37% of coded statements, followed by PS knowledge, with 33.6%, and PS pedagogical knowledge, with 28.2%. Among the elective subjects, the distribution varies more, with higher percentages of coded statements relating to PS pedagogical knowledge and to PS competency (Table V).

**Table V.** Frequency of knowledge components in the teaching guides for teaching PS, for compulsory and elective subjects

	Compulsory		Elective	
	Documents N=114	Statements coded N=1092	Documents N=37	Statements coded N=160
PS competency	103 (90,4%)	412 (37,7%)	26 (70,3%)	60 (37,5%)
PS knowledge	100 (87,8%)	367 (33,6%)	17 (45,9%)	31 (19,4%)
PS pedagogical knowledge	103 (90,4%)	308 (28,2%)	23 (62,2%)	66 (41,3%)
Theme	5 (4,4%)	5 (0,5%)	3 (8%, 1)	3 (1,9%)
Coded documents	108 (95%)		31 (83,8%)	
Non-coded documents	6 (5%)		6 (16,2%)	

Source: The authors

Regarding the presence of PS in the curricular levels, the categories (PS competency and Professional knowledge, i.e. PS knowledge and PS pedagogical knowledge<sup>2</sup>) are mainly located in the first two levels (learning expectations and contents), while PS competency is the most important component in the levels of learning expectations and methodology (see Figure II). The national guidelines (Order ECI/3857/2007, of 27 De-

<sup>2</sup> In this Figure, the sub-categories have been combined.

ember) state that a general competency in the area of Mathematics is that of “Posing and solving problems related to everyday life” (p. 53759). Thus, PS competency mainly corresponds to learning expectations, although in some cases it is particularised to a specific field (such as numbers, measurement and geometry) or without mentioning everyday life (e.g., “Being able to solve problems that involve the connection between different areas of subject content”).

Regarding content, the teaching guides refer to various theoretical and practical aspects of PS. In some cases, a subject termed “Problem Solving” is mentioned, but no further details are given.

In general, the methodology and assessment levels focus on describing organisational aspects of the subject, which explains the relatively meagre presence of knowledge components. Specifically, when PS is included in the methodology level, it usually takes the form of practical sessions, as a development of the theoretical sessions. In other cases, a problem-based learning method is used in combination with other approaches.

Finally, the documents that mention PS in the assessment section include it as part of the written tests in which students are required to solve problems, for example, the “End-of-Course Conceptual Questions, Exercises and Problems” exam, which is included in the subject *Mathematics in Primary Education* at the University of the Balearic Islands. PS is also among the assessable activities performed during the course.

**Figure II.** Matrix of relationships between components and curricular levels.

Code System	Learning expectations	Contents	Methodology	Assesment
PS competency	220	55	126	63
PS knowledge	156	177	49	16
PS pedagogical knowledge	209	102	39	14

Source: The authors

### PS Knowledge

Regarding the categories of PS knowledge, Table VI illustrates a similar balance between compulsory and elective subjects. Among the teaching guides analysed, the category most frequently observed is that of types of tasks (for example, the University of Granada teaching guide for the subject Foundations of Mathematics for Primary Education includes *Types of arithmetic problems*). Present to a lesser extent are the categories of tasks with no known solving procedure (for example, the PS topic of the subject Mathematics and its Didactics I, at the University of the Basque

Country, includes “Distinguishing problems from exercises”), PS stages (for example, “Phases in the resolution of a verbal arithmetic problem” within the subject PS and Mathematical Connections at the University of Almería) and strategies in PS (for example, Heuristic strategies within the PS theme of the subject Reasoning and Mathematical Activity in Primary Education, at the University of Barcelona). Indeed, 95% of teaching guides for compulsory subjects include different types of tasks that can be posed as problems. In addition, about 26% of the documents refer to the notion of a problem as a task with no known solving procedure, to solution phases and to strategies. There is scant presence (less than 3%) of the remaining categories in these subjects.

Regarding elective subjects, the pattern is similar, but with a significantly lower presence in relation to PS knowledge. The most common category continues to be type of tasks posed as problems, present in 35% of documents, followed by tasks with no known solving procedure (13.5%), solution phases (10.8%), and strategies (8.1%). The remaining categories are not present in any teaching guide for elective subjects.

**Table VI.** Frequency of components within PS knowledge

		Compulsory		Elective	
		Documents N=114	Statements coded N=1092	Documents N=37	Statements coded N=160
Problem characteri- zation	Task with no known solving procedure	31 (27,2%)	48	5 (13,5%)	5
	Problem solver's consideration	2 (1,8%)	3	0	0
	Types of tasks posed as problems	95 (83,3%)	258	13 (35,1%)	16
PS process	PS phases and their characterization	29 (25,4%)	46	4 (10,8%)	6
	Strategies in PS	30 (26,3%)	57	3 (8,1%)	4
	Metacognition in PS	3 (2,6%)	6	0	0
	Non-cognitive factors in PS	3 (2,6%)	5	0	0
Disposition	Acceptance of the PS challenge	2 (1,8%)	2	0	0

Source: The authors

### PS Pedagogical Knowledge

The category of PS pedagogical knowledge most frequently found in the teaching guides is that of problem posing (Table VII). This presence is to be expected since, as observed above, according to the national guidelines, one of the general competencies required in the area of Mathematics is “Posing and solving problems related to everyday life” (Order ECI/3857/2007, 27 December, p. 53759). However, it is only found in 66.7% of the guides for compulsory subjects.

The second most prevalent category in both subject groups is that of discourse on PS, for example the core competency “Students should be able to convey information, ideas, problems and solutions to both specialised and non-specialised audiences”, within the Mathematics and its Didactics III course at the University of Salamanca.

The documents for the compulsory subjects also mention the questions of problem selection (for example, the learning outcome “Extracting problems from everyday life to pose learning situations” within the subject Mathematical Knowledge in Primary Education II, at the University of Cádiz), of PS models and strategies (for example, the content “Analysis of PS strategies” within Mathematics and its Didactics I, at the University of La Rioja) and of teaching approaches to PS (such as the content “Guidelines for the application of knowledge and PS in the Primary Education classroom” within the subject Mathematics and its Didactics I, at the Autonomous University of Madrid).

**Table VII.** Frequency of components within PS pedagogical knowledge

		Compulsory		Elective	
		Documents	Statements coded	Documents	Statements coded
Student as problem solver	Characteristics of successful solvers	5 (4,4%)	5	4 (10,8%)	5
	Difficulties and mistakes	10 (8,8%)	15	3 (8,1%)	4
PS as a worthwhile task	Problem selection	28 (24,6%)	42	7 (18,9%)	8
	PS models and strategies	22 (19,3%)	54	3 (8,1%)	6
	Problem posing	76 (66,7%)	125	9 (24,3%)	18

		Compulsory		Elective	
		Documents	Statements coded	Documents	Statements coded
Non-cognitive factors	Knowledge of non-cognitive factors affecting PS	6 (5,3%)	6	0	0
PS instructional practice	Teaching approaches to PS	19 (16,7%)	34	8 (21,6%)	15
	Discourse on PS	48 (42,1%)	71	9 (24,3%)	15
	Stuck satet in PS	0	0	2 (5,4%)	3
	Assessment in PS	1 (0,9%)	2	1 (2,7%)	1
	Resources on PS	11 (9,6%)	14	8 (21,6%)	9

Source: The authors

### Teaching guides and knowledge components

As remarked above, most of the documents make some reference to PS. However, they do not mention each of the components, and so we performed a cluster analysis to obtain groupings of teaching guides that present similarities. Since the MAXQDA tool requires the method and number of clusters to be specified a priori, that is, it uses non-hierarchical methods, the analysis was performed by increasing the number of clusters until new groupings were produced by dividing the minority groups. This procedure led us to identify the clusters presented in Tables VIII and IX. The first of these shows the four groups of compulsory subjects arising from the analysis.

The first cluster contains just one programme, the Introduction to the Teaching and Learning of Mathematics in the Primary Education course offered at the University of Huelva, which includes all the categories in its document except for PS instructional practice and acceptance of the PS challenge. Cluster 2 consists of the seven programmes that only include statements referring to PS competency, mentioning no other type of knowledge. The third cluster also consists of seven documents, and was the only one that included the PS disposition category. It also included the PS instructional practice, in every case. Finally, Cluster 4, the largest, contained the remaining 99 programmes, which had no mention either of the PS disposition category or of non-cognitive factors.

**Table VIII.** Cluster of compulsory subjects

		<b>Cluster 1 N=1</b>	<b>Cluster 2 N=7</b>	<b>Cluster 3 N=7</b>	<b>Cluster 4 N=99</b>
PS competency		1 (100%)	1 (14%)	7 (100%)	94 (95%)
PS knowledge	Problem characterization	1 (100%)	0	7 (100%)	91 (92%)
	PS process	1 (100%)	0	7 (100%)	33 (33%)
	Disposition	0	0	2 (28.5%)	0
PS pedagogical knowledge	Student as problem solver	1 (100%)	0	4 (57%)	6 (6%)
	PS as a worth-while task	1 (100%)	0	7 (100%)	79 (80%)
	Non-cognitive factors	1 (100%)	0	5 (71.5%)	0
	PS instructional practice	0	0	7 (100%)	58 (59%)

Source: The authors

Table IX shows the three clusters formed from the elective courses. A notable feature is that none of these courses includes the disposition category or that of non-cognitive factors. Cluster 1 groups five programmes and is characterised by the non-inclusion of the PS instructional practice category. Cluster 2 contains ten programmes and is characterised by mentions of the PS process category. Cluster 3, the largest, consists of programmes that only include PS competency, problem characterization, PS as a worth-while task and PS instructional practice.

**Table IX.** Clusters of elective subjects

		<b>Cluster 1 N=5</b>	<b>Cluster 2 N=10</b>	<b>Cluster 3 N=22</b>
PS competency		5 (100%)	10 (100%)	11 (50%)
PS knowledge	Problem characterization	5 (100%)	9 (90%)	2 (9%)
	PS process	0	5 (50%)	0
	Disposition	0	0	0
PS pedagogical knowledge	Student as problem solver	1 (20%)	3 (30%)	0
	PS as a worth-while task	2 (40%)	9 (90%)	5 (23%)
	Non-cognitive factors	0	0	0
	PS instructional practice	0	10 (100%)	9 (41%)

Source: The authors.

## Discussion and conclusions

This paper has two main objectives: to describe the subjects offered in the Didactics of Mathematics programme within official university degree courses that qualify holders to teach in primary education in Spain; and to describe the presence and teaching of PS in the teaching guides provided for these programmes.

Regarding the first objective, the national guidelines (Order ECI/3857/2007, of 27 December) include Mathematics within the Didactics module. However, they do not specify the number of subjects or ECTS credits for Mathematics, only the total number of credits (100 ECTS) for the module. In our analysis, median values of three compulsory subjects and one elective subject per Mathematics programme were calculated; for credits, the corresponding values were eighteen for the compulsory subjects and three for the elective ones. Although the number of credits has increased since the previous training plan for primary school teachers, we consider it to be still insufficient due to the strong presence of mathematics in primary education (RDL 157/2022, of 1 March, and also because pre-service teachers still obtain poor results, both in practical tests and in research studies such as TEDS-M (Instituto Nacional de Evaluación Educativa, 2012). Furthermore, only half of these universities offer any elective subjects related to the area of Mathematics; where this option is absent, future teachers do not have the opportunity to further their university training in the Didactics of Mathematics.

Regarding the organisation of compulsory subjects, most of the universities considered design their curriculum either focusing on the development of mathematical content knowledge, or on developing pedagogical content knowledge. The first case is reflected in models of mathematics knowledge such as those of Ball et al. (2008) and Carrillo et al. (2018), which separate content knowledge and pedagogical content knowledge into two domains. The second type of model combines both areas of knowledge in all the subjects offered, which are separated according to the mathematical content considered, although some may focus solely on content-related aspects, leaving aside didactics considerations.

Regarding the second objective, most of the teaching guides make some mention of PS, especially for the compulsory subjects and within the sections on competencies, objectives and/or content. In this regard, the Spanish training plans differed from those elsewhere, for example in Chile PS is very present in the stated objectives and methodology (Felmer & Perdomo-Díaz, 2017).

The documents considered almost always include the different types of tasks that are posed as problems, and fundamentally seek to promote

PS competency, as required at the national level (Order ECI/3857/2007, of 27 December): “Pose and solve problems related to everyday life.” However, posing problem is often omitted from the teaching guides, which mention PS pedagogical knowledge only with respect to the elective subjects. Furthermore, aspects considered relevant in the literature on PS knowledge (Chapman, 2015), such as problem solver’s consideration, non-cognitive factors and the disposition of the problem solver, as well as aspects of PS pedagogical knowledge such as stuck state in PS, assessment in PS and resource on PS, are absent from the documents for both types of subject.

The cluster analysis conducted on the compulsory subjects shows that some programmes focus on different aspects of teacher training in this field. Thus, only Cluster 3 considers the problem solver, despite the crucial importance of this figure in teaching PS skills; failure to take the solver into consideration could mean that the tasks posed do not represent real problems. Moreover, Cluster 2 only considers PS competency; this aspect, while valuable may not be sufficient, since the PS proficiency does not guarantee effective teaching. Regarding the clusters obtained for the elective subjects, Cluster 1 focuses on the tasks posed as problems, their difficulties and mistakes, and on the selection criteria applied. Cluster 2 considers a greater range of knowledge but ignores non-cognitive factors, although this is a critical aspect of PS teaching, as these factors may determine how the solver approaches the task. For example, students whose PS experience consists only of exercises that are rapidly solved may come to believe that all problems should be resolved equally quickly, and may not continue working on problems they could solve given more time (Schoenfeld, 2013).

Our analysis shows that, since only the competencies that must be acquired are stipulated at the national level, teaching guides vary widely. Each university has interpreted the teaching and learning of PS from its own perspective, meaning that teacher training in this context is extremely heterogeneous. These contrasts can lead to differences in future teachers’ classroom performance. As has been shown in previous research (Son & Lee, 2021), teachers’ conceptions of PS are related to their prior instruction. Furthermore, these teachers may continue to transmit traditional views of PS, since most teaching guides only address procedural aspects of the task.

Since the new teacher training guidelines for primary education include dedicating some time during the school day to PS (RDL 157/2022, of 1 March), and since there is a known association between what is studied in training and the specialised knowledge used in practice (Morris & Hiebert, 2017), we believe future teachers should be appropriately

prepared, in terms of their professional knowledge, for teaching PS, and that this preparation should go beyond the mere acquisition of the competency in question (Contreras, 2022). For example, training programmes could be based on the model proposed by Piñeiro et al. (2019), which details the professional knowledge needed for teaching PS, or take inspiration from teaching guides that detail much of the knowledge required in the latter model, such as the programme for the subject Introduction to the Teaching and Learning of Mathematics in Primary Education, at the University of Huelva, or the subject Problem Solving and Mathematical Connections, at the University of Almería.

A final observation is that the present study provides information on the initial PS training of primary school teachers in Spain, in an analysis that may be useful for designing future training plans. However, we are aware that in many cases the teaching guides are too brief and primarily focus on organisational aspects. Therefore, further research is needed to expand the information we present, for example via interviews with teacher trainers, considering the training programme, or through the analysis of classroom notes and observations.

## Bibliography

- Ball, D. L., Thames, M. H. & Phelps, G. (2008). Content knowledge for teaching: What makes it special? *Journal of Teacher Education*, 59(5), 389-407. <http://dx.doi.org/10.1177/0022487108324554>
- Bowen, G. A. (2009). Document analysis as a qualitative research method. *Qualitative Research Journal*, 9(2), 27-40. <https://doi.org/10.3316/QRJ0902027>
- Britz, J. & Richard, N. (1992). *Problem solving in the early childhood classroom*. National Education Assn.
- Carrillo, J., Climent, N., Montes, M., Contreras, L. C., Flores-Medrano, E., Escudero-Ávila, D., Vasco, D., Rojas, N., Flores, P., Aguilar- González, Á., Ribeiro, M. & Muñoz-Catalán, M. C. (2018). The mathematics teacher's specialised knowledge (MTSK) model. *Research in Mathematics Education*, 20(3), 236-253. <https://doi.org/10.1080/14794802.2018.1479981>
- Contreras, L. C. (2022). La nueva propuesta curricular & la formación del profesor. In T. F. Blanco, C. Núñez-García, M. C. Cañadas & J. A. González-Calero (Eds.), *Investigación en Educación Matemática XXV* (pp. 63-79). SEIEM.
- Chapman, O. (2015). Mathematics teachers' knowledge for teaching problem solving. *LUMAT*, 3(1), 19-36. <https://doi.org/10.31129/lumat.v3i1.1049>

- Felmer, P. & Perdomo-Díaz, J. (2017). Un programa de desarrollo profesional docente para el nuevo currículo de matemática: la resolución de problemas como eje articulador. *Educación Matemática*, 29(1), 201-217. <https://doi.org/10.24844/em2901.08>
- Fraenkel, J. R., Wallen, N. E. & Hyun, H. H. (2011). *How to design and evaluate research in education*. McGraw-Hill.
- Gourdeau, F. (2019). Problem solving as a subject and as a pedagogical approach, and the ongoing dialogue between mathematics and mathematics education. In P. Felmer, P. Liljedahl & B. Koichu (Eds.), *Problem solving in mathematics instruction and teacher professional development* (pp. 23-42). Springer. [https://doi.org/10.1007/978-3-030-29215-7\\_21](https://doi.org/10.1007/978-3-030-29215-7_21)
- Hähkiöniemi, M. & Francisco, J. (2019). Teacher guidance in mathematical problem-solving lessons: Insights from two professional development programs. In P. Felmer, P. Liljedahl & B. Koichu (Eds.), *Problem solving in mathematics instruction and teacher professional development* (pp. 279-296). Springer.
- Hiebert, J., Berk, D., Miller, E., Gallivan, H. & Meikle, E. (2019). Relationships between opportunity to learn mathematics in teacher preparation and graduates' knowledge for teaching mathematics. *Journal for Research in Mathematics Education*, 50(1), 23-50. <https://doi.org/10.5951/jresmetheduc.50.1.0023>
- Instituto Nacional de Evaluación Educativa. (2012). *TEDS-M Informe Español. Estudio Internacional sobre la formación en matemáticas de los maestros. IEA. Informe español*. Secretaría General Técnica. <https://www.educacionyfp.gob.es/dctm/inee/internacional/teds-mlinea.pdf?documentId=0901e72b8143866e>
- Lester, F. K. & Cai, J. (2016). Can mathematical problem solving be taught? preliminary answers from 30 years of research. In Felmer et al. (Eds.), *Posing and solving mathematical problems, research in mathematics education* (pp. 117-126). Springer. [https://doi.org/10.1007/978-3-319-28023-3\\_8](https://doi.org/10.1007/978-3-319-28023-3_8)
- Leung, S. S. (2016). Mathematical problem posing: A case of elementary school teachers developing tasks and designing instructions in Taiwan. In P. Felmer et al. (Eds.), *Posing and solving mathematical problems, research in mathematics education* (pp. 327-344). Springer.
- Liljedahl, P. & Cai, J. (2021). Empirical research on problem solving and problem posing: A look at the state of the art. *ZDM*, 53(4), 723-735. <https://doi.org/10.1007/s11858-021-01291-w>
- Mellone, M., Pacelli, T. & Liljedahl, P. (2021). Cultural transposition of a thinking classroom: To conceive possible unthoughts in mathematical

- problem solving activity. *ZDM*, 53, 785-798. <https://doi.org/10.1007/s11858-021-01256-z>
- Morris, A. K. & Hiebert, J. (2017). Effects of teacher preparation courses: Do graduates use what they learned to plan mathematics lessons? *American Educational Research Journal*, 54(3), 524-567. <https://doi.org/10.3102/0002831217695217>
- Orden ECI/3857/2007, de 27 de diciembre, por la que se establecen los requisitos para la verificación de los títulos universitarios oficiales que habiliten para el ejercicio de la profesión de Maestro en Educación Primaria. *Boletín Oficial del Estado*, 312, de 29 de diciembre de 2007. <https://www.boe.es/eli/es/o/2007/12/27/eci3857>
- Piñeiro, J. L., Castro-Rodríguez, E. & Castro, E. (2019). Componentes de conocimiento del profesor para la enseñanza de la resolución de problemas en educación primaria. *PNA*, 13(2), 124-139. <https://doi.org/10.30827/pna.v13i2.7876>
- Piñeiro, J. L., Castro-Rodríguez, E. & Castro, E. (2022). What problem-solving knowledge is required in mathematical teaching? *A curricular approach. Curriculum Perspective*, 42, 1-12. <https://doi.org/10.1007/s41297-021-00152-6>
- Pólya, G. (1981). *Cómo plantear & resolver problemas* (2a ed.). Trillas.
- Real Decreto 157/2022, de 1 de marzo, por el que se establecen la ordenación & las enseñanzas mínimas de la Educación Primaria. *Boletín Oficial del Estado*, 52, de 2 de marzo de 2022. <https://www.boe.es/eli/es/rd/2022/03/01/157/con>
- Rico, L. (2013). *Análisis Didáctico en Educación Matemática. Metodología de investigación, formación de profesores e innovación curricular* (pp. 23-58). Comares.
- Rico, L., Gómez, P. & Cañadas, M. (2014). Formación Inicial en educación matemática de los maestros de primaria en España, 1991-2010. *Revista de Educación*, 363, 35- 59. <https://doi.org/10.4438/1988-592X-RE-2011-362-152>
- Schoenfeld, A. H. (2013). Reflections on problem solving theory and practice. *The Mathematics Enthusiast*, 10(1-2), 9-34.
- Son, J. W. & Lee, M. Y. (2021). Exploring the relationship between preservice teachers' conceptions of problem solving and their problem-solving performances. *International Journal of Science and Mathematics Education*, 19, 129-150. <https://doi.org/10.1007/s10763-019-10045-w>
- Tatto, M. T. (2018). The mathematical education of primary teachers. In M. T. Tatto, M. C. Rodriguez, W. M. Smith, M. D. Reckase & K. Bankow (Eds.), *Exploring the mathematical education of teachers using*

*TEDS-M data* (pp. 205-256). Springer. [https://doi.org/10.1007/978-3-319-92144-0\\_8](https://doi.org/10.1007/978-3-319-92144-0_8)

Tyler, R. (1986). *Principios básicos del currículo*. Troquel.

Wake, G., Swan, M. & Foster, C. (2016). Professional learning through the collaborative design of problem-solving lessons. *Journal of Mathematics Teacher Education*, 19(2), 243-260. <https://doi.org/10.1007/s10857-015-9332-9>





# University autonomy: a multilevel analytical model from a Latin American perspective

## Autonomía universitaria: un modelo analítico multinivel desde una perspectiva latinoamericana

<https://doi.org/10.4438/1988-592X-RE-2025-412-735>

**Mario Alarcón Bravo**

<https://orcid.org/0000-0002-3939-724X>

*Universidad Diego Portales*

**José Joaquín Brunner Ried**

<https://orcid.org/0000-0002-7832-0237>

*Diego Portales University and University of Tarapacá*

### Abstract

This essay examines university autonomy as a dynamic and multifaceted concept, influenced by a range of political, economic, and social factors. It proposes an analytical model that approaches the phenomenon from three levels: macro (political-systemic), meso (institutional), and micro (individual). Unlike traditional approaches that focus on general or static definitions, this model takes a relational and dynamic approach, examining the interactions between the state, the market, university organizations, and individual academic actors. The model's primary contribution is the identification of specific analytical variables for each level, which facilitates the empirical comparison of autonomy dynamics across different contexts. The macro-level analysis focuses on the relationship between universities, the state, and the market, highlighting how reforms based on New Public Management have promoted a model of regulated autonomy that strains the balance between institutional independence and accountability. At the meso level, it examines the changes in university governance marked by the shift from the collegiate model to managerial structures, which have created tensions in the distribution of internal power. Finally, at the micro level, it analyses how external pressures affect the individual autonomy of academics, particularly in their research, teaching, and professional activities. The essay emphasizes that university autonomy in Latin America has evolved historically, from its consolidation with the Cordoba Reform in 1918 to the contemporary challenges posed by

globalization, neoliberal policies, and the commodification of higher education. It also highlights the need for comprehensive analytical approaches that recognize the internal tensions and dynamics of universities in the current context. The concluding reflections emphasize the importance of defending university autonomy as an essential principle for the development of democratic and just societies, stressing that it must be understood as a relational construct, subject to constant negotiation between universities and their environments.

*Keywords:* new public management, university autonomy, managerialism, academic profession, academic freedom, multilevel model, Latin America.

### **Resumen**

Este ensayo examina la autonomía universitaria como un concepto dinámico y multifacético, influenciado por factores políticos, económicos y sociales. Propone un modelo analítico que aborda este fenómeno desde tres niveles: macro (político-sistémico), meso (institucional) y micro (individual). A diferencia de enfoques tradicionales centrados en definiciones generales o estáticas, este modelo adopta una perspectiva relacional y dinámica, examinando las interacciones entre el Estado, el mercado, las organizaciones universitarias y los actores académicos individuales. Su principal contribución radica en identificar variables analíticas específicas para cada nivel, lo que facilita la comparación empírica de las dinámicas de autonomía en distintos contextos. El análisis a nivel macro se centra en la relación entre las universidades, el Estado y el mercado, destacando cómo las reformas basadas en la Nueva Gestión Pública han promovido un modelo de autonomía regulada que tensiona el equilibrio entre independencia institucional y rendición de cuentas. A nivel meso, se exploran las transformaciones en la gobernanza universitaria, marcadas por el desplazamiento del modelo colegiado hacia estructuras gerenciales, lo que ha generado tensiones en la distribución del poder interno. Por último, a nivel micro, se analiza cómo las presiones externas afectan la autonomía individual de los académicos, especialmente en sus actividades de investigación, docencia y participación profesional. El ensayo subraya que la autonomía universitaria en América Latina ha evolucionado históricamente, desde su consolidación con la Reforma de Córdoba en 1918 hasta los desafíos contemporáneos impuestos por la globalización, las políticas neoliberales y la mercantilización de la educación superior. Asimismo, destaca la necesidad de enfoques analíticos integrales que reconozcan las tensiones y dinámicas internas de las universidades en el contexto actual. Las reflexiones finales destacan la defensa de la autonomía universitaria como principio esencial para sociedades democráticas y equitativas, concebida como un constructo relacional en constante negociación con el entorno.

*Palabras clave:* nueva gerencia pública, autonomía universitaria, gerencialismo, profesión académica, libertad académica, modelo multinivel, Latinoamérica.

## Introduction

Since the 1980s, New Public Management (NPM) has transformed the organization of universities worldwide (Leišytė, Marquina, & Jones, 2025; Shattock, 2014; Veiga et al., 2018). These changes can be seen in the following features: greater integration and efficiency; a rebalancing of power between academic and administrative units; the increasing professionalization of management; the appointment rather than election of authorities; the incorporation of external actors into governing bodies; the centralization of strategic decisions; an emphasis on individual responsibility rather than collective work; and finally, the strengthening of middle management positions. These processes have been widely documented in Europe (Bruckmann & Carvalho, 2018; Krüger et al., 2018; Veiga et al., 2018), Asia (Da Wan et al., 2019), North America (Lavigne, 2018; Stephenson et al., 2025), and Latin America (Acosta, 2020; Brunner & Alarcón, 2024; Marquina et al., 2022).

The resulting academic debate has centered on faculty participation in governance, teaching, research, outreach, management tasks, and internationalization, as well as the adoption of management tools and job satisfaction (Teichler et al., 2022). These issues have primarily been examined within the contexts of academic capitalism (Brunner et al., 2022; Slaughter & Leslie, 1997) and professional theory (Abbott, 1988). Comparative studies have revealed variations in working conditions and professional status in both OECD countries (Kezar et al., 2019; Musselin, 2007; Teichler et al., 2022) and, increasingly, in Latin America (Alarcón, Brunner & Labraña, 2025; Brunner & Alarcón, 2024; Santin et al., 2022).

Extensive international surveys have driven this agenda, including the Carnegie (1991–1993) and Changing Academic Profession (CAP, 2007) studies, as well as the APIKS (2018–2019) initiative. The latter broadens the focus to include careers and governance (Teichler et al., 2022), and encompasses Argentina (Fuksman & Nosiglia, 2020), Mexico (Estévez Nenninger et al., 2020), and Chile (Brunner & Alarcón, 2025). Academic professionalization takes different forms in countries such as Germany, Canada, the United States, India, the United Kingdom, Russia, China, Japan, and Brazil. Finkelstein and Jones (2017) identify four factors that explain this diversity: the importance given to research, the extent to which the career is institutionalized, the balance between collegial and managerial governance, and the internal heterogeneity of each system.

In contrast, the impact of managerialism on academic autonomy and freedom has received less attention. Recent studies have highlighted the redistribution of power and the resulting tensions between managers, ac-

ademics, and professionals (Kenny, 2018). Findings are mixed: collegial influence persists in the Netherlands (Kallenberg, 2015), Portugal (Carvalho & Videira, 2019), and Argentina (Marquina et al., 2022), while a loss of autonomy has been recorded in Finland (Kallio, Kallio, & Blomberg, 2020) and France (Mignot-Gérard et al., 2023). Therefore, there are still analytical gaps regarding the factors that influence university autonomy. These gaps make it imperative to conduct comparative empirical studies (Carvalho & Videira, 2019) and formulate revised conceptual frameworks (Pekkola et al., 2018).

This essay proposes a multilevel analytical model for studying university autonomy from a Latin American perspective. It posits that university autonomy is a multifaceted concept requiring analysis at three levels: political-systemic (macro), institutional (meso), and individual (micro). The essay is structured into four sections. First, the conceptual framework used to construct the analysis is described. Second, the evolution of academic autonomy in Latin America since the Córdoba Reform of 1918 is analyzed. Finally, the proposed multilevel analytical model is presented.

## **Conceptual framework**

### **University autonomy**

University autonomy is a guiding principle of higher education systems, linking institutional independence with social responsibility. It means each university can decide on governance, administration, teaching, and research without undue external interference (Maassen et al., 2017). In the United States and Europe, this autonomy is closely linked to academic freedom, which is understood as the right of individuals to teach and conduct research guided by intellectual rigor and the pursuit of truth (Bernasconi, 2021; Lerch et al., 2024). In Latin America, however, autonomy is often viewed as a corporate attribute that can hinder accountability and responsiveness to public demands in some instances (Bernasconi, 2024; Mendoza, 2020). This contrast reflects divergent historical trajectories: Anglo-Saxon systems prioritize individual rights and favor flexibility and innovation. In contrast, Latin American systems subordinate academic freedom to institutional autonomy, thereby reducing the scope of action of teachers and researchers.

The literature distinguishes between two types of autonomy: substantive autonomy, which is the ability to set goals, curricula, and research priorities, and procedural autonomy, which is the power to define the means to achieve these goals (Kohtamäki & Balbachevsky, 2018; Maas-

sen et al., 2017; Wen & Marginson, 2023). Estermann & Steneil (2011), on the other hand, identify four interdependent dimensions: organizational, relating to the design of internal structures and governing bodies; academic, relating to the freedom to structure programmes, methodologies, and lines of research; financial, referring to the autonomous management of resources and the ability to generate income; and personnel, focused on the power to hire, evaluate, and promote academic and administrative staff.

Over the last decade, the concept has evolved towards the notion of 'living autonomy', which emphasizes that effective autonomy depends more on self-governance practices and the strategic capacity of universities to respond to their environment than on legal provisions (Maassen et al., 2017; Maassen, 2024). In Europe, growing demands from governments, the private sector, and other external actors have strengthened oversight mechanisms. In Latin America, however, these pressures are combined with the internal dynamics of politicization — the 'militant university' described by Brunner (2015) — which can restrict both institutional autonomy and academic freedom (Mendoza, 2020).

Considering these tensions, a multilevel analytical approach (macro, meso, and micro) is necessary to examine how power is reconfigured in university governance and the implications of these changes for contemporary academic autonomy and freedom (Kwiek, 2019).

### **Macro-level changes: university autonomy and its relationship with the state and the market**

At this level, autonomy is understood as the ability of institutions to govern themselves, remain free from direct state intervention, and act as guardians of knowledge and expertise. During the 17th and 18th centuries, universities enjoyed considerable autonomy, operating as independent entities in what Clark (1983) termed the 'ivory tower', in line with Humboldt's notion of freedom and isolation. This concept posits that the pursuit of truth necessitates freedom from external interference and a certain degree of self-absorption, enabling one to establish one's own research and teaching agendas.

However, with the emergence of the nation-state in the 19th century and the expansion of modern universities, this autonomy began to come under strain. Universities became part of the state structure, receiving public funding and participating in the training of professionals for public service. While they retained some autonomy, their relationship with the state became more complex as they had to balance funding with academic freedom (Finkelstein & Jones, 2017).

Since the 1980s, reforms inspired by New Public Management (NPM) have radically altered this relationship (Veiga et al., 2018). These policies, designed to increase efficiency and reduce costs, introduced external control mechanisms that limited traditional autonomy. Evaluation and accountability systems were implemented, encouraging self-regulation but also imposing limits on institutional freedom. This new regime, known as 'regulated autonomy', has been widely documented in Europe (Krüger et al., 2018) and Latin America (Acosta-Silva, 2022; Alarcón & Brunner, 2024) and is characterized by the state defining regulatory frameworks within which universities can manage themselves.

In this scenario, university autonomy is under strain from state control and growing market pressures. The decline in direct public funding has prompted universities to diversify their income streams, adopting strategies associated with academic capitalism. Initially developed by Slaughter and Leslie (1997) based on studies in the United States, the United Kingdom, and Canada, this conceptual framework has recently been re-interpreted from a comparative perspective that recognizes the varieties of academic capitalism present in Latin America (Brunner et al., 2022).

From this perspective, universities compete for external resources, form alliances with industry, and strive to attract international students, thereby intensifying the commercialization of higher education. While this process may increase institutional financial autonomy, it also limits universities' ability to define their own academic and research priorities independently, as evidenced by studies in Europe (Carvalho, 2017) and Latin America (Pérez Mora et al., 2022).

### **Changes at the meso level: university governance and redistribution of power**

At this level, academic autonomy is articulated within higher education institutions themselves. The internal distribution of power has been altered by the same reforms that affected the macro level. Traditionally, university governance has been based on collegial structures, in which academics play a central role in decision-making through senates and committees (Deem & Brehony, 2005; Carvalho & Santiago, 2015). In Latin America, this model has historically been characterized by a strong emphasis on university self-governance and extensive institutional autonomy, which is regarded as fundamental to the university's identity as an academic institution (Alarcón & Brunner, 2024).

However, reforms inspired by managerialism have introduced a more hierarchical and professionalized structure, shifting decision-making power to rectors and deans who are often selected based on management

criteria rather than academic merit (Carvalho, 2017). In Latin America, unlike in Europe, these changes have been more gradual and conflictive, resulting in hybrid models that combine emerging managerial practices with elements of the existing collegial academic culture (Alarcón, 2020).

At the same time, the professionalization of management has led to the emergence of a layer of non-academic administrators who have decision-making power over key issues. This scenario has caused friction between academic bodies and management teams, as well as a widespread perception among faculty of a loss of control. Nevertheless, this tendency is less pronounced in Latin America than in Europe, as evidenced by recent comparative studies (Jones & Weinrib, 2019; Marquina et al., 2022).

### **Changes at the micro level: the fragmentation of the academic profession and the resulting loss of individual autonomy**

Academic autonomy at this level refers to the freedom of academics to define their own research agendas, teaching methods, and professional activities. However, this freedom has been eroded by the fragmentation and stratification of the academic profession (Finkelstein & Jones, 2019). Horizontal fragmentation, resulting from the increasing specialization of educational activities, and vertical fragmentation, characterized by the rise of precarious employment categories, have further eroded individual autonomy. Precarious employment forces academics to focus their efforts on activities that ensure contract renewal, thereby restricting their ability to pursue their own lines of research (Kezar et al., 2019).

In Latin America, the normative ideal of the academic profession has shifted progressively towards the academic-researcher figure, as incentive systems have become more focused on scientific productivity (Estévez-Nenninger et al., 2020; Bernasconi, Berríos, and Véliz, 2021). While this transformation has strengthened research capacities in various institutions, it has also generated persistent tensions regarding the balance between teaching and research functions, as well as intensified pressure on academics to align with performance standards defined by external agendas.

While these changes have not explicitly suppressed individual academic autonomy, they have reconfigured it under the parameters of efficiency, evaluation, and administrative control. Consequently, there has been a gradual loss of professional control over the conditions and direction of academic work in an institutional environment where managerial logic tends to prevail over traditional principles of disciplinary self-regulation (Brunner & Alarcón, 2024).

## **Academic Autonomy in Latin America: 100 Years After the Córdoba Reform**

The Córdoba Reform of 1918 was a significant event in the history of Latin American higher education, establishing university autonomy as a guiding principle. Promoted by students at the National University of Córdoba amid oligarchic and clerical domination, the reform sought to democratize institutions controlled by elites (Balbachevsky, 2014; Bernasconi, 2024). The Preliminary Manifesto proclaimed autonomy as a safeguard against state and ecclesiastical control, establishing universities as environments conducive to critical thinking and innovation (Bergel, 2008; Marsiske, 2017).

This concept was structured around three pillars: administrative, enabling the democratic election of authorities with tripartite participation; academic, granting autonomy to define curricula, teaching methods, and research agendas; and financial, allowing for independent resource management (Marsiske, 2010; 2017).

Córdoba's influence quickly spread to Mexico, Chile, Peru, Brazil, and other countries, adapting to different contexts. In Mexico, for example, autonomy emerged from student mobilizations calling for democratization and disengagement from political power. In other cases, however, it was restricted by military dictatorships during the 1960s and 1970s, resulting in the repression of students and academics, as well as severe state intervention (Burbano, 2011; Marsiske, 2017).

Following the return to democracy in the 1980s, the defense of autonomy was reinvigorated, establishing it as a fundamental principle, albeit one subject to new pressures. In the 21st century, globalization, technological advances, and neoliberal policies have generated additional tensions. On the one hand, autonomy is limited by the requirements of external evaluations and quality standards promoted by commercial organizations (Acosta-Silva, 2021; Sisto, 2020). Conversely, the expansion of the private sector has intensified competition for resources, students, and prestige (Burbano, 2011; Cantard, 2015).

In this context, autonomy is viewed as a relational, dynamic, and contextualized construct, emerging from the interactions between universities, the state, and societal actors, and influenced by historical and political contexts (Ordorika, 2010). This perspective highlights autonomy as both a safeguard against external interference and a space for negotiation, ensuring academic freedom and intellectual independence. Acosta-Silva (2020) describes it as a conflictual process between the demands of universities and the interests of the state. In contrast, Del Bello

(2022) emphasizes its role in generating and disseminating knowledge, as well as safeguarding institutional integrity in the face of external pressures.

Although autonomy takes different forms, Marsiske (2017) identifies common elements in its exercise, such as self-governance through democratic elections and co-governance, which have been consolidated in Mexico and Argentina, as well as academic freedom to define content, methods, and programs, and financial autonomy to manage budgets and generate income. These mechanisms have established universities as platforms for discourse and societal change, thereby gaining legitimacy in the eyes of the state (Ordorika, 2010).

However, the exercise of autonomy has generated controversy. Initially conceived as a defense against political power, its limited accountability, opacity, and poor response to social demands have been questioned (Mendoza, 2020). In some cases, its use as a means of confronting the state has distorted its original meaning.

Public universities today face the challenge of balancing institutional independence with social responsibility, while also contributing to the economic and cultural development of their communities. Neoliberal policies have exacerbated this tension by treating them as service providers rather than public goods (Marsiske, 2010; Villar, 2015). Although autonomy is enshrined in several constitutions, allowing them to withstand political and economic crises, their dependence on public funding and pressure from external evaluation mechanisms continues to limit their autonomy (Burbano, 2011; Ordorika, 2010; Villar, 2015).

## **Academic autonomy on three levels: an analytical model**

The proposed model aims to address the limitations of traditional approaches, which have typically focused on the legal-formal or institutional level and adopted a one-dimensional or normative perspective when examining university autonomy. Based on a relational and dynamic conception of autonomy, the model enables us to observe how the margins of university autonomy are configured, negotiated, and strained across three interconnected levels: macro (political-systemic), meso (institutional), and micro (individual).

Each level is operationalized through specific analytical variables formulated based on a critical review of literature and a comparative, empirical analysis of Latin American universities. The following table summarizes the model's structure and explains the analytical focus and key variables considered at each level.

**Table 1:** Three-level analytical model

Level	Analytical focus	Key analytical variables
Macro	Relationship between the state, the market and the university	Academic self-governance; legal basis and academic autonomy; state funding and academic autonomy
Meso	Institutional governance and redistribution of power	Control over the hiring of academics; control over academic careers (promotion and rank); peer review
Micro	Individual academic working conditions	Freedom to define research agendas; freedom in teaching methods; freedom to participate in professional activities

This model contributes to the analytical clarification of the dimensions of university autonomy, offering a valuable tool for its comparative empirical study. Its relational approach enables us to understand how Latin American universities negotiate their autonomy within an increasingly complex landscape, caught between the logics of external control, internal institutional reorganization, and evolving conditions of academic work.

However, the multilevel scheme has three limitations that should be made explicit: (i) it is based mainly on secondary literature, which risks homogenizing heterogeneous national realities; (ii) the rigid compartmentalization of the macro, meso, and micro levels reduces the density of interactions between the state, the market, and institutions; and (iii) it lacks empirical corroboration, meaning its explanatory power is conjectural. Recognizing these limitations opens up an agenda for refining and testing the model through comparative studies and mixed methodologies.

## Macro level: political-systemic

In higher education, this level is where the state, the market, and higher education institutions converge to shape the operational framework of universities. Clark's triangle (1983) illustrates these power relations, associating the state with regulation and public policy, the market with supply and demand dynamics, and institutions with autonomy.

This conceptual framework enables us to analyze how universities are simultaneously subject to pressure from the state, the market, and their own internal structures. The balance between these forces varies according to historical, political, and social context. In Latin America, for example, university autonomy has played a crucial role in safeguarding academic freedom and fostering research excellence (Bernasconi, 2024), establishing itself as a vital means of addressing societal needs.

## **Academic self-governance**

The cornerstone of university autonomy in Latin America, academic self-governance enables university communities to make key decisions independently of external intervention. This prerogative includes electing authorities, defining curricula, setting faculty conditions, and managing resources autonomously, thus balancing academic freedom, political independence, and administrative management (Acosta-Silva, 2008).

However, since the 1980s, socio-economic and political changes have transformed the meaning and practice of self-governance. Economic crises, structural adjustments, and accountability policies have limited its scope by demanding efficiency, quality, and social relevance (Acosta-Silva, 2008). This environment has introduced management models based on management by objectives, performance indicators, and incentives, which put pressure on the principles of autonomy (Acosta-Silva, 2022; Labraña & Brunner, 2022).

The involvement of new stakeholders, such as pressure groups and the private sector, has complicated university governance and challenged the cohesion and legitimacy of self-government. The challenge for universities in the region is to preserve autonomy while responding to social and political demands (Acosta-Silva, 2008; Bernasconi, 2015).

## **Legal bases and academic autonomy**

Legal bases protect and regulate university autonomy, providing a framework that shields universities from external interference. In Latin America, for example, autonomy is enshrined in the constitutions of several countries, granting legal independence from the state (Bernasconi, 2024; Marsiske, 2017). This legal framework safeguards fundamental academic values, including academic freedom, independent research, and institutional self-governance.

In Mexico, for example, university autonomy was formalized in 1929 when the National Autonomous University of Mexico (UNAM) was granted autonomy, setting a precedent for other universities in the region (Acosta-Silva, 2021). However, legal autonomy does not always guarantee complete independence, as universities may still be subject to external regulations, such as those imposed by accreditation agencies, which can limit curricular or administrative decision-making processes (Valadés, 2015).

Educational reforms promoted by international organizations encourage the standardization of education systems to adjust them to a competitive global market (Marquina, 2020). While these reforms aim to modernize institutions, they may also hinder their capacity to adapt policies and practices to local contexts, thereby disrupting the equilibrium between autonomy and external intervention (Marsiske, 2017).

## **State funding and academic autonomy**

State funding is crucial for the autonomy of public universities in Latin America (Bernasconi, 2021; Kohtamäki & Balbachevsky, 2018; Mendoza, 2020). However, dependence on state funds can limit self-governance, as governments often make resources conditional on the fulfillment of political or development objectives (Marsiske, 2017). This dilemma creates a fundamental tension: while funding is essential, it can restrict academic freedom if it is dependent on efficiency and productivity indicators.

Financial autonomy necessitates striking a balance between institutional independence and accountability to resource providers (Bernasconi, 2020; Marsiske, 2010; Tünnermann, 2008). While state funding has promoted equity and inclusion in access to higher education, it has also introduced control mechanisms that challenge university autonomy.

Currently, universities are under pressure to adopt managerial logic and respond to performance indicators imposed by external funders. These demands have altered the relationship between universities and the state, presenting new challenges to preserving the autonomy and core values of higher education (Sisto, 2020).

## **Meso level: institutional**

Analyzing higher education institutions at this level involves addressing the tensions between two management models: collegiality and managerialism. Here, universities must balance their traditional teaching and research functions with the need for efficiency, accountability, and competitiveness. These tensions directly impact academic autonomy, influenced by three key factors: control over hiring, academic careers, and peer review.

### **Control over the hiring of academics**

Control over hiring is essential for academic autonomy. In the collegial model, academics were traditionally selected based on strictly academic criteria, with decisions made by the scholarly community (Bernasconi, 2021; Kohtamäki & Balbachevsky, 2018; Mendoza, 2020). However, managerialism introduces strategic criteria that seek to align hiring with external funding goals and positioning in international rankings, thereby shifting control to administrators or external boards.

In Chile, for example, the adoption of performance-based funding instruments has led to an increasing centralization of selection processes within rector's offices, thereby limiting the effective participation of academic bodies. By contrast, in Mexico, programs such as the PIFI (Com-

prehensive Institutional Strengthening Program) and the SNI (National System of Researchers) have introduced economic incentives that prioritize profiles demonstrating high scientific productivity and the capacity to secure funding, thereby subordinating purely disciplinary criteria to considerations of efficiency and competitiveness.

Meanwhile, the National University of Córdoba continues to operate with tripartite selection committees, which partly preserves the collegial logic. However, the recent establishment of a Secretariat for Strategic Planning means that new positions are allocated based on organizational goals. This new arrangement has created a hybrid model in which academic deliberation and strategic management coexist (Brunner & Alarcón, 2024). These cases illustrate how the interplay between external incentives and institutional self-governance capacity reshapes autonomy in academic hiring.

### **Control over academic career: promotions and ranks**

Control over academic careers, particularly promotion and rank, also affects academic autonomy (Bernasconi, 2021; Kohtamäki & Balbachevsky, 2018; Mendoza, 2020). In the collegiate model, academic committees evaluate performance based on the quality of research, teaching, and service to the academic community when making decisions. This system ensures high academic standards and a fair process (Mignot-Gérard et al., 2022).

However, the adoption of managerial practices has made promotion more dependent on productivity indicators, such as the number of publications and external funding (Acosta-Silva, 2021). While these criteria aim to professionalize academia, they can also distort the process by prioritizing quantity over quality, creating tensions between managerial logic and academic autonomy (Villanueva, 2022). In Latin America, having autonomous control over promotion is crucial to avoid external interference in strategic decisions (Marquina, 2020).

The cases of Mexico, Chile, and Argentina demonstrate a converging trend towards the use of quantitative indicators, albeit with varying degrees of intensity and effect. In Mexico, incentive programs such as the SNI have reinforced a culture of evaluation oriented towards measurable results, generating academic trajectories that are more closely aligned with productivity parameters. In Chile, competitive funding schemes such as FONDECYT (the National Fund for Scientific and Technological Development) have introduced similar requirements, albeit with less systematic application. By contrast, Argentina's academic committees retain greater decision-making power, maintaining a tradition of comprehensive evaluation that considers a variety of merits.

These differences reflect how managerialism is redefining the criteria and mechanisms of academic hierarchy, adjusting the scope of autonomy according to the strength of external incentives and the institutionalization of collegial norms (Brunner & Alarcón, 2024).

### **Peer review**

Peer review is the traditional mechanism for ensuring the quality of academic work. This system enables decisions about research and teaching to be made by individuals with the most excellent understanding of their value, thereby reinforcing academic autonomy (Bernasconi, 2021). However, peer review has been supplemented, or in some cases replaced, by evaluation systems based on performance metrics, such as bibliometric indicators or rankings (Acosta-Silva, 2021). While these metrics introduce objectivity, they can also erode academic autonomy by reducing academics' control over the evaluation of their work and creating a culture of performativity that prioritizes quantity over quality (Bruckmann & Carvalho, 2018). In Latin America, peer review remains crucial for safeguarding academic autonomy against external pressures and for maintaining high academic standards (Marquina, 2020).

However, its practical application varies significantly between countries. In Chile, for example, peer review of teaching and research is relatively rare, reflecting a greater reliance on formal and managerial quality assurance systems. In contrast, Argentina and Mexico have higher levels of academic participation in these processes, though they too face mounting pressure to adopt standardized criteria. Mexico, for example, has extensively institutionalized student evaluation, particularly in teaching, raising questions about the balance between pedagogical legitimacy and professional control (Brunner & Alarcón, 2024).

Ultimately, comparative evidence shows that the tension between collegiality and managerialism does not result in one replacing the other, but rather in hybrid configurations emerging. When resource allocation and hiring are linked to external indicators, as in Chile and to a greater extent in Mexico, the academic voice is reduced, and autonomy depends on the bargaining power of the base units. Where robust collegial mechanisms persist (as in Argentina), managers must reconcile performance logic with disciplinary deliberation. Understanding these different trajectories is essential for designing policies that bolster participatory governance while maintaining transparency and institutional accountability.

## **Micro level: individual**

The analysis of academic autonomy at the micro level focuses on the personal sphere, where academics carry out their work in areas such as research, teaching, and complementary professional activities. At this level, autonomy is defined as an academic's ability to make informed decisions based on expert judgment, free from external coercion by the university administration, the state, or private funders (Bernasconi, 2021; Kohtamäki & Balbachevsky, 2018; Mendoza, 2020). This form of autonomy is crucial for the comprehensive development of academics and for ensuring that public universities remain spaces for critical thinking, knowledge creation, and social responsibility (Marsiske, 2010; Acosta Silva, 2020; Marquina, 2020).

Three main factors that condition academic autonomy at this level have been identified: freedom in terms of research agenda, teaching methods, and participation in professional activities. Each of these elements significantly impacts how academics structure their work and, ultimately, the ability of universities to fulfill their mission of knowledge creation and transfer.

### **Autonomy in the research agenda**

The freedom to define research topics and questions is an essential component of individual academic autonomy. This capacity enables academics to develop projects relevant to their discipline and society without external interference, promoting the advancement of knowledge by facilitating the exploration of emerging and innovative areas that might otherwise be marginalized (Kohtamäki & Balbachevsky, 2018).

However, research autonomy is facing increasing restrictions. Dependence on external funding imposes conditions that tend to prioritize areas of research established by funding agencies, which may not always align with academic interests or local challenges (Pérez-Mora et al., 2022). In this context, safeguarding autonomy requires institutional mechanisms that ensure access to diversified sources of funding aligned with research freedom.

In Latin America, recent reforms influenced by global evaluation and funding models have exacerbated tensions between international priorities and local agendas. Academics are under pressure to align their choice of topics with external strategic priorities or to respond to evaluation systems that favor specific approaches over others (Marquina, 2020; Acosta-Silva, 2021).

Nevertheless, national systems exhibit varying configurations (Alarcón et al., 2025). In Mexico, for example, the focus is on productivity and

market links. In contrast, in Chile, institutional policies promote research with social impact and international projection, albeit under significant pressure to secure competitive funding. In Argentina, however, there is greater freedom to define agendas based on academic deliberation, with less interference from external criteria, despite the partial adoption of global standards. These differences suggest that research autonomy is negotiated at the intersection of institutional logics, funding requirements, and national regulatory frameworks.

### **Autonomy in teaching methods**

Academic freedom is also expressed in teaching, where academics must be able to design and structure their courses according to advances in their discipline and the characteristics of their students (Kohtamäki & Balbachevsky, 2018; Mendoza, 2020). This autonomy is essential for providing relevant, up-to-date training that is aligned with emerging knowledge and the training needs of each context (Marsiske, 2010). Preserving decision-making margins in teaching enables teachers to experiment with methodologies, adapt learning paces, and implement pedagogical approaches that promote more meaningful learning processes (Marquina, 2020).

However, in several Latin American institutions, this autonomy is restricted by curriculum standardization and evaluation processes driven by quality assurance policies. While these mechanisms aim to ensure minimum levels of institutional performance, they can also limit teachers' ability to innovate pedagogically or adapt their teaching to specific contexts (Acosta-Silva, 2021). The expansion of international teacher evaluation metrics has reinforced these trends by imposing performance criteria that do not always reflect local conditions or the specific needs of students (Marquina, 2022).

Regulations on teaching vary significantly between countries. In Argentina and Mexico, for example, teaching organizations are highly regulated in terms of timetables and external evaluation requirements. In these systems, teaching supervision usually falls to the heads of academic units, and student evaluation is widely institutionalized. In contrast, the Chilean system has lower levels of formal regulation regarding teaching times and course structure, opening up potential spaces for more diverse teaching approaches. However, this greater operational autonomy is offset by intense student scrutiny of teaching (Brunner & Alarcón, 2024; Alarcón et al., 2025).

In all cases, face-to-face teaching remains the predominant mode, although hybrid and virtual learning experiences are also emerging. These

transitions require us to rethink teaching autonomy, viewing it not only as academic freedom but also as the ability to adapt pedagogical tools to new teaching and learning environments critically. Therefore, the challenge lies in resisting standardization and finding institutional ways to safeguard and promote autonomy as a condition for reflective and innovative teaching.

### **Autonomy in academic and professional activities**

The third factor that influences academic autonomy at an individual level is the freedom to engage in additional educational and professional activities, such as attending conferences, joining academic networks, and undertaking consultancy work (Bernasconi, 2021; Kohtamäki & Balbachevsky, 2018; Mendoza, 2020). Academics not only teach and conduct research, but also actively participate in their field of study through external activities that keep them up to date with the latest developments and enable them to apply their knowledge in practical contexts (Pérez-Mora et al., 2022).

Participation in consulting, technical advisory services, and other external activities is also a significant expression of an academic's professional autonomy. In various Latin American countries, it is common for university teachers to collaborate with public agencies, social organizations, the media, or the productive sector, acting as specialists in their respective fields. These interactions allow academics to disseminate university knowledge more widely and also provide them with feedback on their own academic practices. These relationships broaden their analytical frameworks and enrich both teaching and research (Acosta Silva, 2022; Bernasconi, 2024).

However, this dimension of autonomy faces structural limitations. The scarcity of resources and insufficient institutional support hinder active participation in international collaboration networks and professional outreach opportunities. This situation is exacerbated by growing pressure on universities to meet institutional performance goals, such as program accreditation and improving efficiency and employability indicators. In this scenario, activities that do not directly contribute to these objectives are often postponed (Acosta Silva, 2020).

Added to this is an intensification of academic service tasks, especially those aimed at student services, the organization and formalization of which varies between countries (Alarcón et al., 2025). In Mexico, these tasks tend to be highly regulated and associated with explicit goals, whereas in Argentina, they are implemented more heterogeneously. In Chile, on the other hand, there is less formalization, reflecting differenc-

es in institutional regulation of these interactions. While these functions play an essential role in student training, they can also reduce the time available for participation in external networks, consultancy work, and outreach activities. However, their impact on professional autonomy appears to be mitigated when academics perceive working conditions that recognize their efforts and promote their professional development.

## Final reflections

This essay has critically reviewed the notion of university autonomy, proposing a multilevel model that captures the inherent complexity and tensions of the concept within the context of Latin American higher education institutions. As we have demonstrated, autonomy cannot be analyzed from a normative or prescriptive perspective alone. However, it must be understood as a dynamic construct influenced by political, economic, and social factors. Addressing the macro, meso, and micro levels enables a more nuanced understanding of the various elements that influence university autonomy and the emerging tensions between academia and university management.

At the macro level, the influence of the state and the market on Latin American universities raises essential questions about these institutions' ability to resist growing pressure to be more efficient and accountable. While universities undoubtedly must be responsible to society, public policies must respect the principles of academic freedom and self-governance, which have been fundamental to the development of knowledge.

At the institutional level, the increasing importance of professional management and the introduction of standardized evaluation mechanisms pose a significant challenge to collegial governance. Although managerialism may offer practical solutions in terms of organizational efficiency, its long-term effects on academic autonomy must be considered. The excessive centralization of power and the imposition of quantitative metrics on academic output risk undermining participatory processes and the educational ethos, thereby compromising academics' ability to influence key university decisions.

Finally, at an individual level, academics are subject to an increasing amount of external pressure that limits their autonomy in conducting research, teaching, and contributing to public debate without restrictions imposed by external agendas. The growing precariousness of academic careers, coupled with an emphasis on productivity and internationalization, has led to fragmentation within the profession. These unstable

labor conditions exacerbate inequalities within the educational community, calling into question the fundamental principles of intellectual and professional freedom.

In conclusion, this conceptual essay underscores the significance of adopting a multilevel perspective to comprehend university autonomy within the Latin American context. While the tensions between autonomy and external demands are not new, they have taken on a new dimension under contemporary reforms. While neoliberal policies and managerial models have undoubtedly transformed institutional governance, universities must find ways to preserve their academic and social missions. As a guiding principle of higher education, autonomy must be defended not only as an abstract ideal, but also as a fundamental condition for developing democratic, equitable, and knowledge-based societies.

## Bibliographic

- Acosta Silva, A. (2008). La autonomía universitaria en América Latina: Problemas, desafíos y temas capitales. *Universidades*, 36, 69–82.
- Acosta Silva, A. (2021). Autonomía universitaria y Estado: La experiencia en Latinoamérica. *Revista de Educación Superior*, 49(193), 1–25.
- Acosta Silva, A. (2022). Autonomía universitaria, gobierno institucional y gobernanza interpretativa en México. *Perfiles Latinoamericanos*, 30(59). <https://doi.org/10.18504/pl3059-016-2022>
- Alarcón, M., Brunner, J. J. y Labraña, J. (2025). Between managerialism and collegiality: The transformation of the academic profession in Ibero-America. *European Journal of Higher Education*, 1–21. <https://doi.org/10.1080/21568235.2025.2491076>
- Alarcón, M. & Brunner, J. J. (2024). Multi-governance in higher education: The case of Chile 2018-2023. *Research in Education*, 119(1), 97–118. <https://doi.org/10.1177/00345237241248535>
- Balbatchevsky, E. (2014). The Latin American university model and the challenges posed by the reforms: Perspectives from the academics. In B. Cantwell and I. Kauppinen (Eds.), *The relevance of academic work in comparative perspective* (pp 239–252). Springer.
- Bergel, M. (2008). Latinoamérica desde abajo. Las redes transnacionales de la Reforma Universitaria (1918–1930). En *La reforma universitaria: Desafíos y perspectivas noventa años después* (pp. 146–184).
- Bernasconi, A. (2021). University autonomy and academic freedom: Contrasting Latin American and US perspectives. *Higher Education Governance and Policy*, 2(1), 56–67.

- Bernasconi, A. (2023). Academic freedom in Latin America and the deceptive comfort of autonomy. *International Higher Education*, 116, 15. <https://doi.org/10.36197/IHE.2021.105.03>
- Bruckmann, S., y Carvalho, T. (2018). Entendiendo el cambio en la educación superior: un enfoque arquetípico. *Higher Education*, 1–19. <https://doi.org/10.1007/s10734-018-0229-2>
- Brunner, J. J. y Alarcón, M. (2025). More managerialism and less collegiality? The management of Chilean universities from the perspective of academics. In L. Leišytė, M. Marquina, and G. A. Jones (Eds.), *University governance, management and the academic profession* (Vol. 26, pp. 137–152). Springer. [https://doi.org/10.1007/978-3-031-86889-4\\_8](https://doi.org/10.1007/978-3-031-86889-4_8)
- Brunner, J. J. y Alarcón, M. (2024). Academic management in Chilean universities: An analysis from academics' perspective. *European Journal of Education*, 59(4), e12798.
- Brunner Ried, J. J., y Alarcón-Bravo, M. (2024). El gobierno de las universidades chilenas en perspectiva internacional comparada. *Revista Colombiana de Educación* (93), 193–217.
- Brunner, J. J., Salmi, J., Labraña, J., Balbachevsky, E., Ganga, F., ... y Schwartzman, S. (2022). *Enfoques de sociología y economía política de la educación superior: Aproximaciones al capitalismo académico en América Latina*. Ediciones Universidad Diego Portales.
- Brunner, J. J. (2015). Ideas y fines de la universidad. *Estudios Públicos*, 139, 155–164.
- Brunner, J. J. (2011). Gobernanza universitaria: Tipología, dinámicas y tendencias. *Revista de Educación*, 355, 137–159.
- Cantard, A. (2014). La autonomía universitaria hoy: un debate necesario. *Unión de Universidades de América Latina y el Caribe*.
- Cardoso, S., Carvalho, T. e Videira, P. (2019). Is it still worth working in academia? The views from Portuguese academics. *Higher Education Policy*, 32(4), 663–679.
- Carvalho, T. (2017). New Public Management and the academic profession. In J. Huisman and M. Tight (Eds.), *Theory and method in higher education research* (Vol. 3, pp. 59–76). Emerald Publishing.
- Carvalho, T. (2018). New Public Management and the academic profession. In P. Teixeira and J. Shin (Eds.), *Encyclopedia of international higher education systems and institutions*. Springer. [https://doi.org/10.1007/978-94-017-9553-1\\_310-1](https://doi.org/10.1007/978-94-017-9553-1_310-1)
- Da Wan, C., Lee, M. N., & Loke, H. Y. (Eds.). (2019). *The Governance and Management of Universities in Asia: Global Influences and Local Responses*. Routledge.

- Del Bello, J. C. (2022). *Los alcances de la autonomía universitaria*. Editorial EUDEBA.
- Estermann, T., & Steinel, M. (2011). University autonomy in Europe. *Beiträge zur Hochschulforschung*, 33(2), 86–105.
- Estévez-Nenninger, E. H., González-Bello, E. O., Valdés-Cuervo, Á., Arcos-Vega, J. L., Ramiro-Marentes, F. y Gutiérrez-Franco, L. E. (2020). Teaching and research of academics in Mexico: Preferences and dedication according to the international survey APIKS. In *Higher Education Forum*, 17, 99–114. Research Institute for Higher Education, Hiroshima University.
- Fuksman, B. U. y Nosiglia, M. C. (2020). El lugar de la investigación en la profesión académica argentina: Hallazgos del estudio internacional APIKS. *Revista Latinoamericana de Políticas y Administración de la Educación*, 12, 61–81.
- Jones, G. A., & Weinrib, J. (2022). The Changing Context of Academic Work: Fragmentation, Institutional Horizontal Diversity, and Vertical Stratification. In S. Lysons and T. Burgess (Eds.), *Research handbook on academic careers and managing academics* (pp. 36–46). Edward Elgar Publishing.
- Jones, G. A., & Finkelstein, M. (2019). Looking across systems: Implications for comparative, international studies of academic work. In M. Finkelstein, G. A. Jones, and M. Huisman (Eds.), *Professorial pathways: Academic careers in a global perspective* (pp. 265–290). Johns Hopkins University Press.
- Kallenberg, T. (2015). Academic middle managers shaping the landscape between policy and practice. In B. Jongbloed, J. Enders, and C. Salerno (Eds.), *Diversity and excellence in higher education* (pp. 201–216). Brill.
- Kallio, T. J., Kallio, K.-M., & Blomberg, A. (2020). From professional bureaucracy to competitive bureaucracy: Redefining universities' organization principles, performance measurement criteria, and reason for being. *Qualitative Research in Accounting and Management*, 17(1), 82–108. <https://doi.org/10.1108/QRAM-10-2019-0111>
- Kenny, J. (2018). Re-empowering academics in a corporate culture: An exploration of workload and performativity in a university. *Higher Education*, 75, 365–380.
- Kezar, A., De Paola, T., & Scott, D. T. (2019). *The gig academy: Reforming higher education: Innovation and the public good*. Johns Hopkins University Press.
- Kohtamäki, V., & Balbachevsky, E. (2018). University autonomy from the past to the present. *Theoretical and methodological perspectives*

*on higher education management and transformation: an advanced reader for PhD students*, 179-194.

- Kwiek, M. (2019). *Changing European academics*. Taylor and Francis.
- Labraña, J. y Brunner, J. J. (2022). La ideología de la nueva gestión pública desde la mirada de los directivos de las universidades chilenas: un estudio de casos múltiples. *Revista Iberoamericana de Educación Superior*, 13(38), 3–23.
- Lerch, J. C., Frank, D. J., & Schofer, E. (2023). The social foundations of academic freedom: Heterogeneous institutions in world society, 1960 to 2022. *American Sociological Review*, 89(1), 88–125.
- Leišytė, L., Jones, G. A., and Marquina, M. (2025). Trends in Higher Education Governance and the Academic Profession. In L. Leišytė, M. Marquina and G. A. Jones (Eds.), *University governance, management and the academic profession* (Vol. 26, pp. 3–10). Springer. [https://doi.org/10.1007/978-3-031-86889-4\\_1](https://doi.org/10.1007/978-3-031-86889-4_1)
- Maassen, P. (2024). Erosion of academic freedom in the European Union. *International Higher Education*, 120.
- Maassen, P., Gornitzka, Å., & Fumasoli, T. (2017). University reform and institutional autonomy: A framework for analyzing the living autonomy. *Higher Education Quarterly*, 71(3), 239–257.
- Marquina, M. (2020). Between the global and the local: The study of the academic profession from a Latin American perspective. *Higher Education Governance and Policy*, 1(1), 63–75.
- Marquina, M., Centeno, C. P., & Reznik, N. (2022). Academic power and institutional control of academia in Argentine public universities within the context of a managerial governance model. In S. Lysons and T. Burgess (Eds.), *Research handbook on academic careers and managing academics* (pp. 47–63). Edward Elgar Publishing.
- Marsiske, R. (2017). La autonomía universitaria en América Latina a 100 años del movimiento estudiantil de Córdoba: una agenda de investigación desde México. *Universidades*, 72, 27–35.
- Mendoza, P. (2020). Autonomy and weak governments: Challenges to university quality in Latin America. *Higher Education*, 80, 719–737.
- Mignot-Gérard, S., Sponem, S., Chatelain-Ponroy, S., & Musselin, C. (2023). Kaleidoscopic collegiality and the use of performance research metrics: The case of French universities. *Higher Education*, 85(4), 887–918.
- Musselin, C. (2007). The transformation of academic work: Facts and analysis. *Research y Occasional Paper Series: CSHE.4.07*. Center for Studies in Higher Education.

- Ordorika Sacristán, I. (2010). La autonomía universitaria: una perspectiva política. *Perfiles Educativos*, 32 (número especial), 79–81.
- Pekkola, E., Carvalho, T., Siekkinen, T., and Johansson, J. E. (2018). The sociology of professions and the study of the academic profession. *ResearchGate*. <https://www.researchgate.net/publication/328107248>
- Pérez Mora, R., Castañeda Bernal, X. Y. y Inguanzo Arias, B. L. (2022). El trabajo científico: Entre la racionalidad económico-burocrática y la racionalidad ético-política. *Analecta Política*, 12(23), 1–26. <https://doi.org/10.18566/apolit.v12n23.a05>
- Sisto, V. (2020). Overwhelmed: Accountability and work intensification in the neo-liberal university. The case of Chile. *Education Policy Analysis Archives*, 28, 7. <https://doi.org/10.14507/epaa.28.4907>
- Slaughter, S., & Leslie, L. L. (1997). *Academic capitalism: Politics, policies, and the entrepreneurial university*. Johns Hopkins University Press.
- Stephenson, G. K., Jones, G. A., Bégin-Caouette, O., & Metcalfe, A. S. (2025). The paradox of increasing managerialism amid historic decentralization: Canadian professors and university governance. In L. Leišytė, M. Marquina and G. A. Jones (Eds.), *University governance, management and the academic profession* (Vol. 26, pp. 71–91). Springer. [https://doi.org/10.1007/978-3-031-86889-4\\_5](https://doi.org/10.1007/978-3-031-86889-4_5)
- Teichler, U., Aarrevaara, T., & Huang, F. (2022). Conclusion: What we know about the teaching-research nexus in the knowledge-based society. In F. Huang, T. Aarrevaara, and U. Teichler (Eds.), *Teaching and research in the knowledge-based society: Historical and comparative perspectives* (pp. 241–253). Springer.
- Valadés, D. (2015). Autonomía y constitución en América Latina—*Unión de Universidades de América Latina y el Caribe*.
- Villar, A. y Ibarra, A. (2014). La autonomía universitaria, una mirada latinoamericana. *UDUAL*.
- Wen, W., & Marginson, S. (2023). Governance in Chinese universities. In R. W. Watermeyer and D. Knight (Eds.), *University collegiality and the erosion of faculty authority* (pp. 171–197). Emerald Publishing.



# Adherence to the PRISMA Statement in meta-Analyses reporting experimental interventions published in Education: A Systematic Meta-Review

## Adherencia a la declaración PRISMA en los meta-análisis de intervenciones experimentales publicados en Educación: una meta-revisión sistemática<sup>1</sup>

<https://doi.org/10.4438/1988-592X-RE-2025-412-736>

### Micaela Sánchez-Martín

<https://orcid.org/0000-0002-9107-574X>

Universidad de Murcia. IMIB\*

### Marta Gutiérrez-Sánchez

<https://orcid.org/0000-0001-7302-9283>

Universidad de Murcia

### Eva María Olmedo-Moreno

<https://orcid.org/0000-0003-0558-1513>

Universidad de Granada

### Fernando Navarro-Mateu

<https://orcid.org/0000-0002-3228-623X>

Universidad de Murcia. IMIB\*

### Abstract

**Objective:** To assess adherence to the PRISMA statement in meta-analyses of educational interventions aimed at improving academic performance between 2009 and 2022. **Method:** Systematic review. Eligibility criteria: Meta-analyses of experimental studies evaluating educational interventions designed to improve academic performance, published in English or Spanish between January 1, 2009, and April 30, 2022. Exclusion criteria: Other designs or outcomes, and inaccessi-

<sup>1</sup> Instituto Murciano de Investigación Biosanitaria Pascual Parrilla

ble publications. Bibliographic search: Conducted in four databases (ERIC, Web of Science, Scopus, and PubMed). Data extraction: Based on a previously developed and registered protocol. Data: Adherence to PRISMA recommendations, risk of bias according to AMSTAR 2, and various characteristics of published studies. Statistical analyses: Chi-square test or Fisher's exact test. As measures of association, odds ratios and  $\beta$  were calculated, with 95% confidence intervals (95% CI) and p-values through logistic and linear regression analyses. **Results:** Out of 2,076 identified studies, 69 were included. The mean PRISMA score was 19.7 (SD = 4.4) out of 27. A total of 51.8% (n = 14) of the recommendations had adherence rates above 75%. Significant differences were found in adherence to the recommendations depending on whether the authors declared compliance with PRISMA in 14 of the 27 items (51.8%). Compliance with 13 of the 27 recommendations (48.1%) was associated with a lower risk of bias. PRISMA adherence was associated with lower risk of bias, more recent publication, health-related education, and interventions involving physical activities in higher education. **Conclusions:** Adherence to PRISMA in published educational meta-analyses of experimental interventions to improve academic performance is clearly improvable. **Funding:** Research Project PID2020-119194RB-I00, funded by MCIN / AEI/10.13039/501100011033. **Protocol:** Registered on Open Science Framework: <https://doi.org/10.17605/OSF.IO/PSKN6>

*Key words:* Meta-review, Systematic Review, Meta-analysis, PRISMA, educational interventions, academic performance.

## Resumen

**Objetivo:** Evaluar la adherencia a la declaración PRISMA de los estudios de meta-análisis sobre intervenciones educativas orientadas a mejorar el rendimiento académico entre 2009 y 2022. **Método:** Revisión sistemática. **Criterios de selección:** metaanálisis de estudios experimentales evaluando intervenciones educativas diseñadas para mejorar el rendimiento académico publicadas en inglés o español entre el 1 de enero de 2009 hasta el 30 de abril de 2022. Criterios de exclusión: otros diseños o resultados y publicaciones no accesibles. Búsqueda bibliográfica en cuatro bases de datos (ERIC, Web of Science, Scopus y PubMed). Extracción de datos mediante protocolo previamente elaborado y registrado. **Datos:** adherencia a las recomendaciones PRISMA, riesgo de sesgos según la AMSTAR 2, y diversas características de los estudios publicados. Análisis estadísticos: Chi-cuadrado o prueba exacta de Fisher. Como medida de asociación, se calculó las odds ratios y  $\beta$ , con sus intervalos de confianza del 95% (IC95%) y p-valores mediante análisis de regresión logística y lineal. **Resultados:** De 2076 estudios identificados se seleccionaron 69. La puntuación media de la PRISMA fue de 19.7 (SD=4.4) sobre 27. Un 51.8% (n=14) de las recomendaciones tuvo una adherencia superior al 75%. Se encontraron diferencias significativas en el cumplimiento de las diversas recomendaciones en los estudios según los autores hayan declarado su adherencia a la PRISMA en 14 de las 27 recomendaciones

(51.8%). El seguimiento de 13 de las 27 recomendaciones (48.1%) se asociaron a un menor riesgo de sesgos. La adherencia a la PRISMA se asoció a un menor riesgo de sesgos, a una publicación más reciente, a la educación en el área sanitaria y a intervenciones con actividades físicas orientadas a la educación superior. **Conclusiones:** La adherencia a la PRISMA en los metaanálisis publicados en educación con intervenciones experimentales para aumentar el rendimiento académico son claramente mejorables.

**Financiación:** Proyecto de Investigación PID2020-119194RB-I00, financiado por MCIN /AEI/10.13039/501100011033.

**Protocolo:** registrado en Open Science Framework: <https://doi.org/10.17605/OSF.IO/PSKN6>

*Palabras clave:* Meta-revisión, Revisión Sistemática, Meta-análisis, PRISMA, intervenciones educativas, rendimiento académico.

## Introduction

In recent decades, the number of scientific publications has grown exponentially across all areas of knowledge, including the field of education (Ahn et al., 2012; Ruiz-Corbella et al., 2014). This proliferation, together with the need to efficiently identify and apply the best scientific evidence in educational decision-making, has highlighted the importance of systematic reviews (SRs) and meta-analyses (MAs) as essential tools in the development of Evidence-Based Education (EBE) (Hederich et al., 2014; Sánchez-Martín et al., 2022). Unlike traditional narrative reviews, SRs and MAs follow a defined scientific method to systematically and reproducibly identify, assess, and synthesize the literature on a given topic. Since Gene V. Glass introduced the term “meta-analysis” in the educational field in 1976, its use has grown significantly within education (Ahn et al., 2012). MAs are systematic reviews that employ statistical techniques to calculate an overall effect size and to assess issues such as publication bias and heterogeneity.

This increase in the number of publications has been accompanied by a growing concern regarding their methodological quality and has fostered the development of studies focused on the methodological assessment of scientific literature, also known as meta-research (Ioannidis, 2018; Ioannidis et al., 2015). This trend has also emerged in the field of education (Anguera, 2023; Blanco-Blanco, 2018; Cook et al., 2011), driven by the need for scientific knowledge to be based on rigorous, reproducible studies with valid results that can be applied to educational practice. This research, also referred to as meta-reviews or methodological reviews, are considered a special type of umbrella review. Their main objective is to assess the quality of the research itself, beyond the findings being

analyzed. This type of study focuses on identifying areas for improvement in methodological quality and in the transparency of published reports.

In medical education, several studies have evaluated the methodological quality of publications (Cook et al., 2011; Howley et al., 2008), and at least three reports have addressed this issue in other educational fields (Ahn et al., 2012; Eser & Yurtçu, 2020; Sánchez-Martín et al., 2024a). Findings from these studies indicate that the methodological quality and risk of bias (RoB) in meta-analyses (MAs) and systematic reviews (SRs) in education do not meet desirable standards. This highlights an urgent need to enhance these designs to increase confidence in their findings. Furthermore, a lack of key information has been identified in many scientific publications. This deficiency is particularly striking considering the development of checklists specifically designed to improve the transparency and reproducibility of MAs and SRs, such as the PRISMA 2020 statement (Preferred Reporting Items for Systematic Reviews and Meta-Analyses) (Page et al., 2021a) and the Meta-Analysis Reporting Standards (MARS) (American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards, 2008; Appelbaum et al., 2018), among others (Rubio-Aparicio et al., 2018; Sánchez-Meca et al., 2021).

Although the use of the PRISMA guideline in the field of education has increased in recent years (Holmqvist & Lantz Ekström, 2024; Moreno et al., 2022), its impact has primarily been evaluated within medical education. Several studies suggest that adherence to PRISMA during the drafting of a systematic review (SR) is associated with a lower risk of bias and, consequently, greater methodological quality (Panic et al., 2013; Sharma & Oremus, 2018). In the broader field of education, results of the first meta-review to date have recently been published (Sánchez-Martín et al., 2024a, 2025). This meta-review used two internationally recognized tools for assessing risk of bias and transparency and clarity of reporting—namely, AMSTAR 2 (A Measurement Tool to Assess Systematic Reviews) (Shea et al., 2017) and the PRISMA statement. This meta-review focused on published meta-analyses based on experimental studies evaluating educational interventions aimed at improving academic performance. The first analysis suggested that an explicit statement by the authors indicating the use of the PRISMA guideline was associated with a lower risk of bias (Sánchez-Martín et al., 2024a). In the second analysis, an explicit statement regarding PRISMA use increased the likelihood of including a flow diagram (Sánchez-Martín et al., 2025). This diagram graphically illustrates the search and selection process, and its presence in the manuscript was associated with a lower risk of bias. However, neither of these analyses assessed the actual completion of each individual PRISMA item,

nor whether the authors' explicit mention of PRISMA adherence was associated with genuine compliance with its recommendations. A previous study analyzing this aspect—without relying on explicit PRISMA mentions in the full text—found no differences in overall PRISMA adherence when comparing two groups of SRs and MAs, one explicitly citing its use and the other not (Panic et al., 2013).

In this third analysis, using the same sample as the two previous reports, the aim is to answer the question of actual adherence to the PRISMA statement in meta-analyses (MAs) of educational interventions aimed at improving academic performance between 2009 and 2022, and whether there are differences in compliance with each of the items between studies that explicitly mention the use of the PRISMA guideline and those that do not. Specifically, the aims were: i) To describe the current compliance with each of the PRISMA recommendations in selected meta-analyses (Objective 1); ii) To analyze the association between item-by-item compliance and the authors' explicit statement of having used the PRISMA guideline (Objective 2); iii) To assess the association between risk of bias (RoB) and compliance with each of the PRISMA recommendations (Objective 3); and iv) To identify characteristics associated with the authors' explicit statement of having used PRISMA (Objective 4a) and with the number of recommendations fulfilled (Objective 4b).

## Method

### Research Design

Meta-review or methodological systematic review designed to assess the transparency and quality of published meta-analyses focused on educational interventions targeting various academic outcomes (Sánchez-Martín et al., 2024a; Sánchez-Martín et al., 2025).

### Protocol and Registration

This study was written in accordance with the PRISMA 2020 guidelines (Page et al., 2021b). The protocol was registered in the Open Science Framework (OSF): <https://doi.org/10.17605/OSF.IO/PSKN6>

### Eligibility Criteria

The inclusion criteria were: i) meta-analyses (MAs) of experimental and quasi-experimental designs evaluating the effectiveness of educational interventions on academic performance outcomes; ii) written in English or Spanish; and iii) published between January 1, 2009, the publication date of the first PRISMA guideline (Moher et al., 2009), and April 30,

2022. The exclusion criteria were: i) MAs based on non-experimental designs, traditional narrative reviews, or other types of publications (e.g., conference proceedings or posters, book chapters, theoretical articles); ii) MAs evaluating the effects of educational interventions on outcomes other than academic performance (e.g., cognitive enhancement, social skills); and iii) articles not available in full-text format.

### **Information Sources and Database Search Strategies**

Searches were conducted in four databases ERIC, Web of Science (WoS), SCOPUS, and PubMed. The specific search strategies for each database have been previously published in detail (Sánchez-Martín et al., 2024a). In general terms, the search string was developed using the following keywords: “meta-analysis” AND “educational intervention” AND “academic outcomes”, along with synonyms and related terms, in both English and Spanish. Additionally, the reference lists of the selected articles were manually searched.

### **Study Selection and Data Extraction**

The references from the identified meta-analyses were imported into Zotero. One reviewer performed the initial screening, removing duplicates and clearly irrelevant records. Subsequently, two independent reviewers selected studies based on title and abstract, followed by full-text review according to the inclusion criteria. This process achieved a Cohen’s Kappa ( $\kappa$ ) of .869. Discrepancies were resolved by consensus and/or through consultation with a third reviewer.

A data extraction protocol was developed before data collection. It included variables such as: journal editors’ suggestion to adhere to PRISMA, PRISMA adherence, evaluation of individual PRISMA items, year of publication, search strategy (regional or national/international), country, continent, area of education (education, psychology, health), intervention type, target population characteristics (educational level), and risk of bias (RoB), assessed using the AMSTAR 2 tool. Two researchers independently extracted data using this predefined protocol. The initial inter-rater agreement was  $\kappa = .897$ . Disagreements were resolved through consensus and/or with input from a third reviewer.

### **PRISMA Statement**

The original version of the PRISMA statement (Moher et al., 2009), in its Spanish adaptation (Urrútia & Bonfill, 2010), was used. This version was chosen because it was the most widely available to the authors of the meta-analyses at the time their studies were conducted. This tool includes a 27-item checklist grouped into seven major sections: I) TITLE:

1. Title; II) ABSTRACT: 2. Structured summary; III) INTRODUCTION: 3. Rationale; 4. Objectives; IV) METHODS: 5. Protocol and registration; 6. Eligibility criteria; 7. Information sources; 8. Search; 9. Study selection; 10. Data collection process; 11. Data items; 12. Risk of bias in individual studies; 13. Summary measures; 14. Synthesis of results; 15. Risk of bias across studies; 16. Additional analyses; V) RESULTS: 17. Study selection; 18. Study characteristics; 19. Risk of bias within studies; 20. Results of individual studies; 21. Synthesis of results; 22. Risk of bias across studies; 23. Additional analyses; VI) DISCUSSION: 24. Summary of evidence; 25. Limitations; 26. Conclusions; VII) FUNDING: 27. Funding.

### **Risk of Bias Assessment**

Our previous study detailed the risk of bias (RoB) assessment using the AMSTAR 2 scale (Sánchez-Martín et al., 2024a). This earlier publication specifically described the AMSTAR 2 tool and the criteria applied for RoB evaluation in more detail. Briefly, a less restrictive approach than the original AMSTAR 2 evaluation criteria (Shea et al., 2017) was used. RoB was defined solely based on the number of critical domains, disregarding non-critical items. RoB was dichotomized as follows: “low risk” (up to two critical weaknesses) and “high risk” (three or more critical weaknesses). This adaptation followed the original author’s recommendations, which suggest the possibility of adjusting the evaluation criteria based on research objectives. This adjustment was necessary because 90% of the 69 included studies were rated as having critically low confidence in the overall AMSTAR 2 assessment when using the original criteria (Sánchez-Martín et al., 2024a).

### **Synthesis of Results**

Inter-rater agreement during both the study selection and data extraction processes was measured using Cohen’s Kappa index ( $\kappa$ ). The PRISMA 2020 flow diagram (Page et al., 2021a), along with the description of included studies and a list of excluded studies with reasons for exclusion, have been previously published (Sánchez-Martín et al., 2024a). An initial descriptive analysis was performed, reporting means and standard deviations for continuous variables, and frequencies and percentages for categorical variables. For instance, adherence to each PRISMA statement recommendation was evaluated as a dichotomous variable (Yes/No).

The chi-square test or Fisher’s exact test was used to analyze differences based on whether the authors explicitly reported PRISMA use, and the risk of bias associated with each individual PRISMA item (Objectives 2 and 3). To assess associations between study characteristics (independent variables) and the authors’ explicit statement of PRISMA use (dependent

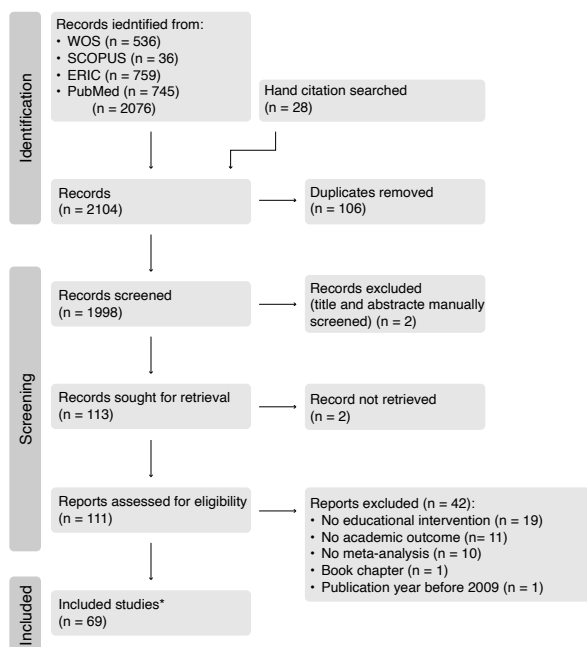
variable, Yes/No), odds ratios (ORs) with 95% confidence intervals (95% CIs) and corresponding p-values were calculated using independent simple logistic regression analyses (Objective 4a). Similarly, associations between the same study characteristics and the overall PRISMA adherence score (dependent variable) were assessed using independent simple linear regression models, reporting  $\beta$  coefficients, 95% CIs, and p-values (Objective 4b).

All statistical tests were two-tailed and used an alpha level of .05. Analyses were conducted using SPSS software (Version 28.0).

## Results

The study selection strategy has been described in detail in a previous publication (Sánchez-Martín et al., 2024a). The flow diagram corresponding to the systematic review is shown in Figure I. In summary, a total of 2,076 studies were initially identified, of which 69 met the inclusion criteria and were incorporated into the final analysis. The list of excluded articles and the rationale for their exclusion have been previously reported (Sánchez-Martín et al., 2024a).

**Figure I.** Flow diagram



\* Source: Sánchez Martín (2024b)

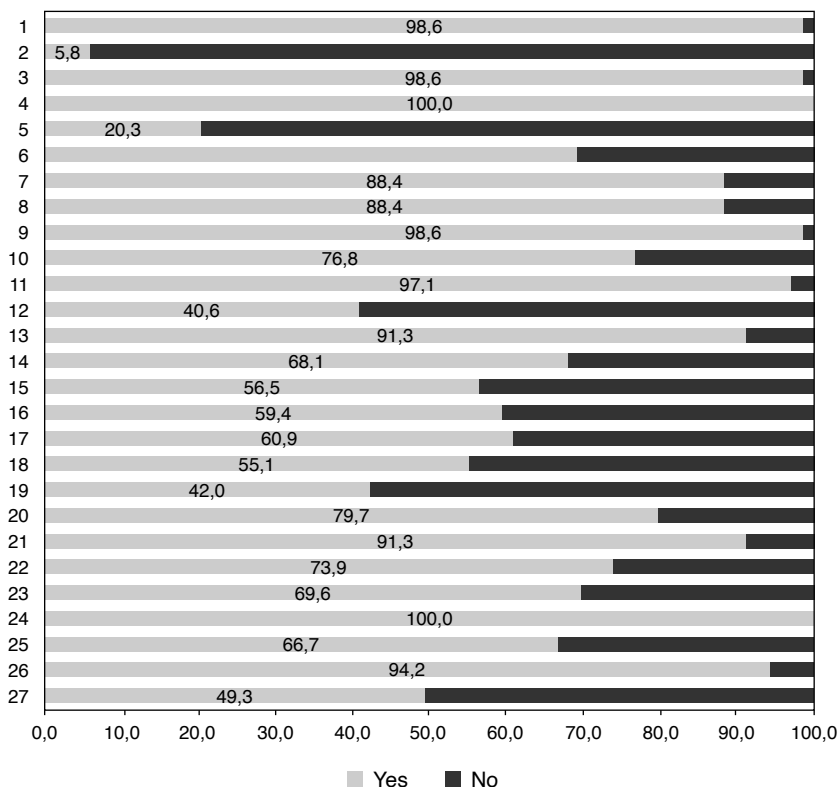
The general characteristics of the included studies are presented in Supplementary Table I. Regarding PRISMA adherence, only 24 studies (34.8%) explicitly mentioned using the guideline in the manuscript, despite the fact that 33 journals (47.8%) recommended it in their author guidelines. The mean PRISMA adherence score was 19.7 (SD = 4.4) out of a maximum of 27. Regarding risk of bias (RoB), only 21 studies (30.4%) were rated as low risk. Most studies were published from 2017 onwards, with the highest number of publications originating from the United States, Turkey, and China, in descending order. Publications were grouped into three main knowledge areas: i) General education, which comprised the majority of studies, ii) Educational interventions in health, and iii) Interventions in the field of psychology. The most commonly analyzed intervention type was methodological strategies. Regarding the target population, 34.8% of the studies (n = 24) addressed interventions targeting a combination of educational levels (primary, secondary or special education, and higher education), while 27.5% (n = 19) included both levels in their analysis.

Figure II and Table I present the overall results regarding adherence to the different PRISMA guideline recommendations (Objective 1), while Supplementary Table II provides adherence data for each individual PRISMA item, as well as the overall score for each included study. Overall, 51.8% of the recommendations (n = 14) showed adherence above 75%. Adherence ranged from a minimum of 6% to a maximum of 100%. The recommendation with the lowest adherence was the structured abstract (6%). In contrast, the recommendations with 100% compliance were: INTRODUCTION: Objectives (Item 4); METHODS: Eligibility criteria (Item 6); and DISCUSSION: Summary of the evidence (Item 24). A total of 18.5% of the recommendations (n = 5) had adherence rates below 50%, the lowest being: ABSTRACT: Structured abstract (Item 2); METHODS: Protocol and registration (Item 5); METHODS: Risk of bias in individual studies (Item 12); RESULTS: Risk of bias within studies (Item 19); and DISCUSSION: Funding (Item 27).

Regarding the second objective -to analyze the association between compliance with each PRISMA item and the authors' explicit statement of PRISMA use- statistically significant differences were found in 14 of the 27 PRISMA recommendations (51.8%) based on whether the authors explicitly declared adherence to the statement (Table I). Publications that reported PRISMA use showed significantly higher compliance with the following recommendations: ABSTRACT: 2. Structured summary; METHODS: 5. Protocol and registration; 7. Information sources; 8. Search; 12. Risk of bias in individual

studies; 15. Risk of bias across studies; RESULTS: 17. Study selection; 18. Study characteristics; 19. Risk of bias within studies; 20. Results of individual studies; 22. Risk of bias across studies; 23. Additional analyses; DISCUSSION: 25. Limitations; FUNDING: 27. Funding.

**Figure II.** Percentage of compliance with PRISMA statement recommendations among the included studies (n = 69)



TITLE: 1. Title; ABSTRACT: 2. Structured summary; INTRODUCTION: 3. Rationale; 4. Objectives; METHODS: 5. Protocol and registration; 6. Eligibility criteria; 7. Information sources; 8. Search; 9. Study selection; 10. Data collection process; 11. Data items; 12. Risk of bias in individual studies; 13. Summary measures; 14. Synthesis of results; 15. Risk of bias across studies; 16. Additional analyses; RESULTS: 17. Study selection; 18. Study characteristics; 19. Risk of bias within studies; 20. Results of individual studies; 21. Synthesis of results; 22. Risk of bias across studies; 23. Additional analyses; DISCUSSION: 24. Summary of evidence; 25. Limitations; 26. Conclusions; FUNDING: 27. Funding

The analysis of the association between risk of bias (RoB) and compliance with each PRISMA recommendation (Objective 3) revealed that

adherence to 13 of the 27 items (48.1%) was significantly associated with lower RoB (Table I). These items were: METHODS: 5. Protocol and registration; 10. Data collection process; 12. Risk of bias in individual studies; 14. Synthesis of results; 15. Risk of bias across studies; 16. Additional analyses; RESULTS: 17. Study selection; 18. Study characteristics; 19. Risk of bias within studies; 20. Results of individual studies; 22. Risk of bias across studies; DISCUSSION: 25. Limitations; FUNDING: 27. Funding. However, four items that showed significant differences based on PRISMA use declaration were not associated with differences in RoB: ABSTRACT: 2. Structured summary; METHODS: 7. Information sources; 8. Search; RESULTS: 23. Additional analyses. Conversely, three items that did not show differences based on PRISMA use declaration were associated with lower RoB. All of them belong to the METHODS section: 10. Data collection process; 14. Synthesis of results; 16. Additional analyses

Table II presents the results of the association analyses between study characteristics and two outcomes: authors' explicit reporting of PRISMA adherence (Objective 4a), and the number of PRISMA recommendations fulfilled (Objective 4b). Four characteristics were significantly associated with a higher likelihood of PRISMA being explicitly mentioned in the manuscript: i) a higher number of recommendations followed (PRISMA score), ii) lower risk of bias (RoB), iii) more recent publication years (2017–2022), and iv) studies focused on health education and interventions related to physical activity. On the other hand, seven characteristics were significantly associated with a higher number of fulfilled PRISMA recommendations: i) the journal's recommendation to use the PRISMA guideline in its author instructions, ii) explicit author declaration of PRISMA adherence, iii) lower RoB, iv) recent publication years (2017–2022), v) studies within the fields of psychology education and health education, vi) interventions related to physical activity, and vii) target populations composed of higher education students, compared to those focused on primary, secondary, or special education levels.

**Table 1.** Compliance with individual PRISMA statement recommendations according to reported use and associated risk of bias (RoB)

	Authors declare the use of the PRISMA statement										RoB #		p-value
	Overall PRISMA Adherence	Non Compliance		Yes Compliance		p-value	High Compliance		Low Compliance		n (%)	n (%)	
		n (%)	No	Yes	No		Yes	No	Yes	No			
<b>TITLE</b>													
1. Title	68 (98.6)	1 (2.2)	0 (0)	44 (97.8)	24 (100.0)	1.000*	1 (2.1)	47 (97.9)	0 (0.0)	21 (100.0)	1.000*		
<b>ABSTRACT</b>													
2. Structured summary	4 (5.8)	45 (100.0)	0 (0.0)	0 (83.3)	4 (16.7)	.012*	47 (97.9)	1 (2.1)	18 (84.7)	3 (14.3)	.081*		
<b>INTRODUCTION</b>													
3. Rationales	68 (98.6)	0 (0.0)	45 (100.0)	1 (4.2)	23 (95.8)	.348*	1 (2.1)	47 (97.9)	0 (0.0)	21 (100.0)	1.000*		
4. Objectives	69 (100.0)	0 (0.0)	45 (100.0)	0 (0.0)	24 (100.0)	1.000*	0 (0.0)	48 (100)	0 (0.0)	21 (100.0)	-		
<b>METHODS</b>													
5. Protocol and registration	14 (20.3)	41 (91.1)	4 (8.9)	14 (58.3)	10 (41.7)	.003*	44 (91.7)	4 (8.3)	11 (52.4)	10 (47.6)	<.001*		
6. Eligibility criteria	69 (100.0)	0 (0.0)	45 (100.0)	0 (0.0)	24 (100.0)	-	0 (0.0)	48 (100.0)	0 (0.0)	21 (100.0)	-		
7. Information sources	61 (88.4)	8 (17.8)	37 (82.2)	0 (0.0)	24 (100.0)	.028*	8 (16.7)	40 (83.3)	0 (0.0)	21 (100.0)	.095		
8. Search	61 (88.4)	8 (17.8)	37 (82.2)	0 (0.0)	24 (100.0)	.044*	7 (14.6)	41 (85.4)	1 (4.8)	20 (95.2)	.419		

	Authors declare the use of the PRISMA statement						RoB #				
	Overall PRISMA Adherence	Non Compliance		Yes Compliance		p-value	High Compliance		Low Compliance		p-value
		n (%)	No	Yes	n (%)		No	Yes	n (%)	No	
9. Study selection	68 (98.6)	1 (2.2)	44 (97.8)	24 (100.0)	1.000*	1 (2.1)	47 (97.9)	0 (0.0)	21 (100.0)	1.000*	
10. Data collection process	53 (76.8)	12 (26.7)	33 (73.3)	20 (83.3)	.349	15 (31.2)	33 (68.8)	1 (4.8)	20 (95.2)	.027	
11. Data items	67 (97.1)	2 (4.4)	43 (95.6)	24 (100.0)	.540	2 (4.2)	46 (95.8)	0 (0.0)	21 (100.0)	1.000*	
12. Risk of bias in individual studies	28 (40.6)	35 (77.8)	10 (22.2)	18 (75.0)	<.001	39 (81.2)	9 (18.8)	2 (9.5)	19 (90.5)	<.001	
13. Summary measures	63 (91.3)	4 (8.9)	41 (91.1)	22 (91.7)	1.000*	3 (6.2)	45 (93.8)	3 (14.3)	18 (85.7)	.359	
14. Synthesis of results	47 (68.1)	17 (37.8)	28 (62.2)	19 (79.2)	.150	21 (43.8)	27 (56.2)	1 (4.8)	20 (95.2)	.001	
15. Risk of bias across studies	39 (56.5)	25 (55.6)	20 (44.4)	19 (79.2)	.006	27 (56.2)	21 (43.8)	3 (14.3)	18 (85.7)	.001	
16. Additional analyses	41 (59.4)	22 (48.9)	23 (51.1)	18 (75.0)	.054	24 (50.0)	24 (50.0)	4 (19.0)	17 (81.0)	.016	
<b>RESULTS</b>											
17. Study selection	42 (60.9)	26 (57.8)	19 (42.2)	23 (95.8)	<.001	26 (54.2)	22 (45.8)	1 (4.8)	20 (95.2)	<.001	
18. Study characteristics	38 (55.1)	27 (60.0)	18 (40.0)	20 (83.3)	.001	26 (54.2)	22 (45.8)	5 (238)	16 (76.2)	.020	
19. Risk of bias within studies	29 (42.0)	36 (80.0)	9 (20.0)	20 (83.3)	<.001	38 (79.2)	10 (20.8)	2 (9.5)	19 (90.5)	<.001	

Authors declare the use of the PRISMA statement										RoB #	
Overall PRISMA Adherence	Non Compliance		Yes Compliance		p-value	High Compliance		Low Compliance		p-value	
	n (%)	n (%)	No	Yes		No	Yes	No	Yes		
20. Results of individual studies	55 (79.7)	14 (31.1)	31 (69.9)	24 (100.0)	.001*	13 (27.1)	35 (72.9)	1 (4.5)	20 (95.2)	.049*	
21. Synthesis of results	63 (91.3)	6 (13.3)	39 (86.7)	24 (100.0)	.085*	6 (12.5)	42 (87.5)	0 (0.0)	21 (100.0)	.167*	
22. Risk of bias across studies	51 (73.9)	15 (33.3)	30 (66.7)	21 (87.5)	.061	3 (12.5)	17 (35.4)	1 (4.8)	20 (95.2)	.008	
23. Additional analyses	48 (69.6)	18 (40.0)	27 (60.0)	21 (87.5)	.018	3 (12.5)	18 (37.5)	3 (14.3)	18 (85.7)	.054	
<b>DISCUSSION</b>											
24. Summary of evidence	69 (100.0)	0 (0.0)	45 (100.0)	24 (100.0)	-	0 (0.0)	48 (100)	0 (0.0)	21 (100.0)	-	
25. Limitations	46 (66.7)	21 (46.7)	24 (53.3)	22 (91.7)	.001	2 (8.3)	20 (41.7)	3 (14.3)	18 (85.7)	.026	
26. Conclusions	65 (94.2)	3 (6.7)	42 (93.3)	23 (95.8)	1.000*	1 (4.2)	3 (6.2)	1 (4.8)	20 (95.2)	1.000*	
<b>FUNDING</b>											
27. Funding	34 (49.3)	31 (68.9)	14 (31.1)	20 (83.3)	<.001	4 (16.7)	31 (64.6)	4 (19.0)	17 (81.0)	<.001	

\* Fisher's exact test; # RoB (Risk of Bias): Low risk (< 2 critical domains); High risk (> 2 critical domains). Source: Own elaboration.

**Table II.** Association between study characteristics and the authors' declaration of PRISMA adherence and the number of PRISMA items fulfilled

	Authors report following PRISMA guidelines					N° of PRISMA recommendations					
	No N (%)	Yes N (%)	OR	Lower 95%CI	Upper 95%CI	p-value	N° (SD)	β	Lower	Upper	p-value
The journal recommends the use of the PRISMA statement											
No	26 (57.8)	10 (41.7)	Ref				18.2 (4.4)	Ref			
Yes	19 (42.2)	14 (58.3)	1.9	.7	5.2	.269	21.3 (3.7)	3.1	1.1	5.1	.003
Authors state reporting according to the PRISMA statement											
No	-	-					17.7 (3.8)	Ref			
Yes	-	-	Ref				23.4 (2.4)	5.7	4.0	7.5	<.001
N° of PRISMA recommendations (SD)	17.7 (3.8)	23.5 (2.4)	1.8	1.3	2.3	<.001	-				
RoB											
High	37 (82.2)	11 (45.8)	Ref				17.8 (3.8)	Ref			
Low	8 (17.8)	13 (54.2)	5.5	1.8	16.6	.003	23.9 (2.0)	6.8	4.3	7.8	<.001
Year of publication											
2009-2016	26 (57.8)	5 (20.8)	Ref				18.3 (4.1)	Ref			

	Authors report following PRISMA guidelines					N° of PRISMA recommendations					
	95%CI					95%CI					
	No N (%)	Yes N (%)	OR	Lower	Upper	p-value	N° (SD)	β	Lower	Upper	p-value
2017-2022	19 (42.2)	19 (79.2)	5.2	1.6	16.4	.005	20.9 (4.3)	2.6	.5	4.6	.014
<b>Search strategy with geographical restrictions</b>											
Regional or national	11 (24.4)	2 (8.3)	Ref				17.8 (5.1)	Ref			
International	34 (75.6)	22 (91.7)	3.6	.7	17.6	.120	20.2 (4.1)	2.4	-2	5	.075
<b>Continents</b>											
Asia and Australia	22 (48.9)	10 (41.7)	Ref				18.7 (5.0)	Ref			
Europe	10 (22.2)	8 (33.3)	1.8	.5	5.8	.353	21.6 (3.4)	2.9	.4	5.4	.023
EEUU	13 (28.9)	6 (25.0)	1.0	.3	3.4	.980	19.6 (3.5)	0.9	-1.5	3.4	.447
<b>Educational area</b>											
Education	35 (77.8)	8 (33.3)	Ref				17.6 (4.0)	Ref			
Psychology	6 (13.3)	2 (8.3)	1.5	.2	8.6	.677	21.0 (1.9)	3.4	.8	6.0	.012
Health	4 (8.9)	14 (58.3)	15.3	4.0	59.1	<.001	24.1 (1.8)	3.2	1.3	8.4	<.001
<b>Type of intervention</b>											

	Authors report following PRISMA guidelines					N° of PRISMA recommendations					
	95%CI					95%CI					
	No N (%)	Yes N (%)	OR	Lower	Upper	p-value	N° (SD)	β	Lower	Upper	p-value
Methodological strategies	21 (46.7)	8 (37.5)	Ref				19.1 (4.1)	Ref			
Organizational strategies	10 (22.2)	7 (29.2)	1.6	.5	5.7	.439	19.8 (4.9)	0.7	-1.8	3.3	.577
Physical activities and others	3 (6.7)	7 (29.2)	5.4	1.1	25.9	.033	22.8 (2.9)	3.7	.6	6.8	.020
Combined	11 (24.4)	1 (4.2)	.2	.0	1.9	.165	18.5 (4.5)	-0.6	-3.5	2.3	.681
<b>Target population</b>											
Primary, secondary, and functional diversity education	6 (13.6)	4 (16.7)	Ref				18.0 (5.2)	Ref			
Higher education	6 (13.6)	9 (37.5)	2.6	.5	13.1	.239	11.7 (3.8)	3.9	.7	7.1	.019
All educational levels included	16 (36.4)	3 (12.5)	.3	.1	1.9	.328	17.6 (4.0)	-0.7	-3.8	2.4	.638
Combination of levels	16 (36.4)	8 (33.3)	.9	.2	3.9	.861	20.4 (3.8)	2.0	-1.0	5.0	.182

Source: Own elaboration

## Conclusions

To our knowledge, this is the first study in education to analyze the transparency and quality of published meta-analyses (MAs) of experimental studies. We assessed this through adherence to the PRISMA statement, and its relationship with risk of bias (RoB), measured using the AMSTAR 2 tool. Both instruments, PRISMA (Moher et al., 2009) and AMSTAR 2 (Shea et al., 2017) are widely accepted across various knowledge fields. Our results suggest that PRISMA adherence in educational MA publications is suboptimal (34.8%). Furthermore, MAs that explicitly declare and comply with PRISMA recommendations are generally associated with higher methodological quality and lower RoB.

Very few studies have examined the methodological quality of MAs in education. Only two independent investigations have been identified (Ahn et al., 2012; Eser & Yurtçu, 2020). Both used author-developed rating scales rather than standardized instruments, which limits the comparability of their results across studies and disciplines. While neither of these studies evaluated adherence to the PRISMA statement, both highlighted deficiencies in the reporting included studies and recommended greater transparency in data presentation.

In medical education, studies assessing the quality of research designs, such as experimental trials, have been published. These studies often used either author-developed scales (Cook et al., 2007) or checklists like STROBE (Strengthening the Reporting of Observational Studies in Epidemiology) (Cook et al., 2011). They consistently concluded that the quality and transparency of publications could be substantially improved. Assessing PRISMA adherence in systematic reviews has been more common in medical education. Findings showed lower-than-desirable usage levels (Cullis et al., 2017; Innocenti et al., 2022; Javidan et al., 2023; Liu et al., 2017; Panic et al., 2013; Sun et al., 2019, 2021; Wasiaak et al., 2017; Yang et al., 2024). A methodological study by the PRISMA developers (Page & Moher, 2017) found that only 57% of 100 meta-analyses indexed in MEDLINE adhered to any of its items or extensions. In exceptional cases, adherence has reached 86% (Sharma & Oremus, 2018).

Furthermore, some studies have reported improvements in publication quality following the introduction of the PRISMA statement (Liu et al., 2019). Evaluations have also been conducted on adherence to the MARS (Meta-Analysis Reporting Standards), developed by the American Psychological Association to enhance the reporting of meta-analyses in social sciences and education (American Psychological Association Publications and Communications Board Working Group on Journal Article

Reporting Standards, 2008; Appelbaum et al., 2018). However, insufficient compliance levels have also been found in industrial and organizational psychology (Kepes et al., 2013; Schalken & Rietbergen, 2017), as well as, more broadly, in psychology (Hohn et al., 2019).

Our analysis revealed significant variability in publication transparency and quality, consistent with previous findings in education (Ahn et al., 2012; Eser & Yurtçu, 2020). However, direct comparisons are challenging due to differences in the instruments used. Additionally, reporting practices vary between educational and medical journals (Blanco-Blanco, 2018). For instance, the recommendation to include a structured abstract (item 2) showed low adherence in our study (5.8%), in contrast to medical literature, where adherence ranges from 59% to 79% (Page & Moher, 2017). Among the five least frequently followed recommendations (ABSTRACT: 2. Structured summary; METHODS: 5. Protocol and registration; 12. RoB in individual studies; RESULTS: 19. RoB in studies; and DISCUSSION: 27. Funding), only the structured summary did not show a statistically significant association with RoB. A possible explanation is that many education journals do not require structured abstracts in their author guidelines.

The recommendation with the second-lowest adherence rate (20.3%) is item 5, regarding study protocol and registration. This is considered an essential component for reducing publication bias and improving transparency. In medical education, adherence to this item reaches 21% in studies published after 2009 (Page & Moher, 2017). Initiatives such as PROSPERO (International Prospective Register of Systematic Reviews) and the OSF (Open Science Framework) have been promoted to support protocol registration and enhance the credibility of systematic reviews (Page et al., 2021b), and journals such as Systematic Reviews have even started publishing study protocols prior to final results (<https://systematicreviewsjournal.biomedcentral.com/>). Finally, systematic reviews with registered protocols have demonstrated greater adherence to PRISMA (Yang et al., 2024).

The assessment of risk of bias (RoB) in the included studies—both in the description of the method used (item 12) and in the presentation of results (item 19)—is a key factor associated with methodological quality. In our study, adherence to these items was 40.6% and 42.0%, respectively, and their fulfillment was associated with lower RoB. In contrast, healthcare studies showed higher adherence levels, reaching 65% for Item 12 and 60% for Item 19 (Page & Moher, 2017). A detailed explanation of how risk of bias (RoB) was assessed, including the tools used and its consideration during result synthesis, is crucial for evaluating the quality

of the meta-analysis outcomes. In recent years, several tools have been developed specifically to assess RoB across various research designs. These include AMSTAR 2 for systematic reviews, the NOS (Newcastle-Ottawa Scale) for observational studies (Stang, 2010), and the MMAT (Mixed Methods Appraisal Tool) for quantitative, qualitative, or mixed-methods studies (Hong et al., 2018). The evaluation of RoB across studies (items 15 and 22) describes the methods used and the results obtained when assessing any risk that could threaten the cumulative evidence, such as publication bias (Thornton & Lee, 2000).

This study has certain limitations. These include restricting the inclusion criteria to meta-analyses (MAs) of experimental interventions in education aimed at improving academic performance. This criterion was adopted due to the growing number of MAs published in the education field and the anticipated heterogeneity among them. The decision to focus on interventions targeting academic performance narrowed the scope to a clearly defined outcome. Future meta-reviews of MAs and systematic reviews (SRs) centered on other relevant outcomes or using different research designs may offer a more comprehensive analysis of the issue and enable broader generalization of findings to other areas of education. Additionally, since the aim was to assess the quality of published studies, the search was limited to four databases, excluding gray literature and direct contact with key authors. Furthermore, RoB categorization was based on justified criteria by classifying AMSTAR 2 items according to the number of critical domains.

The PRISMA statement (Moher et al., 2009) was developed to provide a checklist of recommendations. These recommendations aim to improve the quality and transparency of reporting in systematic reviews (SRs) and meta-analyses (MAs), guiding researchers, editors, and peer reviewers. This statement is continuously evolving. In addition to its recent update, PRISMA 2020 (Page et al., 2021b), several extensions have been developed, such as PRISMA-P 2015 for protocols (Moher et al., 2015). The EQUATOR Network (Enhancing the QUALity and Transparency Of health Research, <https://www.equator-network.org/>) facilitates access to these tools and promotes their use among researchers and journal editors (Sánchez-Martín et al., 2024b; Struthers et al., 2021). While many checklists originate from the health sciences, their implementation in education may also be beneficial (Pandis & Fedorowicz, 2011; Sánchez-Martín et al., 2023). Replicability and transparency in educational research require joint efforts from researchers, editors, and science policy makers. Initiatives such as GoodReports (<https://www.goodreports.org/>) aim to promote the use of these tools from the early stages of scientific

writing. Moreover, international journals like *The Lancet* and *PLOS* have begun to mandate the use of reporting checklists. In Spain, journals, such as *ESPIRAL. Cuadernos del Profesorado* and *Educación XXI*, have adopted similar strategies.

In summary, the growth in scientific publications underscores the need for SRs and MAs as key tools for advancing knowledge. In education, ensuring the quality of these studies is essential. This study evaluates adherence to PRISMA guidelines in MAs on educational interventions aimed at improving academic performance between 2009 and 2022. The results suggest that adherence to PRISMA remains low. Articles following its guidelines are generally associated with a lower risk of bias. To strengthen educational research, it is crucial to promote meta-reviews. These can identify current methodological limitations and encourage strategies to genuinely increase adherence to the best available recommendations, such as PRISMA 2020. This will improve the transparency and quality of research publications.

### **Funding**

This study was carried out within the framework of the research project “Transformation of learning in hybrid contexts for the educational and employment inclusion of especially vulnerable groups, with emphasis on UFM” (TYNDALL/UFM), reference: PID2020-119194RB-I00, funded by MCIN/AEI/10.13039/501100011033.

### **Declaration of Interests**

The authors declare no conflict of interest.

### **Data Availability Statement**

The protocol was previously registered on the Open Science Framework (OSF): <https://doi.org/10.17605/OSF.IO/PSKN6>. All data used in this study are included in the article.

### **Author Contributions**

MSM and FNM designed the study. MSM and MGS independently participated in the selection process and data extraction. EOM and FNM served as third reviewers when needed. MSM and FNM conducted the data analysis, interpreted the results, and drafted the manuscript. All authors critically reviewed the manuscript and approved the final version for publication.

## Bibliography

- Ahn, S., Ames, A. J., & Myers, N. D. (2012). A Review of Meta-Analyses in Education: Methodological Strengths and Weaknesses. *Review of Educational Research*, 82(4), 436-476. <https://doi.org/10.3102/0034654312458162>
- American Psychological Association Publications and Communications Board Working Group on Journal Article Reporting Standards. (2008). Reporting standards for research in psychology: Why do we need them? What might they be? *Am. Psychol.*, 63, 839-851. <https://doi.org/10.1037/0003-066X.63.9.839>
- Anguera, M. T. (2023). Revisiting Systematic Reviews from a Methodological Perspective [Revisitando las revisiones sistemáticas desde la perspectiva metodológica]. *RELIEVE-Revista Electrónica de Investigación y Evaluación Educativa*, 29(1), art. M4. <https://doi.org/10.30827/relieve.v29i1.27758>
- Appelbaum, M., Cooper, H., Kline, R. B., Mayo-Wilson, E., Nezu, A. M., & Rao, S. M. (2018). Journal article reporting standards for quantitative research in psychology: The APA Publications and Communications Board task force report. *American Psychologist*, 73(1), 3-25. <https://doi.org/10.1037/amp0000191>
- Blanco-Blanco, Á. (2018). The State of Scientific Practices and Educational Research: Possible Challenges for the Next Decade [Estado de las prácticas científicas e investigación educativa. Posibles retos para la próxima década]. *Revista de Educación*, 381, 207-230. <https://doi.org/10.4438/1988-592X-RE-2017-381-386>
- Cook, D. A., Beckman, T. J., & Bordage, G. (2007). Quality of reporting of experimental studies in medical education: A systematic review. *Medical Education*, 41(8), 737-745. <https://doi.org/10.1111/j.1365-2923.2007.02777.x>
- Cook, D. A., Levinson, A. J., & Garside, S. (2011). Method and reporting quality in health professions education research: A systematic review. *Medical Education*, 45(3), 227-238. <https://doi.org/10.1111/j.1365-2923.2010.03890.x>
- Cullis, P. S., Gudlaugsdottir, K., & Andrews, J. (2017). A systematic review of the quality of conduct and reporting of systematic reviews and meta-analyses in paediatric surgery. *PloS One*, 12(4), e0175213. <https://doi.org/10.1371/journal.pone.0175213>
- Eser, M. T., & Yurtçu, M. (2020). A Review of Meta-Analysis Articles in Educational Sciences Conducted Between 2010 and 2019 in Turkey.

- International Journal of Progressive Education*, 16(6), 15-32. <https://doi.org/10.29329/ijpe.2020.280.2>
- Glass, G. V. (1976). Primary, Secondary, and Meta-Analysis of Research. *Educational Researcher*, 5(10), 3-8. <https://doi.org/10.2307/1174772>
- Hederich, C., Martínez, J., & Rincón, L. (2014). Towards Evidence-Based Education [Hacia una educación basada en la evidencia]. *Revista Colombiana de Educación*, 66, 19-54. <https://doi.org/10.17227/01203916.66rce19.54>
- Hohn, R. E., Slaney, K. L., & Tafreshi, D. (2019). Primary Study Quality in Psychological Meta-Analyses: An Empirical Assessment of Recent Practice. *Frontiers in Psychology*, 9, 2667. <https://doi.org/10.3389/fpsyg.2018.02667>
- Holmqvist, M., & Lantz Ekström, M. (2024). A systematic review of research on educational superintendents. *Cogent Education*, 11(1), 2307142. <https://doi.org/10.1080/2331186X.2024.2307142>
- Hong, Q. N., Gonzalez-Reyes, A., & Pluye, P. (2018). Improving the usefulness of a tool for appraising the quality of qualitative, quantitative and mixed methods studies, the Mixed Methods Appraisal Tool (MMAT). *Journal of Evaluation in Clinical Practice*, 24(3), 459-467. <https://doi.org/10.1111/jep.12884>
- Howley, L., Szauter, K., Perkowski, L., Clifton, M., McNaughton, N., & Association of Standardized Patient Educators (ASPE). (2008). Quality of standardised patient research reports in the medical education literature: Review and recommendations. *Medical Education*, 42(4), 350-358. <https://doi.org/10.1111/j.1365-2923.2007.02999.x>
- Innocenti, T., Feller, D., Giagio, S., Salvioli, S., Minnucci, S., Brindisino, F., Cosentino, C., Piano, L., Chiarotto, A., & Ostelo, R. (2022). Adherence to the PRISMA statement and its association with risk of bias in systematic reviews published in rehabilitation journals: A meta-research study. *Brazilian Journal of Physical Therapy*, 26(5), 100450. <https://doi.org/10.1016/j.bjpt.2022.100450>
- Ioannidis, J. P. A. (2018). Meta-research: Why research on research matters. *PLOS Biology*, 16(3), e2005468. <https://doi.org/10.1371/journal.pbio.2005468>
- Ioannidis, J. P. A., Fanelli, D., Dunne, D. D., & Goodman, S. N. (2015). Meta-research: Evaluation and Improvement of Research Methods and Practices. *PLoS Biology*, 13(10), e1002264. <https://doi.org/10.1371/journal.pbio.1002264>
- Javidan, A., Alaichi, J., Nassar, Y., Li, A., Balta, K. Y., & Najj, F. (2023). Completeness of reporting in systematic reviews and meta-analyses

- in vascular surgery. *Journal of Vascular Surgery*, 78(6), 1550-1558.e2. <https://doi.org/10.1016/j.jvs.2023.04.009>
- Kepes, S., McDaniel, M. A., Brannick, M. T., & Banks, G. C. (2013). Meta-analytic Reviews in the Organizational Sciences: Two Meta-analytic Schools on the Way to MARS (the Meta-analytic Reporting Standards). *Journal of Business and Psychology*, 28(2), 123-143.
- Liu, H., Zhou, X., Yu, G., & Sun, X. (2019). The effects of the PRISMA statement to improve the conduct and reporting of systematic reviews and meta-analyses of nursing interventions for patients with heart failure. *International Journal of Nursing Practice*, 25(3), e12729. <https://doi.org/10.1111/ijn.12729>
- Liu, X., Kinzler, M., Yuan, J., He, G., & Zhang, L. (2017). Low Reporting Quality of the Meta-Analyses in Diagnostic Pathology. *Archives of Pathology & Laboratory Medicine*, 141(3), 423-430. <https://doi.org/10.5858/arpa.2016-0144-OA>
- Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *BMJ*, 339, b2535. <https://doi.org/10.1136/bmj.b2535>
- Moher, D., Shamseer, L., Clarke, M., Ghersi, D., Liberati, A., Petticrew, M., Shekelle, P., & Stewart, L. A. (2015). Preferred reporting items for systematic review and meta-analysis protocols (PRISMA-P) 2015 statement. *Systematic Reviews*, 4(1). <https://doi.org/10.1186/2046-4053-4-1>
- Moreno, E. M., Montilla-Arechabala, C., & Maldonado, M. A. (2022). Effectiveness and characteristics of programs for developing oral competencies at university: A systematic review. *Cogent Education*, 9(1), 2149224. <https://doi.org/10.1080/2331186X.2022.2149224>
- Page, M. J., McKenzie, J. E., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T., Loder, E. W., Mayo-Wilson, E., McDonald, S., ... Moher, D. (2021a). The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ*, 372(71). <https://doi.org/10.1136/bmj.n71>
- Page, M. J., & Moher, D. (2017). Evaluations of the uptake and impact of the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA) Statement and extensions: A scoping review. *Systematic Reviews*, 6(1), 263, 1-14. <https://doi.org/10.1186/s13643-017-0663-8>
- Page, M. J., Moher, D., Bossuyt, P. M., Boutron, I., Hoffmann, T. C., Mulrow, C. D., Shamseer, L., Tetzlaff, J. M., Akl, E. A., Brennan, S. E., Chou, R., Glanville, J., Grimshaw, J. M., Hróbjartsson, A., Lalu, M. M., Li, T.,

- Loder, E. W., Mayo-Wilson, E., McDonald, S., ... McKenzie, J. E. (2021b). PRISMA 2020 explanation and elaboration: Updated guidance and exemplars for reporting systematic reviews. *BMJ*, 372(160). <https://doi.org/10.1136/bmj.n160>
- Pandis, N., & Fedorowicz, Z. (2011). The international EQUATOR network: Enhancing the quality and transparency of health care research. *Journal of Applied Oral Science: Revista FOB*, 19(5). <https://doi.org/10.1590/s1678-77572011000500001>
- Panic, N., Leoncini, E., de Belvis, G., Ricciardi, W., & Boccia, S. (2013). Evaluation of the endorsement of the preferred reporting items for systematic reviews and meta-analysis (PRISMA) statement on the quality of published systematic review and meta-analyses. *PloS One*, 8(12), e83138. <https://doi.org/10.1371/journal.pone.0083138>
- Rubio-Aparicio, M., Sánchez-Meca, J., Marín-Martínez, F., & López-López, J. A. (2018). Guidelines for Reporting Systematic Reviews and Meta-analyses. *Anales de Psicología*, 34(2), 412. <https://doi.org/10.6018/analesps.34.2.320131>
- Ruíz-Corbella, M., Galán, A., & Diestro, A. (2014). Scientific Education Journals in Spain: Evolution and Future Perspectives [Las revistas científicas de Educación en España: Evolución y perspectivas de futuro]. *RELIEVE*, 20(2). <https://doi.org/10.7203/relieve.20.2.4361>
- Sánchez-Martín, M., Gutiérrez-Sánchez, M., Olmedo-Moreno, E. M. & Navarro-Mateu, F. (2025). Quality of flowcharts in meta-analyses in education: A systematic review. *Espiral Cuadernos del Profesorado*, 18(37), 89-110. <https://doi.org/10.25115/ecp.v18i37.10159>
- Sánchez-Martín, M., Gutiérrez-Sánchez, M., Olmedo-Moreno, E. M., & Navarro-Mateu, F. (2024a). A systematic review to evaluate the risk of bias of meta-analyses reporting experimental educational interventions focused on academic performance. *Cogent Education*, 11(1), 2385785. <https://doi.org/10.1080/2331186X.2024.2385785>
- Sánchez-Martín, M., Navarro-Mateu, F., & Sánchez-Meca, J. (2022). Systematic reviews and evidence-based education. *Espiral. Cuadernos del Profesorado*, 15(30), 108-120. <https://doi.org/10.25115/ecp.v15i30.7860> Autor... (2022)
- Sánchez-Martín, M., Olmedo Moreno, E. M., Gutiérrez-Sánchez, M., & Navarro-Mateu, F. (2024b). EQUATOR-Network: a roadmap to improve the quality and transparency of research reporting. *Espiral. Cuadernos del Profesorado*, 17(35). <https://doi.org/10.25115/ecp.v17i35.9529>
- Sánchez-Martín, M., Ponce, A., Plana, M., Ibáñez-López, F., & Navarro-Mateu, F. (2023). There are many lists of research, but... not all serve the

- same. *Espiral. Cuadernos del Profesorado.*, 16(33), 81-91. <https://doi.org/10.25115/ecp.v16i33.9334>
- Sánchez-Meca, J., Boruch, R. F., & Petrosino, A. (2002). The Campbell Collaboration and Evidence-Based Practice [La Colaboración Campbell y la Práctica Basada en la Evidencia]. *Papeles del Psicólogo*, 83, 44-48.
- Sánchez-Meca, J., Marín-Martínez, F., López-López, J. A., Núñez-Núñez, R. M., Rubio-Aparicio, M., López-García, J. J., López-Pina, J. A., Blázquez-Rincón, D. M., López-Ibáñez, C., & López-Nicolás, R. (2021). Improving the reporting quality of reliability generalization meta-analyses: The REGEMA checklist. *Research Synthesis Methods*, 12(4), 516-536. <https://doi.org/10.1002/jrsm.1487>
- Schalken, N., & Rietbergen, C. (2017). The Reporting Quality of Systematic Reviews and Meta-Analyses in Industrial and Organizational Psychology: A Systematic Review. *Frontiers in Psychology*, 8, 1395. <https://doi.org/10.3389/fpsyg.2017.01395>
- Sharma, S., & Oremus, M. (2018). PRISMA and AMSTAR show systematic reviews on health literacy and cancer screening are of good quality. *Journal of Clinical Epidemiology*, 99, 123-131. <https://doi.org/10.1016/j.jclinepi.2018.03.012>
- Shea, B. J., Reeves, B. C., Wells, G., Thuku, M., Hamel, C., Moran, J., Moher, D., Tugwell, P., Welch, V., Kristjansson, E., & Henry, D. A. (2017). AMSTAR 2: A critical appraisal tool for systematic reviews that include randomised or non-randomised studies of healthcare interventions, or both. *BMJ (Clinical Research Ed.)*, 358, j4008. <https://doi.org/10.1136/bmj.j4008>
- Stang, A. (2010). Critical evaluation of the Newcastle-Ottawa scale for the assessment of the quality of nonrandomized studies in meta-analyses. *European Journal of Epidemiology*, 25(9), 603-605. <https://doi.org/10.1007/s10654-010-9491-z>
- Struthers, C., Harwood, J., de Beyer, J. A., Dhiman, P., Logullo, P., & Schlüssel, M. (2021). GoodReports: Developing a website to help health researchers find and use reporting guidelines. *BMC Medical Research Methodology*, 21(1), 217. <https://doi.org/10.1186/s12874-021-01402-x>
- Sun, X., Wang, D., Wang, M., Li, H., & Liu, B. (2021). The reporting and methodological quality of systematic reviews and meta-analyses of nursing interventions for chronic obstructive pulmonary disease. A systematic review. *Nursing Open*, 8(3), 1489-1500. <https://doi.org/10.1002/nop2.767>
- Sun, X., Zhou, X., Zhang, Y., & Liu, H. (2019). Reporting and Methodological Quality of Systematic Reviews and Meta-Analyses of Nursing Inter-

- ventions in Patients With Alzheimer's Disease: General Implications of the Findings. *Journal of Nursing Scholarship: An Official Publication of Sigma Theta Tau International Honor Society of Nursing*, 51(3), 308-316. <https://doi.org/10.1111/jnu.12462>
- Thornton, A., & Lee, P. (2000). Publication bias in meta-analysis: Its causes and consequences. *Journal of Clinical Epidemiology*, 53(2), 207-216. [https://doi.org/10.1016/s0895-4356\(99\)00161-4](https://doi.org/10.1016/s0895-4356(99)00161-4)
- Urrútia, G., & Bonfill, X. (2010). The PRISMA Statement: A Proposal to Improve the Reporting of Systematic Reviews and Meta-Analyses [Declaración PRISMA: una propuesta para mejorar la publicación de revisiones sistemáticas y metaanálisis]. *Med Clin (Barc)*, 135(11), 507-511.
- Wasiak, J., Tyack, Z., Ware, R., Goodwin, N., & Faggion, C. M. (2017). Poor methodological quality and reporting standards of systematic reviews in burn care management. *International Wound Journal*, 14(5), 754-763. <https://doi.org/10.1111/iwj.12692>
- Yang, Q., Xian, H., Cheng, X., Wu, X., Meng, J., Chen, W., & Zeng, Z. (2024). Methodological and reporting quality assessment of systematic reviews and meta-analyses in the association between sleep duration and hypertension. *Systematic Reviews*, 13(1), 211. <https://doi.org/10.1186/s13643-024-02622-0>

## Appendices

**Supplementary table I.** Overview of included studies

Total	Overall sample (n, %)
	69 (100.0)
<b>The journal recommends the use of the PRISMA statement</b>	
No	36 (52.2)
Yes	33 (47.8)
<b>The authors state they reported according to the PRISMA statement</b>	
No	45 (65.2)
Yes	24 (34.8)
<b>N° of PRISMA recommendations (Mean, SD)</b>	
	19.7 (4.4)
<b>RoB *</b>	
High	48 (69.6)
Low	21 (30.4)
<b>Publication year (Mean, SD)</b>	
	2017.5 (2.9)
2009-2016	23 (32.9)
2017-2022	47 (67.1))
<b>Search strategy with geographical restrictions</b>	
Regional or national	56 (80.0)
International	14 (20.0)
<b>Countries</b>	
EEUU	18 (26.1)
Canadá	1 (1.4)
UK	6 (8.7)
Países bajos	6 (8.7)
Portugal	1 (1.4)
Noruega	1 (1.4)
Finlandia	1 (1.4)
España	3 (4.3)
Australia	1 (1.4)
Turquía	12 (17.4)
Irán	2 (2.9)
China	11 (15.9)
Taiwan	2 (2.9)

<b>Total</b>	<b>Overall sample</b>
	<b>(n, %)</b>
	<b>69 (100.0)</b>
India	3 (4.3)
Filipinas	1 (1.4)
<b>Continents</b>	
Australia & Asia	32 (46.4)
Europa	18 (26.1)
Norteamérica	19 (27.5)
<b>Educational area</b>	
Education	44 (62.3)
Psychology	8 (11.6)
Health	18 (26.1)
<b>Intervention type</b>	
Methodological strategies	30 (43.5)
Organizational strategies	17 (24.6)
Physical activities and others	10 (14.5)
Combination	12 (17.4)
<b>Population origin</b>	
Primary, secondary and functional diversity	10 (14.5)
Higher education	15 (21.7)
All levels included	19 (27.5)
Combination of levels	24 (34.8)

Source: Own elaboration

**Supplementary table II.** Adherence to PRISMA statement recommendations

First autor (publication year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	N° of rec- ommen- dations fulfilled
Aydin (2021)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	N	N	Y	N	N	N	N	Y	Y	Y	N	Y	N	Y	N	14
Akar (2020)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	N	Y	Y	N	Y	Y	Y	Y	N	14
Aktamis (2016)	Y	N	Y	Y	N	Y	N	Y	Y	N	Y	N	Y	Y	Y	Y	N	N	N	N	Y	Y	Y	Y	N	Y	N	16
Alegre- Ansuategui (2018)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	N	Y	Y	N	Y	Y	N	N	N	16
Alvarez- Bueno (2017)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	25
Aspiranti (2018)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	N	N	N	N	Y	Y	N	N	N	Y	N	Y	Y	Y	N	14
Ayaz (2015)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	N	N	N	Y	Y	Y	Y	Y	N	Y	N	17
Bai (2020)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	22
Balakrishnan (2021)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Bas (2016)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	N	N	N	Y	N	Y	Y	Y	N	Y	N	Y	N	17
Bedard (2019)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Brierly (2021)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Capar (2015)	Y	N	Y	Y	N	Y	N	Y	Y	N	Y	N	Y	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	10
Cartiff (2021)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	23

First autor (publication year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	N° of rec- ommen- dations fulfilled
Celilo (2011)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	N	N	N	N	N	N	N	Y	N	Y	Y	Y	Y	N	13
Cen (2021)	Y	N	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	22
Chandran (2022)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Chauhan (2016)	Y	N	Y	Y	N	Y	N	Y	Y	Y	Y	N	Y	Y	N	N	Y	Y	N	Y	Y	N	Y	Y	Y	Y	N	17
Chen (2019)	Y	N	Y	Y	N	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	N	N	N	N	Y	N	Y	Y	N	Y	N	15
Chen (2018)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Cheng (2018)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	20
Cheung (2013)	Y	N	Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	Y	N	N	N	Y	N	N	Y	Y	Y	Y	Y	Y	N	17
Costa (2021)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	N	N	N	Y	N	N	Y	Y	N	Y	Y	N	Y	Y	16
Dagva (2015)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	N	N	N	N	Y	Y	N	Y	N	Y	N	15
De Boer (2014)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	N	18
De Greeff (2017)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	24
Dietrichson (2017)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	23
Donker (2013)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	N	19
Double (2019)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	21
Ergen (2017)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	N	21

First autor (publication year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	N° of rec- ommen- dations fulfilled
Faramatzi (2015)	Y	N	Y	Y	N	Y	N	N	Y	Y	N	N	Y	N	N	N	N	Y	N	N	N	N	Y	Y	Y	Y	Y	12
García- Hermoso (2021)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	26
Hew (2018)	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Hu (2020)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	23
Hu (2018)	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	26
Jacobse (2011)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	N	N	N	Y	Y	N	Y	Y	Y	Y	N	18
Kaçar (2021)	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	Y	Y	Y	N	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	18
Kalalan (2017)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	N	N	Y	Y	N	Y	Y	Y	Y	Y	19
Karagöi (2019)	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	Y	N	N	N	Y	N	N	Y	Y	Y	N	Y	N	Y	N	14
Karich (2014)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	21
Kim (2021)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	N	Y	Y	Y	Y	N	21
King-Sears (2021)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	22
Korpershoek (2016)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	Y	N	N	N	N	N	N	Y	Y	Y	Y	Y	17
Lag (2019)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	23
Leung (2015)	N	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	N	Y	Y	Y	Y	N	20

First autor (publication year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	N° of rec- ommen- dations fulfilled
Linden (2018)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	22
May (2021)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Moore (2018)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	25
Norris (2019)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	N	N	Y	Y	Y	Y	Y	Y	Y	Y	23
Oh-Young (2015)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	N	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	20
Orhan (2019)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	N	Y	Y	Y	Y	Y	N	Y	N	18
Petersen- Brown (2019)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	24
Phelps (2019)	Y	N	Y	Y	N	Y	Y	N	Y	N	Y	Y	Y	Y	N	Y	N	N	N	Y	Y	Y	Y	Y	N	N	Y	17
Robbins (2009)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	N	N	N	N	N	Y	Y	Y	Y	Y	Y	N	18
Saw (2021)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	22
Savvyah (2017)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	24
Semerici (2015)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	N	N	N	N	N	N	N	Y	N	Y	N	12
Sneck (2019)	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	23
Steele (2016)	Y	N	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	22

First autor (publication year)	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26	27	N° of rec- ommen- dations fulfilled
Sugano (2020)	Y	N	Y	Y	N	Y	N	N	Y	Y	N	N	N	N	N	N	N	N	N	N	Y	N	N	Y	N	Y	N	9
Sung (2016)	Y	N	Y	Y	N	Y	Y	N	Y	N	Y	N	Y	N	Y	N	N	N	N	N	Y	Y	N	Y	N	N	Y	13
Tan (2017)	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	26
Tokac (2019)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	N	Y	Y	21
Warren (2012)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	Y	N	Y	N	N	N	N	N	N	Y	Y	Y	Y	N	15
Watson (2017)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	24
Wilson (2019)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	24
Sang (2020)	Y	N	Y	Y	Y	Y	Y	Y	Y	Y	Y	N	Y	N	N	N	Y	N	Y	Y	Y	Y	Y	Y	N	Y	Y	20
Yoiro (2012)	Y	N	Y	Y	N	Y	N	N	Y	Y	Y	N	Y	Y	N	Y	Y	Y	N	Y	Y	Y	Y	Y	Y	Y	N	17
Zheng (2016)	Y	N	Y	Y	N	Y	Y	Y	Y	N	Y	N	Y	Y	Y	N	N	N	Y	Y	Y	Y	N	Y	N	Y	N	17

Y: Yes, N: No

Source: Own elaboration





# **Efficacy of Psycho-socio-educational Programs applied in Juvenile Justice in Spain: A Systematic Review**

## **Eficacia de los Programas Psico-socio-educativos aplicados en Justicia Juvenil en España: Revisión Sistemática**

<https://doi.org/10.4438/1988-592X-RE-2025-412-737>

**Mery Estefanía Buestán-Játiva**

<https://orcid.org/0000-0002-9023-1966>

*Universidad de Almería*

**Elena Ortega-Campos**

<https://orcid.org/0000-0003-1994-7109>

*Universidad de Almería*

**Leticia de la Fuente-Sánchez**

<https://orcid.org/0000-0002-8320-3078>

*Universidad de Almería*

**Juan García-García**

<https://orcid.org/0000-0002-0123-8497>

*Universidad de Almería*

### **Abstract**

The study of intervention in juvenile antisocial behavior acquires relevance for the implementation of socio-educational measures within the juvenile justice system. However, there is little information available on the factors related to the effectiveness of different practices and intervention programs in juvenile justice. The aim of this systematic review is to determine the effectiveness of psycho-socio-educational programs applied to juveniles in conflict with law in Spain. The search for studies was carried out in electronic databases and specialized journals and online resources. The inclusion criteria used for the selection of papers

in this systematic review were: studies in which a psycho-socio-educational intervention program was implemented in the Spanish juvenile justice system; the sample had to be made up of juveniles who were completing a socio-educational measure under Organic Law 5/2000 during the psycho-socio-educational program; and the effectiveness of the program had to be evaluated and quantitative data on its effectiveness had to be presented. This work includes a total of 18 independent studies with a total sample of 6753 juveniles, which provide results in areas such as satisfaction, reduction of aggressive behaviour, greater assumption of responsibility, or improvements in the family environment, among other aspects. Of the total number of studies, eleven studies present quantitative data indicating an overall reduction in the repetition of criminal behavior, among 3.4% and 28.8%. This underlines the importance of evaluating psycho-socio-educational programs aimed at juveniles in conflict with the law and of gathering information on effectiveness indicators for the development of appropriate programs in the juvenile justice field.

**Keywords:** Effectiveness, Juvenile Justice, Psycho-socio-educational Programs, Repeated Offenses, Young Offenders.

### **Resumen**

El estudio sobre la intervención en la conducta antisocial penada en la adolescencia adquiere relevancia para la implantación de medidas socioeducativas dentro del sistema de justicia juvenil. Sin embargo, es escasa la información disponible sobre los factores relacionados con la eficacia de las diferentes prácticas y programas de intervención en justicia juvenil. El objetivo de esta revisión sistemática es conocer la eficacia de los programas psico-socio-educativos aplicados a menores en conflicto con la ley en España. La búsqueda de estudios se ha realizado en bases de datos electrónicas y en revistas y recursos online especializados. Los criterios de inclusión utilizados para la selección de los trabajos han sido: estudios en los que se implante un programa de intervención psico-socio-educativo en el sistema de justicia juvenil español; la muestra debía estar formada por menores que estuvieran cumpliendo una medida socioeducativa bajo la Ley Orgánica 5/2000 durante la realización del programa psico-socio-educativo; y se debía evaluar la eficacia del programa y presentar datos cuantitativos sobre la eficacia del mismo. En este trabajo se incluyen 18 estudios independientes con una muestra total de 6.753 jóvenes, que aportan resultados de áreas como satisfacción, disminución de conductas agresivas, mayor asunción de responsabilidad o mejoras en el ámbito familiar, entre otros aspectos. Del total de estudios, once presentan datos cuantitativos que indican una reducción general de la reiteración de la conducta delictiva, entre un 3,4% y un 28,8%. Esto subraya la importancia de evaluar los programas psico-socio-educativos dirigidos a menores en conflicto con la ley y de recabar información sobre los indicadores de eficacia para la elaboración de programas adecuados en el ámbito de justicia juvenil.

**Palabras clave:** Eficacia, Jóvenes Infractores, Justicia Juvenil, Programas Psico-socio-educativos, Reiteración Delictiva.

## Introduction

Punishable antisocial behavior encompasses all actions that violate social norms and rights, constituting a breach of the law (De la Peña, 2005; García et al., 2010; Rutter et al., 2000; Villafuerte-Díaz et al., 2022). Research on punishable antisocial behaviour in adolescents is relevant for the planning and implementation of public policies related to the intervention and prevention of violence, the psychoeducational perspective being the most effective (Badasa et al., 2019; García et al., 2012; Organización Mundial de la Salud, 2020; Souverein et al., 2019).

In Spain, during 2024, a total of 13491 juveniles between the ages of 14 and 17 were convicted by final sentence, representing an increase of 3.6% compared to the previous year. 79.7% of the convicted juveniles were males. Offenses with highest incidence were injuries (31.6%), thefts (15.8%), and threats (9.8%) (National Institute of Statistics, 2024). Organic Law 5/2000 regulates the legal responsibility of minors, between 14 and 17 years, who have committed acts classified as offenses of Penal Code. This law emphasizes that juveniles who have committed punishable antisocial behavior must receive a comprehensive education and are required to participate in formative, educational, and occupational activities. The educational measures imposed by juvenile court are oriented to the effective social reintegration of the juvenile from a psychoeducational perspective (Barroso-Hurtado y Serrano, 2019; García et al., 2012).

The problem of violence in juveniles is not an isolated phenomenon; by contrast, it is linked to other problematic behaviors and psychosocial risk factors (Moffitt, 2018; Young et al., 2017). Some of the risk factors related to punishable antisocial behavior in juveniles include disadvantaged social situations, school failure, truancy, dissocial peers, substance use, dysfunctional families, or prior antisocial behaviors (Aazami et al., 2023; Anjaswarni et al., 2019; Jolliffe et al., 2017; Lahey et al., 2003; Mampaso et al., 2014; Salazar et al., 2011; Van der Put et al., 2011). The models used to explain and address punishable antisocial behavior in youth are based on identifying the risk and protective factors presented by each juvenile offender (Andrews & Bonta, 2010; Ward & Brown, 2004).

The *Risk-Need-Responsivity* model (Andrews & Bonta, 2010) emphasizes the assessment of the risks and needs presented by the juvenile with the purpose of planning an intervention adapted to their specific needs. Complementarily, the Good Lives Model (Ward & Brown, 2004) underlines the need to develop the capacities and strengths of juveniles who engage in punishable behaviors. The current trend is focused on psychoeducational interventions aimed at working with the needs and

strengths of each youth, identified in evaluations before interventions carried out in juvenile justice (Calvo, 2007; Lipsey, 2014; Puig y Capdevila, 2016).

The *What Works?* movement emerges in the adult justice system in response to the Martinson (1974) report, which presented a review of 231 rehabilitation programs in prisons, concluding that “nothing works” in reducing reoffending. The research on What works? became popular in the 1980s and 1990s within the juvenile justice system (Krisberg, 2005), proposing the analysis of interventions carried out in adult justice with the aim of evaluating the effectiveness of the programs and identifying the key or core elements for the intervention to be successful (Lab & Whitehead, 1988). This idea has been transferred to the field of juvenile justice with the premise of evaluating the effectiveness of interventions carried out, as well as its contribution to the psychoeducational reintegration of juveniles (Dowden & Andrews, 1999; Evans-Chase & Zhou, 2014; Pappas & Dent, 2023; Zahn et al., 2009). This approach arises as a response to the need to support the decisions of juvenile justice system on empirical evidence (Krisberg, 2005; Latessa, 2004). The aim of What Works is to identify and develop interventions, programs and policies based on scientific evidence that have been shown to be effective in reducing reoffending of punishable antisocial behavior (Krisberg, 2005).

The development of What Works is related to the increase in evaluative research and the search for answers regarding the effectiveness and efficiency of practices and policies in the justice system (Elliott et al., 2020; Latessa, 2004; Smith, 2005). This approach has shown the importance of identifying risk and protective factors associated with reoffending of punishable antisocial behavior, which are fundamental to planning and developing specific programs for juvenile offenders (Garrido et al., 2006a; Lipsey, 2009; Lipsey et al., 2007; Ortega et al., 2014; Pappas & Dent, 2023; Redondo et al., 2011).

Following the premises of the *What Works?* model, some studies have been conducted, notably the meta-analysis by Lipsey (2009) where the effectiveness of 548 programs applied to juvenile offenders was analyzed. The results showed differences in the reoffending of punishable antisocial behavior according to the type of intervention, with lower reoffending in the guidance and counseling programs ( $\phi=0.066$ ), multidisciplinary programs ( $\phi=0.062$ ), prosocial skills development and behavior management programs ( $\phi=0.060$ ) and restorative justice programs ( $\phi=0.050$ ). Subsequently, Pappas & Dent (2023) evaluated the effectiveness of psycho-socio-educational programs applied in juvenile justice, finding a statistically significant reduction in reoffending in

punishable antisocial behavior ( $\phi=.09$ ) among juveniles who participated in a program, compared to juveniles who did not participate in any psycho-socio-educational program.

Redondo and Sánchez-Meca (2003) conducted a meta-analysis of 17 European programs applied to 2775 juvenile offenders and found an overall effect size of  $r=.18$ , which represented an eighteen-point reduction in reoffending in punishable antisocial behavior in the intervention groups. The most relevant findings were obtained in educational programs ( $r=.42$ ), diversion programs under the judicial system ( $r=.29$ ), therapeutic community programs ( $r=.14$ ), and cognitive-behavioral programs ( $r=.14$ ). A couple of years later, Garrido et al. (2006a) analyzed the effectiveness of 30 programs applied to juvenile offenders and obtained an overall effect size  $d=.14$ , which meant a seven-point reduction in punishable antisocial behavior. Cognitive-behavioral ( $d=.22$ ) and cognitive ( $d=.12$ ) interventions showed greater effectiveness.

Traditionally, intervention has focused on risk factors for juvenile offenders, however, current research emphasizes the need for interventions based on individual strengths of each youth, with the goal of achieving their social reintegration (Day & Howells, 2002; Looman & Abracen, 2013; Menon & Cheung, 2018; Zelaya, 2020). The inclusion of a psycho-educational perspective, crucial in situations of risk and social exclusion, gains greater relevance in the context of juvenile justice, guiding the re-education of these juveniles according to their personal, family, and social circumstances (Caride, 2005; Dionne y Altamirano, 2012).

One of the factors traditionally studied as an indicator of the effectiveness of programs applied in juvenile offenders has been reoffending in punishable antisocial behavior (García et al., 2010, 2012; Ortega et al., 2014). In this regard, the assessment of the risk of reoffending is a fundamental axis during and after the intervention process (Carbonell et al., 2016; Redondo y Martínez, 2013; Schwalbe, 2007). Specific instruments have been developed to measure the risk of reoffending in juvenile offenders, highlighting among the most used the Youth Level of Service Case Management Inventory (Hoge y Andrews, 2006), translated into Spanish as *Inventario de Gestión e Intervención con jóvenes (IGI-J)* (Garrido et al., 2006b), and the Structured Assessment of Violence Risk in Youth (SAVRY) (Borum et al., 2003), adapted into Spanish by Vallés y Hilterman (2006).

Despite the development of psycho-socio-educational programs in the Spanish juvenile justice system, knowledge about the effectiveness of these programs is limited, making it necessary to investigate in which indicators can serve as a reference for the effectiveness of applied pro-

grams, with the aim of providing information on the key factors for the success of interventions (Barroso-Hurtado y Serrano, 2019; Redondo et al., 2011, 2012a). The scarcity of studies on the effectiveness of psycho-socio-educational programs takes on particular importance for intervention planning to prevent and/or reduce punishable antisocial behavior in juveniles.

Therefore, the objective of this study is to conduct a systematic review on the effectiveness of psycho-socio-educational programs applied to juvenile offenders in Spain, focussing on the evaluation and indicators of effectiveness as reoffending of punishable antisocial behavior.

## Method

For the development of this systematic review study, the PRISMA guidelines were followed (Page et al., 2021).

### Search strategy

The information sources used in the search for studies were: a) electronic databases (*Web of Science*, *Scopus*, *Science Direct*, *Redalyc*, *PubMed*, *Psycinfo* y *Dialnet*); b) specialized journals (*Bienestar y Protección Infantil* [Child Welfare and Protection], *Boletín Criminológico* [Criminology Bulletin], *Revista Española de Investigación Criminológica* [Spanish Journal of Criminological Research] and *Anuario de Psicología Jurídica* [Yearbook of Legal Psychology]); c) specialized online resources (*Observatorio de la Infancia y Adolescencia* [Child and Adolescent Observatory], *Observatorio de la Delincuencia* [Delinquency Observatory], *Observatorio de Justicia Penal Juvenil* [Juvenile Justice Observatory] e *International Juvenile Justice Observatory*); d) reports issued by relevant authorities and competent institutions in juvenile justice within the Spanish autonomous communities (*Agencia para la Reeducación y Reinserción del Menor Infractor* [Agency for the Re-education and Reintegration of Juvenile Offenders]; *Centro de Estudios Jurídicos y Formación Especializada* [CEJFE; Centre for Legal Studies and Specialized Training]; *Diagrama Foundation*, *GINSO* and *ADIS-Meridianos*); and e) review of references of the studies included in this work, as well as other systematic reviews and meta-analyses conducted. The literature search was conducted from October to December 2023.

### Palabras clave

The search for studies was conducted using specifically selected descriptors. The keywords used were: *Young offenders*, *Effectiveness*, *Evalua-*

*tion, Intervention, Juvenile Justice y Programs*. In Spanish were: Jóvenes Infractores, Eficacia, Evaluación, Intervención, Justicia Juvenil and Programas. The Boolean operators used were AND, OR.

### **Eligibility criteria**

The inclusion criteria used for the selection of studies in this systematic review were: (1) Studies in which a psycho-socio-educational program was applied in the Spanish juvenile justice system; (2) Studies in which the sample consisted of juvenile offenders subject to a socio-educational measure under the Organic law 5/2000 during the implementation of psycho-socio-educational program; (3) eligible studies must evaluate the effectiveness of program; and (4) report quantitative results related to this effectiveness.

(2) Studies in which juvenile offenders were subject to judicial measures under legislation prior to Organic law 5/2000; y (3) mixed-sample studies, that is, those involving adolescents both with and without judicial measures, as well as those including adolescents and adults.

### **Data extraction and coding of studies**

The variables extracted from the selected studies were: authors and year of publication, type of publication (report, journal article, doctoral thesis, book or manual), study location (Spanish autonomous community, CCAA), study design (experimental, quasi-experimental, prospective, retrospective), sample size, age, gender (% males), nationality (% Spanish), regime of educational measure by juvenile court (open, closed, or mixed), type of judicial measure, type of offense, type of intervention program, assessment instruments, prior records (%), and reoffending in punishable antisocial behavior (%).

### **Search procedure**

The search yielded a total of 15 studies with 18 independent studies on programs applied to juveniles under socio-educational measures within the juvenile justice system. The search began in the electronic databases, yielding 68 results. After removing duplicates and non-empirical studies, 40 studies were identified for screening. Of these, 25 studies were examined in detail and 15 were excluded for different reasons: application of the program under legislation prior to Organic Law 5/2000 (n=4), not reporting data on program results (n=5), or not reporting data on program evaluation (n=6). Consequently, ten independent studies were included.

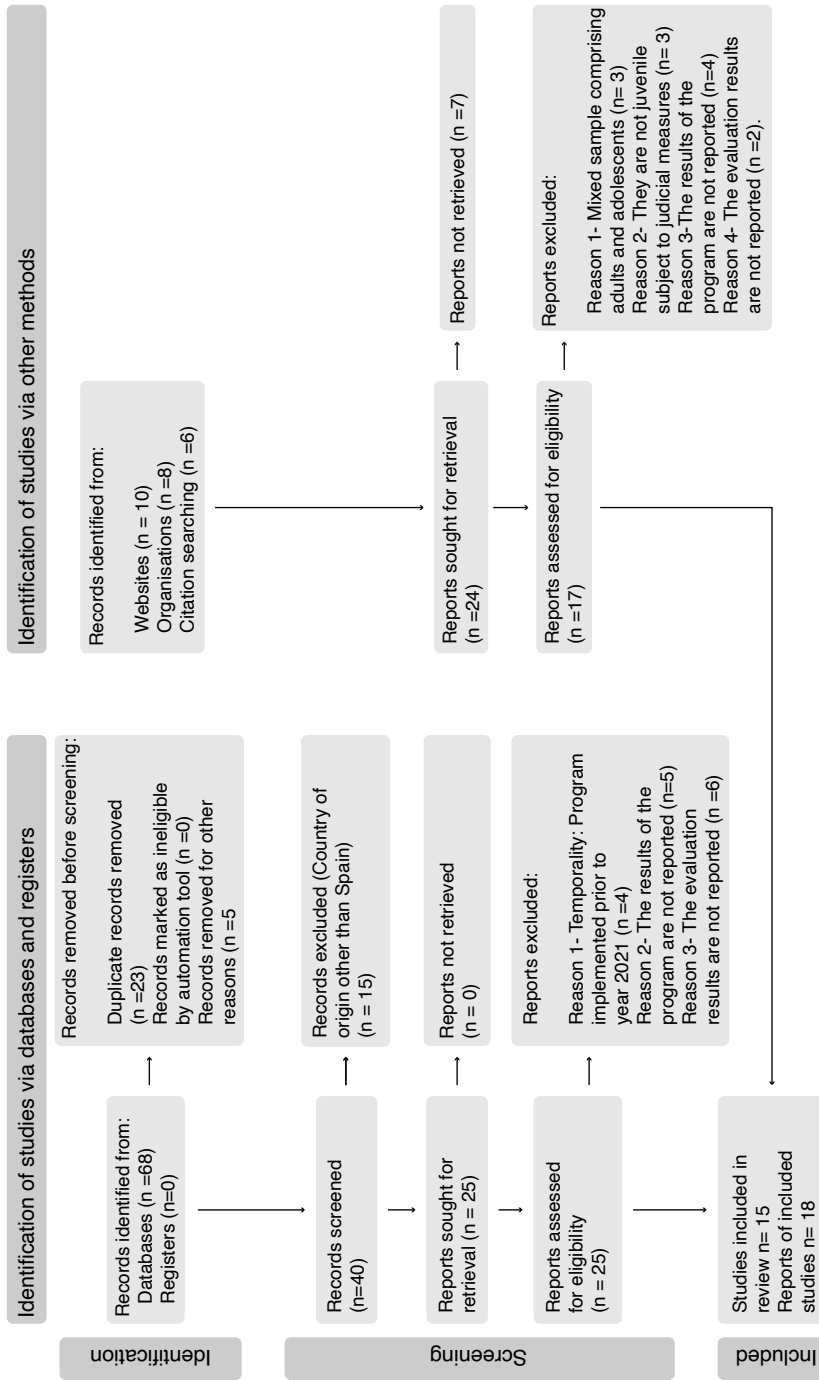
On the other hand, 24 studies were identified through specialized journals and regional authorities such as the Centro de Estudios Jurídicos y

Formación Especializada [CEJFE; Centre for Legal Studies and Specialised Training] or the Agencia para la Reeducación y Reinserción del Menor Infractor [Agency for the Re-education and Reintegration of Young Offenders], and the review of the references of the studies included in this paper. A total of 17 studies were assessed for eligibility, of which 12 were excluded for the following reasons: mixed samples including adolescents and adults ( $n=3$ ), adolescents not subjected to judicial measures ( $n=3$ ), did not report data on program results ( $n=4$ ), did not report data on program evaluation ( $n=2$ ). Five independent studies were included. The study selection process is illustrated in figure I.

### **Data analysis**

A descriptive analysis (qualitative and quantitative) of the most relevant results of the variables included in the systematic review has been conducted. For the quantitative part, effect sizes were calculated for each study or independent sample to assess the effectiveness of psycho-socio-educational programs. A forest plot was generated to display the estimated effect size for each independent sample and the pooled average effect size, with the aim of providing an exploratory analysis of the effectiveness of intervention programs in juvenile justice. The effect sizes were estimated from the proportions of reoffending in punishable anti-social behavior, using a fixed-effects model, deemed most suitable for the sample size and the descriptive meta-analytical approach employed to the subject of study. The Jamovi 2.5 software was used for data analysis.

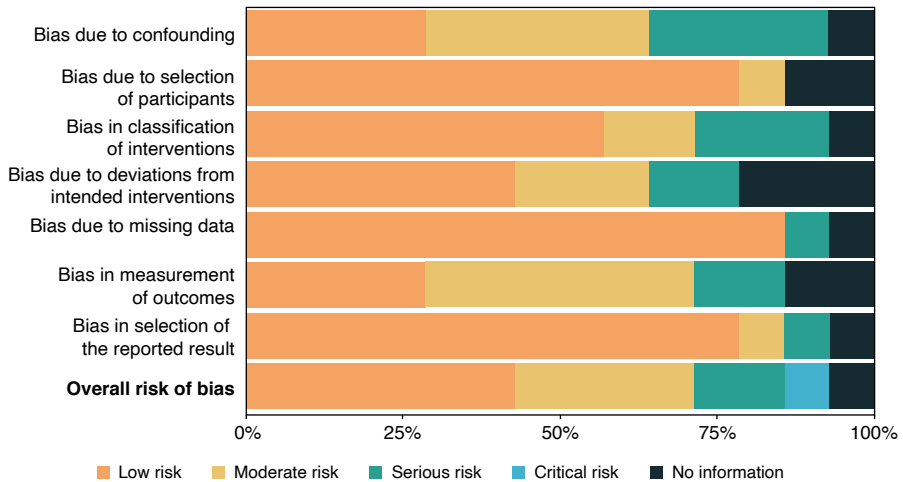
**Figure 1.** Flow chart



## Risk of Bias Assessment

The risk of bias in systematic review refers to the potential presence of systematic errors in included primary studies, which can compromise the validity of the finding and lead to inaccurate or misleading conclusions (Drucker et al., 2016). In this study, the risk of bias was assessed using the ROBINS-I tool (Risk of Bias in Non-randomized Studies-of Interventions) (Sterne et al., 2016), which evaluates the potential for bias in the results of non-randomized intervention studies that compare the effects of two or more interventions. This tool was selected to establish a criterion for assessing the methodological rigour of the studies included in this systematic review. The results indicate a risk of bias that ranges from moderate to low (Figure II). The Robvis software was employed to generate the corresponding risk of bias graph (McGuinness & Higgins, 2021).

Figure II. Summary plot



Note. Bar chart of the distribution of risk of bias in each assessed domain.

## Results

The search process yielded 15 studies comprising 18 independent studies on socio-educational programs implemented within the Spanish juvenile justice system, with a sample size of 6753 adolescents. The main characteristics of each study, together with the corresponding results, are represented in Tables I and II, as well as Figure III.

### Global characteristics of studies

First, Table I provides an overview of the global and integrated characteristics of the 18 studies included in the systematic review, with respect

to the following variables: type of publication, year of publication, study location, sample size, type of program, gender (% males), nationality (% spanish), regime of educational measure imposed, prior records and measure of reoffending in punishable antisocial behavior.

With respect to type of publication, nine of the eighteen studies are published as reports (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Camps y Cano, 2006a,b; Capdevila et al., 2017a,b,c; Junta de Extremadura, 2021), four as doctoral theses (Fernández, 2016; Luna et al., 2015; Pintado, 2012; Sánchez, 2008), four as journal articles (Burcet i Solé et al., 2019; Ocáriz, 2013; Palanques-Alegre et al., 2022; Redondo et al., 2012b) and one as a treatment manual (González et al., 2013). The studies were published between 2006 and 2023, predominantly between 2011 and 2020; specifically, 28% of the studies were published between 2011-2015 and 33% between 2016-2020. Cataluña is the Autonomous Community with the most studies carried out (55%), followed by Madrid and Valencia, each with 11%, whereas in the remaining Autonomous Communities a single study has been carried out. Regarding the sample size of the studies, it ranges between 9 and 2022 juveniles (Capdevila et al., 2017; Pintado, 2012), the majority involved relatively small samples. Specifically, 44% of studies included fewer than 50 participants, 22% involved between 51-100 participants and 17% between 101-1000 y >1000, each of them.

With respect to characteristics of the adolescents, males were the predominant gender (81.23%), 28% of the studies reported between 71% and 80% males, 22% reported between 81% and 90%, and 17% included more than 91% males. Regarding the nationality, 73% were spanish; specifically, 28% of the studies reported at least 75% spanish, 22% reported between 76% and 90% and 11% reported more than 91%. The mean age of the juveniles was 16.8 years.

Regarding the regime of educational measure by juvenile court, el 33% of the studies applied psycho-socio-educational program while the juveniles were in close internment (Camps y Cano, 2006b; Fernández, 2016; González et al., 2013; Junta de Extremadura, 2021; Palanques-Alegre et al., 2022; Pintado, 2012), 61% of the studies the juveniles were in an open setting or without deprivation of liberty (Alba et al., 2007; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Camps y Cano, 2006a; Capdevila et al., 2017a,b,c; Luna et al., 2015; Ocáriz, 2013; Redondo et al., 2012b; Sánchez, 2008), and one study included juveniles from both systems (Barnes et al., 2018).

According to the area of intervention, 28% of studies implemented and evaluate Mediation and Reparation programs (MRM) (Blanch y Mancho, 2023; Capdevila et al., 2017a,b,c; Ocáriz, 2013), 22% implemented specific

programs of child-to-parent violence and supporting families (Barnes et al., 2018; Burcet i Solé et al., 2019; González et al., 2013; Sánchez, 2008), another 22% implemented programs targeting social skills and prosocial thinking (Alba et al., 2007; Luna et al., 2015; Pintado, 2012; Redondo et al., 2012b), 17% focused on self-control and impulsivity programs (Camps y Cano, 2006a,b; Palanques-Alegre et al., 2022), and 11% addressed substance use and addictions programs (Fernández, 2016; Junta de Extremadura, 2021).

With respect to criminological variables, 61% of the studies indicate the prior records of the juveniles (Blanch y Mancho, 2023; Camps y Cano, 2006a,b; Capdevila et al., 2017a, b,c; Fernández, 2016; González et al., 2013; Ocáriz, 2013; Palanques-Alegre et al., 2022; Redondo et al., 2012b). Similarly, 61% of studies reported of the juvenile recurrence in punishable antisocial behavior (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Camps y Cano, 2006a; Capdevila et al., 2017a,b,c; Fernández, 2016; González et al., 2013; Ocáriz, 2013).

Furthermore, with respect to study design, only two of the eighteen studies adopted a quasi-experimental design with control groups (Palanques-Alegre et al., 2022; Redondo et al., 2012b), whereas the remaining studies employed retrospective designs. The evaluation instrument employed have been diverse, ranging from specific tools for assessing the risk of repetition of punishable antisocial behavior such as the IGI-J (Garrido et al., 2006b) or the SAVRY (Vallés y Hilterman, 2006), along with instruments that assess additional dimensions related to the components addressed in the programs such as the Gismero Social Skills Scale of (Gismero, 2000), the Teen-Addiction Severity Index (T-ASI) (Kaminer et al., 1991), the Criminal and Social History: Juvenile Version (HCS-J) (Graña y González, 2008), or general instruments like the MACI (Millon, 1993), or the Emotional Quotient Inventory: Youth version (EQi-YV) (Baron & Parker, 2000).

**Table I.** Characteristics of the studies included in the systematic review.

<b>Variable</b>	<b>N(%)</b>	<b>Variable</b>	<b>N(%)</b>
<b>Type of publication</b>		<b>Year of publication</b>	
Report	9(50)	2006-2010	4(22)
Journal article	4(22)	2011-2015	5(28)
Doctoral thesis	4(22)	2016-2020	6(33)
Book or manual	1(6)	2021-2023	3(17)
<b>Gender (% males)</b>		<b>Nationality (% spanish)</b>	
<70%	4(22)	<60%	2(11)
71-80%	5(28)	61 - 75%	3(17)

Variable	N(%)	Variable	N(%)
81-90%	4(22)	76 – 90%	4(22)
>91%	3(17)	>91%	2(11)
Not reported	2(11)	Not reported	7(39)
Study location	Type of program		
Cataluña	10(55)	Mediation and Reparation	5(28)
Madrid	2(11)	Child-to-parent violence and family context	4(22)
Valencia	2(11)	Social Skills and prosocial thinking	4(22)
País Vasco	1(6)	Self-control and impulsivity	3(17)
Extremadura	1(6)	Drug use and addictions	2(11)
Asturias	1(6)	Sample size	
Others CCAA	1(6)	<50	8(44)
Regime of measure by juvenile court		51 – 100	4(22)
Abierto	11(61)	101 – 1000	3(17)
Cerrado	6(33)	<1001	3(17)
Mixto	1(6)		
Prior records	Reoffending reported		
YES	11(61)	YES	11(61)
NO	7(39)	NO	7(39)

Note: The data shown in the table pertain to the 18 studies included in the systematic review.

## Effectiveness of programs

The effectiveness of programs has been reported in different ways in the reviewed studies. Table II shows the disaggregated data in each study. Thus, among the eleven studies reporting quantitative data on reoffending in punishable antisocial behavior in juveniles as an indicator of program effectiveness, reoffending rates ranged between 3.40% and 28.80% in the intervention group (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Camps y Cano, 2006a; Capdevila et al., 2017a,b,c; Fernández, 2016; González et al., 2013; Ocariz, 2013). A reduction in risk level of 37.79% was observed, with a pre-test mean of 23.10 and a post-test mean of 14.37 (Palanques-Alegre et al., 2022). In the study by Barnes et al. (2018), 19.5% reoffending in punishable antisocial behavior was reported, of which 7.3% were associated with situations of family abuse.

Additionally, four studies have reported program attendance as a measure of effectiveness, showing attendance data ranging between 77% and 90.61%, compulsory attendance was not required in any of the studies (Alba et al., 2007; Barnes et al., 2018; Burcet i Solé et al., 2019; Camps y Cano, 2006), while other authors report on the level of satisfaction with

the program, such as Camps y Cano (2006a,b) who recorded scores of 2.6 and 2.8, on a scale of 1 to 3, or the studies by Luna et al. (2015) and the Junta de Extremadura (2021), which report that 95% of participants are satisfied with the program implemented.

Regarding the specific programs on child-to-parent violence, the study by Sánchez (2008) reports that, after intervention, 53% of juveniles acknowledge responsibility for the acts committed, 34.8% treat their parents with respect, whereas 25.8% have resumed verbal aggression, 18.2% display habitual verbal aggression, 13.6% engage in mild aggression, while 3% have damaged property at home or threatened with a weapon and 4.5% have committed serious aggression. The study by Barnes et al. (2018) reported that the relationship with parents improved in 52.4% of cases, whereas the parents' self-perception of change reached 66.70%.

Intervention programs targeting the improvement of self-control and impulsivity in juveniles have reported 70.5% indicator an increase in training on managing violent behavior (Camps y Cano, 2006a, b). The study by Palanques-Alegre et al. (2022) found that the propensity for impulsivity was reduced by 23.81%, showing a mean pre-test score of 80.02 and post-test score of 60.96, increase the conflict resolution score, from 52.18 to 60.89, and likewise coping positive showed a similar pattern, from 51.75% to 55.53%.

In interventions to improve prosocial thinking and social skills, Redondo et al. (2012b) reported a significant effect on social skills, aggression, and self-esteem. Alba et al. (2007) recorded lower aggression in 40% of cases, and 60% of the juveniles managed to verbalize their feelings. Pintado (2012) reported that 55.5% of juveniles who finished of program exhibited positive changes in the personal domain.

Psychoeducational programs targeting addiction treatment, the PIMICA program (Junta de Extremadura, 2021) recorded a favourable progression of discharges. Fernández (2016) reported a high effect size ( $\eta^2 = .55$ ) in the quasi-experimental group and the reoffending in punishable antisocial behavior was lower (28.80%) in contrast to the group that did not participate in the program (37.50%). The IGI-J a 10% reduction of total risk was recorded.

With respect to effectiveness of Mediation and Reparation programs, the results indicated that 82.6% of the mediations finalized in 2012 were positive, showing a reoffending rate of 5.9%, compared to the 18% rate of cases that finalized mediation unsuccessfully (Ocáriz, 2013). Capdevila et al. (2017a,b,c) reported that reoffending in punishable antisocial behavior in 2005, 2008 and 2010, recorded values of 14.40%, 26.10% and 27.50, respectively. Lastly, Blanch y Mancho (2023) reported a 26.70% of reoffending in punishable antisocial behavior.

**Table II.** Characteristics of studies and reoffending in antisocial behavior

Authors and Year	Sample	Judicial measure	Type of offense	Program	Assessment	Prior records	Reoffending
Camps y Cano (2006a)	N = 19 Age = 18.6 Nationality = 63.2	Socio-educational tasks and probation	Assaults 15.79%; sexual offense 47.37%; homicide 26.32%; domestic violence 5.26%	Control of violent behavior	Clinical impressions form, T-ASI, IGI-J, HCS-J.	68,4%	10%
Camps y Cano (2006b)	N = 10 Age = 19 Nationality = 80	Custodial measure	Assaults 80%	Self-control and assertiveness	Clinical impressions form, T-ASI, IGI-J, HCS-J.	90%	
Alba et al. (2007)	N = 10 (5 GC y 5 TG) Age = 16.7 Gender = 100	Socio-educational tasks and probation		PPS-VCJ	External indicator questionnaires and assessment, IGI-J.		20% GT y 80% GC
Sánchez (2008)	N = 85 (19 CG y 66 TG) Age = 15.5 Gender = 67.1	Socio-educational tasks	Threat or insult 100%; property damage 83.5%; weapon threat 20%; minor assault 75.3%; serious assault 24.7%	Familiar intervention	Interview, Dynamic Questionnaire, intelligence test, Personality Questionnaire.		
Pintado (2012)	N = 9 Age = 14-20 Gender = 88.8	Custodial measure		PPS-VCJ	Learning styles Questionnaire, final assessment Questionnaire, external indicators questionnaire, IGI-J.		
Redondo et al. (2012b)	N = 28 (11 CG y 17 TG) Age = 17.7 TG y 17.4 CG Gender = 63.6 TG y 70.6 CG	Socio-educational tasks and probation	Violent offense TG 76.5%; violent offense CG 33.3%	PPS	Interview, Rosenberg self-esteem scale, IRI, EHS, ATPV, AGQ, BIS-10.	41.2% (TG) 85.7% (CG)	

Authors and Year	Sample	Judicial measure	Type of offense	Program	Assessment	Prior records	Reoffending
González et al. (2013)	N = 263 Age = 17 Gender = 65 Nationality = 82	Custodial measure		Educational and therapeutic treatment program for upward family abuse	Progress assessment forms	56%	3,40%
Ocáriz (2013)	N = 408 Age = 17 Gender = 77.5		Theft 9.55%; threats 10.78%; Larceny 17.15%; property damage 22%; assault 28.2%; offenses against property 4.4%; offenses against persons 7.8%	MRM	Reports	8,3%	24,20%
Luna et al. (2015)	N = 33 Age = 16-18 Gender = 79 Nationality = 57	Socio-educational tasks and probation		Social Skills program	Reports		
Fernández (2016)	N = 92 (40 CG y 52 TG) Age = 17 Gender = 100 Nationality = 39.1		Violent robbery 32.7%; sexual offense 17.44%; violent offense 5.45%; public health offenses 1.1%; domestic violence 1.1%	Educational and therapeutic treatment program for drug use	T-ASI, IGI-J	33,7%	28.8% TG 38% CG
Capdevila et al. (2017a)	N = 934 Age = 15.7 Gender = 87.2 Nationality = 91.9		Offense against persons 20.1%; violent offense against property 7.8%; non-violent offense against property 40.5%	MRM	SAVRY	22,8%	14,40%
Capdevila et al. (2017b)	N = 2022 Age = 16 Gender = 80.5 Nationality = 76.4		Offense against persons 36.7%; violent offense against property 8.3%; non-violent offense against property 26.9%	MRM	SAVRY	26,7%	26,10%

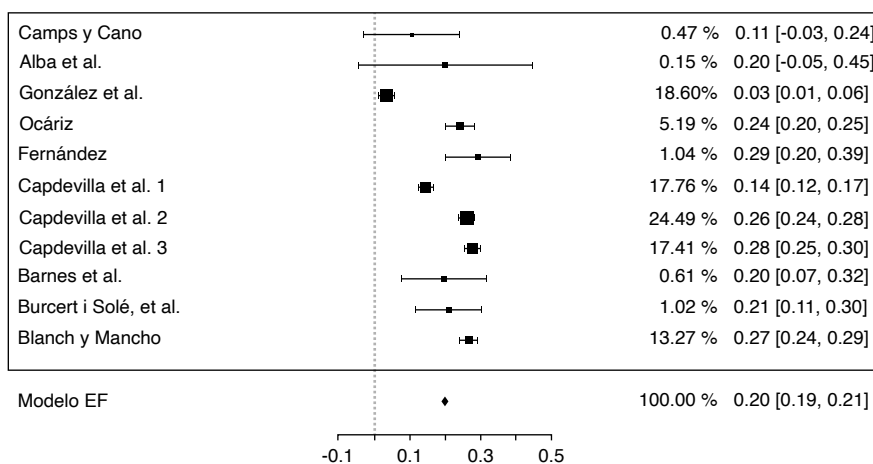
Authors and Year	Sample	Judicial measure	Type of offense	Program	Assessment	Prior records	Reoffending
Capdevila et al. (2017c)	N = 1486 Age = 15.2 Gender = 79 Nationality = 71.1		Offense against persons 38.2%; violent offense against property 8.1%; non-violent offense against property 27.2%	MRM	SAVRY	27,5%	27,50%
Barnes et al. (2018)	N = 41 Age = 16 Gender = 82.9 Nationality = 85.4	Probation, therapeutic treatment, custodial measure	Domestic violence 31.7%; intimate partner violence 2.4%	Mirall program, child-to-parent violence	Session log, records, satisfaction Questionnaire, SAVRY, URICA.		19,50%
Burceti Solé et al. (2019)	N = 72 Age = 15-18 Gender = 76			Mirall program, child-to-parent violence	Observation sheets		21,42%
Junta de Extremadura (2021)	N = 36 Age = 16.5 Gender = 91.67		Theft 36.11%; breach of a court order 25%; child-parent abuse 13.89%; Assaults 13.89%; sexual offenses 5.56%; public health offenses 2.78%; gender-based violence 2.78%	PIMICA	Activity logs, interview, admissions		

Authors and Year	Sample	Judicial measure	Type of offense	Program	Assessment	Prior records	Reoffending
Palanques-Alegre et al. (2022)	N = 93 (48 CG y 45 TG) Age = 17.1 TG y 16.8 CG Gender = 81.73 TG y 85.4 CG Nationality = 91.1 (TG) and 87.9 (CG)	Therapeutic residential care	TG: 40% Domestic violence, 17.8% attempted homicide, 13.3% violent theft, 11.1% sexual offenses, 6.7% robbery, 6.7% breach of court-order, 4.4% collective security. CG: 25.5% Domestic violence, 23.2% offense against privacy, 12.8% sexual offense, 12.8% robbery, 12.8% collective security, 4.3% violent theft, 4.3% offense against freedom and 4.3% attempted homicide	VR-Assited CBT for impulse control	YLS/OMI, MACI, and EQi; YV, ACS.	53.3% TG 68.7% CG	
Blanch y Mancho (2023)	N = 1112 Age = 14-17 Gender = 79.8% Nationality = 65.4%		Offenses against persons 40.1%; offenses against property 14.9%; non-violent offenses against property 27.6%; others 17.4%	MRM	SAVRY	47%	26,70%

Note: N: sample size; TG: Treatment Group; CG: Comparison group; Gender (%): Percentage of males; Nationality (%): Percentage of spanish; ACS: Coping Scale for Adolescents; AGQ: Aggression Questionnaire; ATIPV: Attitude Toward Interpersonal Peer Violence Scale; BIS-10: Barrat impulsivity; IRI: Interpersonal Reactivity Index; EHS: Social Skills Scale; URICA: University Rhode Island Change Assessment Scale; PPS\_VCJ: Programa de Pensamiento Prosocial Versión Corta para Jóvenes [Short Version of the Prosocial Thinking Program for Young People]; PIMICA: Programa de Intervención con Menores Infractores con Conductas Adictivas [Intervention Program for Juvenile Offenders with Addictive Behaviors]; CBT: Cognitive Behavioral Therapy; VR: Virtual Reality; MRM: Mediation and Reparation programs

For a better visualization and understanding of effectiveness found in psychoeducational programs applied within the Spanish juvenile justice system, a forest plot was presented, displaying the different effect sizes and the average effect estimated through a fixed-effect model. To generate this plot, we have been treated as effect sizes the proportions of repetition, recurrence, reoffending of punishable antisocial behavior in the group that participated in the Psychoeducational program. A global effect size of .20 ( $p < .001$ ) was found, with a 95% confidence interval ranging between .19 and .21 (figure III).

**Figure III:** Forest Plot



Note. FE Model: Fixed-effect model

## Discussion

The purpose of this study has been to analyze the information regarding the effectiveness of psychoeducational programs implemented in the Spanish juvenile justice system. A total of 15 studies with 18 independent studies were included, providing qualitative and quantitative data on their effectiveness. Doctoral theses (Alba et al., 2007; Fernández, 2016; Pintado, 2012; Sánchez, 2008) and studies published by CEJFE stand out for their greater elaboration and empirical support, half of the programs have been carried out in juvenile centres in Catalonia, largely due to the existence of CEJFE, an agency dedicated to research and intervention in the justice system (Capdevila et al., 2017).

The primary aim of this review is to evaluate the effectiveness of programs applied to juvenile offenders, an issue closely linked to the *What*

*Works?* Approach (Pappas & Dent, 2023). In this regard, the reoffending in punishable antisocial behavior is one of the most used indicators in evaluating the effectiveness of intervention programs with juvenile offenders. The non-entry of juvenile offenders into the juvenile justice system is considered an indicator of success in the reintegration process of juveniles (Carbonell et al., 2016; Horcajo-Gil et al., 2019; Ortega et al., 2014; Redondo et al., 2011). However, not all reviewed studies report this data, only 61% reported a relevant global moderate reduction in the comparison group or pre-test. Studies that do not report information on the reoffending in punishable antisocial behavior but report other outcome indicators such as the percentage of participants who finished the programs, attendance, satisfaction and interest in the program. Although these measures are not, from our point of view, the most direct indicators, they represent factors associated with the effectiveness of the program and, therefore, these domains should be considered in its evaluation (Redondo et al., 2011).

The methodology of the studies and the use of assessment instruments are essential factors to be considered in the analysis of the effectiveness of psychoeducational programs applied in juvenile justice (Pappas & Dent, 2023; Redondo et al., 2011). As has been observed, some of the most frequent instruments across different programs were the IGI-J and the SAVRY, widely used in assessment risk and protective factors in juvenile justice (Redondo et al., 2011). The comprehensive assessment of the needs of juvenile's offenders provides relevant information within the framework of juvenile risk and protective factors, which supports the development of interventions adapted to their specific needs (Andrews & Bonta, 2010; Ward & Brown, 2004).

Research focused exclusively on protective factors is still under development (Lipsey, 2014; Pyle et al., 2019). Of all the studies included, only Barnes et al. (2018) and Burcet i Solé et al. (2019) considered protective factors during the intervention, applying a parallel program to the parents of juvenile participants, and Junta de Extremadura (2021) with the inclusion of healthy leisure activities. The studies carried out by Alba et al. (2007), Pintado (2012) and Redondo et al. (2012b) focused on developing and promoting social skills of juveniles (Lipsey, 2009; Redondo et al., 2011). In this regard, the Mirall, PPS-VCJ, and MRM programs stand out, as they have been evaluated in multiple editions and efforts continue toward their improvement (Alba et al., 2007; Barnes et al., 2018; Blanch y Mancho, 2023; Burcet i Solé et al., 2019; Capdevila et al., 2017a,b,c; Ocariz, 2013; Pintado, 2012).

## Conclusions

Regarding the limitations of this study, there are inherent difficulties in conducting reviews, such as publication bias, which has hindered the acquisition of sufficient data required for its calculation; the heterogeneity of studies found, since doctoral theses, reports, intervention manuals, and journal articles. Selection bias is another important consideration, as the studies were identified from different sources, including electronic databases, specialized agencies or organizations, and references of prior studies. Another limitation has been the difficulty in finding journal articles on the effectiveness evaluation of programs in Spain, highlighting the lack of a strong culture of effectiveness evaluation for interventions in applied public policy. Despite the limitations, this systematic review provides relevant Conclusions on the effectiveness of Psychoeducational intervention programs applied to juvenile offenders, with useful practical implications for the development of future intervention programs in juvenile justice.

The results of different programs support the idea that intervention based on the social reintegration of juveniles and psychoeducational approach works it reduces the reoffending in punishable antisocial behavior (Barroso-Hurtado y Serrano, 2019; Redondo et al., 2012a). Detailed assessment of juveniles is highlighted to identify risk and protective factors, which allows addressing their specific needs through program adaptation (Lipsey, 2014; Puig y Capdevila, 2016; Wilson y Lipsey, 2024). In this regard, it is essential to incorporate the methodology and evaluation, outlining the phases, methods, and instruments used (Redondo et al., 2011). Specifically, the use of pre-post measures, group comparison, and post-intervention follow-up is emphasized for a complete evaluation of the effectiveness of the program. However, the effectiveness relies not only on methodology, but also on implementation, highlighting the Need for staff training and sustained supervision. Furthermore, the inclusion of protective factors emerges as a cornerstone in intervention, underlining the inclusion of family, School, and Community, which plays a comprehensive role in juvenile support and not only with their behavior (Aazami et al., 2023; Pappas & Dent, 2023).

These Conclusions have significant implications for both research and practice in juvenile justice. It is suggested to explore new lines of research, such as the study of effectiveness indicators in the implementation of juvenile justice programs, the assessment and intervention of risk and protective factors in antisocial behavior, and the analysis of profiles of juvenile offenders. Reporting of results after implementation of programs is crucial for understanding success factors and limitations in intervention with juvenile offenders (Barroso-Hurtado y Serrano, 2019).

Researchers must be willing to take actions aimed at modifying, improving, and developing programs adapted to the changing and specific needs of the juvenile population in the juvenile justice system. Likewise, it is essential to deepen the study of effectiveness indicators, such as the reoffending in punishable antisocial behavior, or other, and continue working on the development of psychoeducational programs with effectiveness evaluation protocols, with the aim of providing necessary tools and resources for professionals working in juvenile justice.

## Bibliography

- Aazami, A., Valek, R., Ponce, A.N., & Zare, H. (2023). Risk and Protective Factors and Interventions for Reducing Juvenile Delinquency: A Systematic Review. *Social Sciences*, *12*(9), 474. <https://doi.org/10.3390/socsci12090474>
- Andrews, D.A., & Bonta, J. (2010). *The psychology of criminal conduct*. Routledge.
- Anjaswarni, T., Nursalam, N., Widati, S., & Yusuf, A. (2019). Analysis of the Risk Factors Related to the Occurrence of Juvenile Delinquency Behavior. *Jurnal Ners*, *14*(2), 129-136. <http://doi.org/10.20473/jn.v14i2.12465>
- Badasa, G., Gameda, A., Gaduda, B. & Wondimu, B. (2019) Juvenile Delinquency: A Need to Multiple Explanations and Interventions. *Open Access Library Journal*, *6*, 1-10. <http://dx.doi.org/10.4236/oalib.1105904>.
- Baron, R., & Parker, J. (2000). *EQI: YV Baron Emotional Quotient Inventory: Youth version. Technical manual*. Multi-Health Systems Inc.
- Barroso-Hurtado, D., & Serrano, J.B. (2019). Revisión de los factores de éxito en la promoción de comportamientos prosociales como estrategia preventiva en la justicia juvenil en España. *Revista Complutense de Educación*, *30*(1), 75-89. <http://dx.doi.org/10.5209/RCED.55509>
- Borum, R., Bartel, P., & Forth, A. (2003). *Structured Assessment of Violence Risk in Youth. Professional Manual*. Pearson.
- Calvo, S. G. (2007). Evolución de la intervención con menores infractores. *En la calle: revista sobre situaciones de riesgo social*, *8*, 4-6.
- Caride, A. (2005). Educación social y valores cívicos. *Revista de ciencias de la educación*, *201*, 25-43.
- Carbonell, A., Gil-Salmerón, A., & Margaix, E. (2016). Evaluación del riesgo de reincidencia en menores infractores: Herramientas para la mejora de estrategias reeducativas en España. *Revista Internacional de Trabajo Social y Bienestar*, *5*, 79-88.

- Day, A., & Howells, K. (2002). Psychological treatments for rehabilitating offenders: Evidence based practice comes of age, *Australian Psychologist*, 37, 39–47.
- De la Peña, M.E. (2005). *Conducta antisocial en adolescentes: Factores de riesgo y protección*. [Thesis, Complutense University of Madrid].
- Dowden, C., & Andrews, D. A. (1999). What works for female offenders: a meta-analytic review. *Crime and Delinquency*, 45, 438–452.
- Dionne, J. & Altamirano, C. (2012). Los desafíos de un verdadero sistema de justicia juvenil: una visión psicoeducativa. *Universitas Psychologica*, 11(4), 1055-1064.
- Drucker, A. M., Fleming, P., & Chan, A.-W. (2016). Research Techniques Made Simple: Assessing Risk of Bias in Systematic Reviews. *Journal of Investigative Dermatology*, 136(11), e109–e114. <https://doi.org/10.1016/j.jid.2016>.
- Elliott, D. S., Buckley, P. R., Gottfredson, D. C., Hawkins, J. D., & Tolan, P. H. (2020). Evidence-based juvenile justice programs and practices: A critical review. *Criminology & Public Policy*, 19(4), 1305–1328. <https://doi.org/10.1111/1745-9133.12520>
- Evans-Chase, M., & Zhou, H. (2014). A Systematic Review of the Juvenile Justice Intervention Literature: What It Can (and Cannot) Tell Us About What Works with Delinquent Youth. *Crime & Delinquency*, 60(3), 451–470. <https://doi.org/10.1177/0011128712466931>
- García, J., Ortega, E., & de la Fuente, L. (2010). Juvenile offenders' recidivism in Spain: A quantitative revision. In M. Frías-Armenta & V. Corral-Verdugo (Eds.), *Bio-psycho-social perspectives on interpersonal violence* (pp. 333–353). Nova Science Publishers.
- García, J., Zaldívar, F., De la Fuente, L., Ortega, E., & Sáinz-Cantero, B. (2012). El Sistema de Justicia Juvenil de Andalucía: descripción y presentación de resultados a través de la investigación empírica. *Edupsykhé*, 11(2), 287-316.
- Garrido, V., Anyela, L., & Sánchez-Meca, J. (2006a). What works for serious juvenile offenders? A systematic review. *Psicothema*, 18(3), 611-619.
- Garrido, V., López, E., Silva, T., López, M., & Molina, P. (2006b). *El Modelo de la competencia social de la ley de menores: Cómo predecir y evaluar para la intervención educativa*. Tirant lo Blanch.
- Gismero, E. (2000). *Manual de la Escala de Habilidades Sociales*. TEA.
- Graña, J.L., & González, L. (2008). *Historial Criminológico y Social-Versión Juvenil: HCS-J*. Agencia de la Comunidad de Madrid para la Reeducación y Reinserción del Menor Infractor.

- Hoge, R., & Andrews, D. (2006). *Youth Level of Service/Case Management Inventory (YLS/CMI). User's manual*. Multi-Health Systems
- Horcajo-Gil, P.J., Dujo-López, V., Andreu-Rodríguez, J.M., & Marín-Rullán, M. (2019). Valoración y gestión del riesgo de reincidencia delictiva en menores infractores: una revisión de instrumentos. *Anuario de Psicología Jurídica*, 29, 41-53. <https://doi.org/10.5093/apj2018a15>
- Instituto Nacional de Estadística (2024). *Estadísticas de Menores Condenados*. Instituto Nacional de Estadística.
- Jolliffe, D., Farrington, D.P., Pirquero, A.R., Loeber, F., & Hill, K.G. (2017). Systematic review of early risk factors for life-course-persistent, adolescence-limited, and late-onset offenders in prospective longitudinal studies. *Aggression and Violent Behavior*, 33, 15-23. <http://dx.doi.org/10.1016/j.avb.2017.01.009>
- Kaminer, Y., Bukstein, O., & Tarte, R.E. (1991). The Teen-Addiction Severity Index: rationale and reliability. *International Journal on Addiction*, 26(2), 219-26. <https://doi.org/10.3109/10826089109053184>
- Krisberg, B. (2005). Lo que funciona en la justicia juvenil. En *Lo que funciona en la justicia juvenil* (pp. 145-162). Publicaciones SAGE, Inc., <https://doi.org/10.4135/9781483328836.n8>
- Lab, S.P., & Whitehead, J.T. (1988). An Analysis of Juvenile Correctional Treatment. *Crime & Delinquency*, 34(1), 60-83. <https://doi.org/10.1177/0011128788034001004>
- Lahey, B., Moffitt, T.E. & Caspi, A. (2003). *The causes of conduct disorder and serious juvenile delinquency*. Guilford.
- Latessa, E.J. (2004). From theory to practice, what works in reducing recidivism? In *State of Crime and Justice in Ohio*. Ohio Office of Criminal Justice Services. [https://www.uc.edu/content/dam/uc/ccjr/docs/articles/Theory\\_Practice.pdf](https://www.uc.edu/content/dam/uc/ccjr/docs/articles/Theory_Practice.pdf)
- Ley Orgánica 5/2000 de 12 de enero, reguladora de la responsabilidad penal de los menores. *Boletín Oficial del Estado*, 11, de 13 de enero del 2000.
- Lipsey, M.W. (2009). The Primary Factors that Characterize Effective Interventions with Juvenile Offenders: A Meta-Analytic Overview. *Victims & Offenders*, 4(2), 124-147. <https://doi.org/10.1080/15564880802612573>
- Lipsey, M.W. (2014). Interventions for Juvenile Offenders: A Serendipitous Journey. *Criminology Public Policy*, 13(1), 1-14.
- Lipsey, M.W., Landenberger, N.A., & Wilson, S.J. (2007). Effects of cognitive-behavioral programs for criminal offenders. *Campbell Systematic Reviews*, 3(1), 1-27.
- Looman, J., & Abracen, J. (2013). The risk need responsivity model of offender rehabilitation: Is there really a need for a paradigm shift? *In-*

- ternational Journal of Behavioral Consultation and Therapy*, 8(3-4), 30-36. <http://dx.doi.org/10.1037/h0100980>
- Mampaso, J., Pérez-Fernández, F., Corbí, B., González, M., & Bernabé, B. (2014). Factores de riesgo y de protección en menores infractores. *Análisis y prospectiva. Psychologia Latina*, 5(1), 11-20.
- Martinson, R. (1974). What works? Questions and answers about prison reform. *The Public Interest*, 10, 22-54.
- Menon, S.E., & Cheung, M. (2018). Desistance-Focused Treatment and Asset-Based Programming for Juvenile Offender Reintegration: A Review of Research Evidence. *Child and Adolescent Social Work Journal*, 35(5), 459-476. <https://doi.org/10.1007/s10560-018-0542-8>
- Millon, T. (1993). *Manual of Millon Adolescent Clinical Inventory*. NCS.
- Moffitt, T.E. (2018) Male antisocial behaviour in adolescence and beyond. *Nature Human Behaviour*, 2, 177-186. <https://pubmed.ncbi.nlm.nih.gov/30271880/>
- McGuinness, L.A., & Higgins, J.P.T. (2021) Risk-of-bias VISualization (robvis): An R package and Shiny web app for visualizing risk-of-bias assessments. *Research Synthesis Methods*, 12(1), 55-61. <https://doi.org/10.1002/jrsm.1411>
- Organización Mundial de la Salud (2020). *Informe sobre la situación mundial de la prevención de la violencia contra los niños*. Organización Mundial de la Salud.
- Ortega, E., García, J., & Frías, M. (2014). Meta-análisis de la reincidencia criminal en menores: estudio de la investigación española. *Revista Mexicana de Psicología*, 31(2), 111-123.
- Page, M.J., McKenzie, J.E., Bossuyt, P.M., Boutron, I., Hoffmann, T.C., Mulrow, C.D., et al. (2021). The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *BMJ*, 372, n71. <http://dx.doi.org/10.1136/bmj.n71>
- Pappas, L.N., & Dent, A.L. (2023). The 40-year debate: a meta-review on what works for juvenile offenders. *Journal of Experimental Criminology*, 19, 1-30. <https://doi.org/10.1007/s11292-021-09472-z>
- Puig, M.F., & Capdevila, M.C. (2016). Intervenciones con jóvenes infractores en el marco de una medida judicial. *Infancia, juventud y ley: Revista de divulgación científica del trabajo con menores*, 7, 61-69.
- Pyle, N., Flower, A., Williams, J., & Fall, A.M. (2019). Social Risk Factors of institutionalized Juvenile Offenders: A Systematic Review. *Adolescente Research Review*, 5, 173-186. <https://doi.org/10.1007/s40894-019-00120-2>.
- Redondo, S., & Martínez, A. (2013). Evaluación criminológica de la justicia juvenil en España. *Cuadernos de Política Criminal*, 2(110), 189-220.

- Redondo, S., Martínez, A., & Andrés, A. (2011). *Factores de éxito asociados a los programas de intervención con menores infractores*. Ministerio de Sanidad, Política Social e Igualdad.
- Redondo, S., Martínez, A., & Andrés-Pueyo, A. (2012a). Intervenciones con delincuentes juveniles en el marco de justicia: investigación y aplicaciones. *EduPsykbé*, 11(2), 143-169.
- Redondo, S., & Sánchez-Meca, J. (2003). Guía de tratamientos psicológicos eficaces para la delincuencia juvenil. En Pérez, M., Fernández Hermida, J.R., Fernández Rodríguez, C., y Amigo Vázquez, I. *Guía de tratamientos psicológicos eficaces III. Infancia y adolescencia* (págs. 183-214). Pirámide.
- Rutter, M., Giller, H., & Hagell, A. (2000). *La conducta antisocial de los jóvenes*. Cambridge University Press.
- Salazar J.G., Torres-López, T.M., Reynaldos, C., Figueroa, N.S., & Araiza, A. (2011). Factores asociados a la delincuencia en adolescentes de Guadalajara, Jalisco. *Papeles de población*, 17(68), 103-126.
- Souverein, F., Dekkers, T., Bulanovaitė, E., Doreleijers, T., Hales, H., Kaltiala-Heino, R., & Van Der Pol, T. (2019). Overview of European forensic youth care: Towards an integrative mission for prevention and intervention strategies for juvenile offenders. *Child and adolescent psychiatry and mental health*, 13, 1-6.
- Schwalbe, C.S. (2007). Risk assessment for juvenile justice: A meta-analysis. *Law and Human Behavior*, 31, 449-462.
- Smith, D. J. (2005). The effectiveness of the juvenile justice system. *Criminal Justice*, 5(2), 181-195. <https://doi.org/10.1177/1466802505053497>
- Sterne, J.A., Hernán, M.A., Reeves, B.C., Savović, J., Berkman, N.D., Viswanathan, M., et al., (2016). ROBINS-I: A tool for assessing risk of bias in non-randomised studies of interventions. *BMJ*, 355, i4919. <https://doi.org/10.1136/bmj.i4919>
- The Jamovi Project (2024). Jamovi (Version 2.5) [Computer Software]. Retrieved from <https://www.jamovi.org>
- Vallés, M.D., & Hilterman, E.L.B. (2006). *SAVRY: Manual per a la valoració estructurada de risc de violència en joves*. Centro de Estudios Jurídicos y Formación Especializado.
- Van Der Put, C.E., Deković, M., Stams, G.J.J., Van Der Laan, P.H., Hoeve, M., & Van Amelsfort, L. (2011). Changes in risk factors during adolescence: Implications for risk assessment. *Criminal Justice and Behavior*, 38(3), 248-262.
- Villafuerte-Díaz, A., Ramos, P., Rivera, F., & Moreno, C. (2022). Conducta antisocial en adolescentes españoles prevalencia y relación con su

- salud global percibida. *Psicología Conductual*, 30(2), 641-661. <https://doi.org/10.51668/bp.8322303s>
- Ward, T., & Brown, M. (2004). The good lives model and conceptual issues in offender rehabilitation. *Psychology, Crime & Law*, 10(3), 243-257, <http://dx.doi.org/10.1080/10683160410001662744>
- Wilson, D.B., & Lipsey, M.W. (2024). Scaling up effective juvenile delinquency programs by focusing on change levers: Evidence from a large meta-analysis. *Criminology & Public Policy*, 23(2), 261-286. <https://doi.org/10.1111/1745-9133.12663>
- Young, S., Greer, B. & Church, R. (2017). Juvenile delinquency, welfare, justice and therapeutic interventions: a global perspective. *National Library of Medicine*, 41 (1), 21-29. <https://doi.org/10.1192%2Fpb.bp.115.052274>
- Zahn, M.A., Day, J.C., Mihalic, S.F., & Tichavsky, L. (2009). Determining What Works for Girls in the Juvenile Justice System: A Summary of Evaluation Evidence. *Crime & Delinquency*, 55(2), 266–293. <https://doi.org/10.1177/0011128708330649>
- Zelaya, A.M. (2020). *A Review of Common Treatments and Their Effectiveness to Reduce Juvenile Recidivism*. [Thesis, The Chicago School of Professional Psychology].



# Sowing the Seeds of a Lifelong Journey that Begins in Childhood: Sociocultural Learning in *Eskola Txikiak*<sup>1</sup>

## Sembrando el proyecto vital desde la infancia: Aprendizaje Sociocultural en *Eskola Txikiak*

<https://doi.org/10.4438/1988-592X-RE-2025-412-738>

**Garazi Ormazabal-Arizkorreta**

<https://orcid.org/0000-0002-2554-9484>

Universidad del País Vasco

**Andoni Arguiñano Madrazo**

<https://orcid.org/0000-0002-9585-8303>

Universidad del País Vasco

**Inaki Karrera Xuarros**

<https://orcid.org/0000-0003-4448-7749>

Universidad del País Vasco

### Abstract

In the pluralistic, ever-changing, and complex world we inhabit, rural schools have become a setting of great interest due to their potential for innovation and improvement through the development of fundamental, pedagogical approaches. Given this scenario, the study at hand aims to broaden the understanding of good practices in a specific context, which is rural education. This paper focuses on four schools in the Basque Country of Spain, which are part of the educational network known as *Eskola Txikiak*. The purpose of this research is to analyse the core pedagogy of *Eskola Txikiak* by highlighting the benefits it offers to both children and teachers. Data collection was conducted through 420 hours of observation sessions, eight focus groups, and an analysis of the schools' internal

---

<sup>1</sup> **FUNDING DETAILS.** This work was supported by the [UPV/EHU] under grant [code: PIF 21/217]; [Basque Government] under grant [code: PIBA\_2023\_1\_0025]; [Basque Government, Education Department] under grant [code: hezkuntza\_23/12].

documentation. After categorising and coding the information, the findings reveal several key factors underlying the sociocultural learning that takes place in the educational workshops at these schools. Specifically, the results obtained indicate the presence of active participation in the learning process, creative distribution of school space, experimentation, and the availability of a wide range of materials and personal resources with which children can interact, all of which foster learning that is child-centred, creative, personalised, and of high quality.

*Keywords:* students, empowerment, well-being, rural schools, workshops.

### Resumen

Dentro del mundo plural, cambiante y complejo en el que nos encontramos, la escuela rural, desarrollando claves pedagógicas de referencia, se ha convertido en un contexto de gran interés por su capacidad de innovación y mejora. En esta línea, el presente estudio nace con la intención de contribuir al conocimiento de buenas prácticas en un contexto particular como es el de la educación rural. Situándonos en cuatro escuelas de la red *Eskola Txikiak* (concepto parecido a las escuelas rurales) del País Vasco, la presente investigación pretende analizar la actividad pedagógica principal de las *Eskola Txikiak*, destacando los beneficios que ésta aporta tanto al alumnado como al profesorado. La recogida de información se ha completado mediante, 420 horas de sesiones de observación, 8 *Focus Groups* y el análisis de la documentación interna. Después de categorizar y codificar la información, los resultados obtenidos muestran varias de las claves que subyacen en los procesos de aprendizaje sociocultural que surgen en los talleres pedagógicos. En concreto, los resultados obtenidos en este estudio indican la participación en el proceso de aprendizaje, así como la distribución creativa del espacio escolar, la experimentación y la disponibilidad de una amplia gama de recursos materiales y personales con los que los estudiantes pueden interactuar, promueven un aprendizaje centrado en el estudiante, creativo, personalizado y de alta calidad.

*Palabras clave:* alumnado, empoderamiento, bienestar, escuelas rurales, talleres.

## Introduction

As pointed out by Bauman (2003) and Díaz (2012), we are currently immersed in a pluralistic world, which is both intricate and volatile. A few decades ago, the term VUCA was used to describe the society of that time, referring to Volatile, Uncertain, Complex, and Ambiguous (Urbano-Carazo, 2022). However, this concept is no longer enough to reflect the new paradigm. For this reason, Cascio (2023) proposed the term BANI to describe both today's society and future events, referring to environments that are Brittle, Anxious, Non-linear, and Incomprehensible.

Given the global context, any type of innovative proposal in the field of education must take into account these characteristics in order to avoid being disconnected from the daily lives of citizens (Martínez & Rogero, 2021). As affirmed by Vergara (2016, p.6), “we need citizens who are able to understand the complexity of situations and the exponential increase in information, and who have the ability to adapt creatively to the speed of change, and to the uncertainty that accompanies such transformations”. Consequently, this new context makes it necessary to develop critical thinking skills in order to assist in the decision-making process and help us engage in problem-solving activities (Pérez, 2019). In doing so, we will be able to reinvent and update our approaches, which will allow us to adapt to the current and future characteristics of society, as well as to the needs that may arise therein (Mena, 2018).

With regard to creative thinking, schools have an obligation to encourage investigative curiosity from childhood onward (European Commission, 2016; Freinet, 1979; Tonucci, 2012), due to the fact that “curiosity is a type of intrinsic motivation that is essential for fostering active learning and spontaneous exploration” (Oudeyer et al., 2016, p.1). Curiosity is the beginning of knowledge, as it drives us to observe, examine, question, discover, test, and interact with the world (Ortiz and Cervantes, 2015). For this reason, we must acknowledge and honour children’s innate curiosity by incorporating the scientific learning process into the school environment (Tonucci, 2001; 2012). The development of inquisitive skills allows children to enhance their abilities regarding creativity, enquiry, innovation, and the ability to create new knowledge (Oquendo, 2019). Moreover, these are essential factors for responding effectively to the challenges of today’s world (Unesco, 2015).

But how can we promote this scientific mindset? Starting in childhood, kids develop their own theories about what is happening around them, similar to the method used by scientists, which is through experiments involving trial and error (Tonucci, 1995). In this regard, we must instil confidence in children in order to make them aware of their current knowledge, as well as their potential to acquire new knowledge (Tonucci, 1995; 2001). However, this implies seeing children as individuals who not only have rights, but a strong potential for development as well (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017).

Likewise, to keep this inquisitive attitude alive, we must acknowledge the importance of children’s opinions. As stated in the Convention on the Rights of the Child, children are entitled to express their opinions, make decisions, and be heard in all matters that concern them (UNICEF, 2015).

Therefore, it is essential to listen to children, give them a voice, and find out what interests them (Basasoro, 2021; Decroly, 2006; Freire, 1997; Rabadan, 2021). In this regard, as pointed out by various leaders of the New School movement, it is essential that children become active participants in their own learning (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Hoyuelos, 2021; Malaguzzi, 1996; Montessori, 1937, 1982; Tonucci, 1995, 2001, 2012, 2017). If we accept this premise, when kids are placed in different situations where they must use various strategies and make their own decisions, they develop the skills needed to deal with any problem. In other words, they learn to manage themselves and gather information autonomously, which allows them to assess and determine the strategies they need to use to solve a problem (Guamán & Espinoza, 2022). Thus, schools must become enriching environments where children have the opportunity to choose between various options (Freinet, 1979; Hall et al., 2014; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017), thereby offering them various learning opportunities.

According to the sociocultural theory of Vygotsky, or interaction-based learning (1978), people learn and develop critical-thinking skills through social interaction. In other words, learning does not take place individually. Instead, it is influenced by a person's social setting, which includes the cultural relationships we establish with our environment. From this point of view, social interaction and participation in daily activities reflect some of the strongest cultural experiences, as they give us access to information, culture, and imagination, among other things (Gómez et al., 2020; Lave & Packer, 2001). However, we must evaluate the conditions for promoting such engagement (Gómez et al., 2020). Along the same lines, as pointed out by Freire (1970), "There are no absolute ignoramuses, nor absolute sages; there are only men who seek to learn more through communication" (p. 87). Thus, egalitarian dialogue is necessary, as this horizontal approach allows children, teachers, and educational agents to jointly construct the teaching-learning process.

As rural schools are considered the ideal setting for innovation and improvement, they are able to take advantage of what have historically been viewed as shortcomings, thereby developing novel and essential pedagogical approaches (Santos, 2011). This has been confirmed by various national and international studies, which highlight the way in which rural schools foster exemplary teaching techniques, including the following: varied groupings of children to promote age diversity (Åberg-Bengtsson, 2009; Boix, 2011; Little, 2001; Santos 2011); autonomous and contextualised learning (Abós, 2015; Boix, 2011), and interdisciplinary (Boix, 2011); collaborative learning (Abós, 2015; Boix, 2011);

Santos, 2011); diverse relational dynamics between children (Matías and Vigo, 2020), as well as horizontal relationships between children and teachers (Santos, 2011); a novel role for teachers (Vigo and Soriano, 2015); a new way of relating and using the setting as a pedagogical resource (Abós, 2019; Boix, 2011); more family participation (Vigo and Soriano, 2015); and finally, an emergent curriculum (Arguiñano, 2019; Karrera and Arguiñano, 2018).

However, in spite of these findings, several studies show that apart from the fact that the situation of rural schools is not well known, these centres of learning have failed to capture the interest of either educational administrations or academia (Little, 2001). In addition, the research at hand confirms that studies carried out to date reflect this attitude. Moreover, this viewpoint is expressed in research conducted in this specific context, which mainly focuses on historical and legislative developments (Aguirregoitia-Güenechea, 2024), and on the key factors involved in making these schools inclusive (Arguiñano, 2019; Karrera et al., 2020). Therefore, in the words of Little (2001), this type of study aims to “make the invisible visible”. In fact, as various research papers have pointed out, it is essential to broaden our understanding of the educational approaches used at various rural schools (Little, 2001; Santos-Gelvasio, 2022). Furthermore, given that these environments have unique characteristics, the pedagogical strategies that take place in these settings require an even deeper understanding (Abós, 2020; De la Vega, 2020).

As such, with the aim of expanding the knowledge regarding good practices in a specific context, this study focuses on various schools in the *Eskola Txikiak* educational network in the Basque Country. Thus, considering the pedagogical features of these schools, the objectives of this research are as follows:

- Analyse the main pedagogical practices at *Eskola Txikiak*.
- Highlight the benefits they bring to both children and teachers.

## Methodology

This research is a critical ethnographic study (Karrera, 2008). Furthermore, ethnography is an appropriate method for investigating any phenomenon related to school organisation, classroom life, and the relationships between schools and their sociocultural environments (Maturana & Garzón, 2015). In this regard, as the objective of this study is to highlight innovative and inspiring educational experiences at *Eskola Txikiak*, the aforementioned method has been chosen. Thus, in addition to bestowing

value and relevance on the emotions, thoughts, and experiences of children and teachers, this research has created different spaces of reflection to raise awareness of the intrinsic value of these participants. As pointed out by Beach and Vigo-Arazola (2021), every person has knowledge, along with the ability to use it. Likewise, referring to the nature of this research, critical ethnography goes beyond mere description and understanding. In fact, it is committed to social change and transformation (McLaren, 2005). Thus, as affirmed by Rodríguez (2022, p. 13), “critical ethnography should be used by us, the educators, in order to turn our campuses and teaching work into spaces for reflection on the needs of students”. Moreover, the critical nature of this research has enabled a collaborative dialogue with the educational community (Denzin, 2018), which has helped to establish shared goals in order to improve the learning experience of both educational and research communities.

## Sample

### Context and participants

*Eskola Txikiak* is a concept similar to rural schools. These are public schools located in rural areas which, based on their framework and context, offer a variety of educational proposals. The participating schools are located in towns with less than 2,000 inhabitants in rural areas of the historical territory known as Gipuzkoa, in the Basque Country of Spain. Regarding the socioeconomic and cultural aspects of these communities, based on data from the Basque Statistics Institute (Eustat, 2021), these towns have a high socioeconomic level and an immigration rate of close to 10%, which is slightly higher than the rate of 8.5% for the entire Basque Autonomous Region (Eustat, 2022).

In terms of economics, the industrial sector dominates in most of these communities, playing a key role in local business development (Eustat, 2024). However, it is important to mention another study, which also explored the features of these and other rural towns in the Basque Autonomous Region. In that study, it was found that these communities need special attention due to their limited services related to the quality of life and well-being of citizens (Basque Government, 2022).

Regarding the participating schools, they are organised around comprehensive workshops (Trueba, 1989), where curricular knowledge and learning converge in a way that is integrated and interdisciplinary. Classrooms become shared spaces dedicated to specific subjects, examples of which include art, science, and languages. Consequently, both the spaces

and the materials are reorganised in a specific way. During the school day, children rotate through the various workshops, with one teacher assigned to each workshop. Another interesting point is that although the children range from 3 to 12 years of age, interact with each other in the same spaces (Eskola Txikiak, 2024). Thus, children in these schools are allowed to choose the workshops they want to attend, and the tasks or projects they wish to carry out.

After two meetings with both the current and former coordinators of the *Eskola Txikiak* network, four schools were chosen for this study. The following aspects were considered for the selection: the educational history of each school; its characteristics; and the pedagogical approach used in each institution. Table 1 shows the schools selected and some of their relevant attributes.

**Table 1.** Participating schools and the key features used for the selection

School	Key Features	Comprehensive workshops at each school
Larraul Infant and Primary School (Larraul, Gipuzkoa).	<ul style="list-style-type: none"> <li>- Children aged 3 to 12</li> <li>- Number of children: 19</li> </ul>	<ul style="list-style-type: none"> <li>28. Library</li> <li>29. Science</li> <li>30. Art</li> <li>31. Carpentry</li> <li>32. Bodily and musical expression</li> <li>33. Symbolic play space</li> </ul>
Zerain Infant and Primary School (Zerain, Gipuzkoa).	<ul style="list-style-type: none"> <li>- Children aged 3 to 7</li> <li>- Number of children: 17</li> </ul>	<ul style="list-style-type: none"> <li>1. Library</li> <li>2. Science</li> <li>3. Art</li> <li>4. Bodily expression</li> <li>5. Carpentry</li> <li>6. Cooking</li> </ul>
Olaberria Infant and Primary School (Olaberria, Gipuzkoa).	<ul style="list-style-type: none"> <li>- Children aged 3 to 12</li> <li>- Number of children: 33</li> </ul>	<ul style="list-style-type: none"> <li>1. Library</li> <li>2. Science</li> <li>3. Art</li> <li>4. Carpentry</li> <li>5. Bodily expression</li> <li>6. Movement</li> </ul>
Ezkio-Itsaso Infant and Primary School (Ezkio, Gipuzkoa).	<ul style="list-style-type: none"> <li>- Children aged 3 to 11</li> <li>- Number of children: 51</li> </ul>	<ul style="list-style-type: none"> <li>1. Language</li> <li>2. Science</li> <li>3. Creativity (art-focused classroom)</li> <li>4. Artists' corner (classroom focused on physical and musical expression)</li> <li>5. Carpentry</li> <li>6. Sewing workshop</li> <li>7. Historical storytelling (classroom focused on symbolic play)</li> </ul>

## Instruments

### Information gathering techniques

These were the techniques used to collect information:

- **Participant observation.** A total of 420 hours were carried out (Hammersley, 2017), in which the research team visited each school for a period of three weeks, spending a full day with the children and teachers. During these sessions, observations were carried out of the entire school itself, the participants, and the various pedagogical practices used in these centres. The research team engaged in hands-on participation in the various workshop activities in order to gain a thorough understanding of both the children's tasks and the context of the work carried out. The information from the observation sessions was recorded using field notes and graphic depictions and was later organised into an observation notebook using a categorical system. A separate section was also included for information that arose spontaneously.
- **Documentation.** The internal documents of each school were gathered, along with those of the educational network. Next, the documents related to each school's educational projects were analysed, in addition to information outlining the network's pedagogical principles.
- **Focus Groups.** Consistent with the proposal of Hammersley (1999), this study includes the experiences, opinions, and actions taken by the children and teachers of these schools through focus groups. Two focus groups were held at each school: one with children and one with teachers. Thus, a total of eight focus groups were conducted involving 28 children and 11 teachers. As the selection of the children was made with the help of the teachers at each school, this procedure adhered to the principle of active participation. Two factors were considered for the selection: gender parity and age diversity. Between six and eight children participated in the children focus groups. As for the teachers, priority was given to those who had more knowledge and experience with the school's educational approach. Thus, given the small staff at each school, a maximum of three teachers participated in each focus group.

## Procedure

### Information analysis

The content analysis was based on the stages proposed by Karrera (2008). In the first stage, the pertinent information from the observations, documentation, and focus groups was selected. Next, inductive categorisation was carried out, followed by subsequent coding of the information with the help of the *Nvivo v.12* programme. After completing this step, the content analysis and triangulation process were initiated. The emerging topics obtained through the collection of information are as follows:

- The learning that took place and the projects that emerged from the comprehensive workshops.
- Pedagogical foundations and contributions to sociocultural learning.

The coding was performed by creating and linking the various codes to the corresponding topics, sources of information, and participants, as seen in the following table:

**Table II.** Codes referring to the topics, sources of information, and participants

TOPICS CODES	
The learning and projects that emerged from the comprehensive workshops	A
Foundations and pedagogical contributions to sociocultural learning	IIP
INFORMATION SOURCE CODES	
Field notes from the observations	LO
Focus Group (children)	IKTE
Focus Group (teachers)	IRTE
Documents	DOC
PARTICIPANT CODES	
Teacher	IR
Children and their ages	IK (age is indicated in quotation marks and italics)

The document was returned to the participants to ensure the accuracy of the findings. As pointed out by Karrera (2008), in order to carry out a meaningful analysis, it is necessary to go beyond the aforementioned procedures and perform diverse triangulations. This guarantees

the value, reliability, and applicability of the research (Vallejos & Finol de Franco, 2009), all of which ensures the accuracy of the results obtained and minimizes potential interpretation biases (Karrera, 2008). In this regard, it is important to note that this research complies with the ethical requirements established by the Ethics Committee for Research with Human Subjects (CEISH-UPV/EHU). In fact, the committee itself approved the study (code: M10/2022/339). Likewise, the approval and informed consent of all the participants, as well as their legal guardians, was procured for this research.

## Results

This section presents the results obtained using the previously defined topics.

### **The learning and projects that emerged from the comprehensive workshops**

Given the variety of comprehensive workshops presented in Table 1, each school offers children the opportunity to choose the space they want to attend, as well as the task or project they wish to carry out. This means that children also decide how to approach the job, in addition to choosing the material and personal resources they will use to complete their work. This is how a typical day starts at one of these schools: *“In the morning, the students and teachers gather at the school entrance to welcome each other and start the day. A teacher greets the children and asks if anyone has anything to share with the group. On this particular day, the children had no specific comment to share, so the teacher proceeded to explain the plan for the morning. She outlined the suggested reading material for that day and the available workshops”* (A\_OLO\_1).

There are several reasons why children choose one workshop over another. On the one hand, most children say they choose certain spaces because they can do the activities they enjoy in those areas. One girl adds that she chooses certain workshops because in addition to finding several features she likes in those spaces, she feels more inspired in those areas. As a result, she sees herself as more skilled in the tasks she performs: *“My favourite places are the creativity, science, and movement workshops. Creativity because I like it a lot, and I’m good at it, and I also love the things you find in that place. I go there a lot and I get many ideas for fun things to do”* (A\_OIKTE\_IK8. “11”\_10). On the other hand, children also attend these workshops when they feel they need to improve in a specific area: *“Well, I really don’t like the library workshop too much, but I usually go quite often because I have to learn”* (A\_LIKTE\_IK 16. “12”\_6-7). The

children point out that through perseverance, they eventually enjoy the process and learn from it: *“I go to the science and language workshops so much that I’ve really started to like it...”* (A\_EIKTE\_IK40. “12”\_1).

Children use various sources of inspiration to start a new project. For example, a child might bring an idea from home: *“Well, maybe I see that I need a coat rack at home, so I decide to make a coat rack, but if I change my mind and think of something else, I make that”* (IIP\_EIKTE\_IK39. “11”\_2-3). In other cases, their classmates become a source of inspiration: *“Maybe I see someone doing something, and I like the idea, so I do it”* (IIP\_LIKTE\_IK22. “9”\_27-28). When the kids are unable to decide what to do, they ask for time to think about it, yet aware that this step also has a limit: *“Thinking... many times, yea, let’s see, maybe not all day, but a lot of times...”* (IIP\_LIKTE\_IK22. “9”\_32). Likewise, children often seek inspiration from books, materials and other resources from the educational workshop: *“Well, sometimes you go to the art workshop, and you don’t know what to do, so you pick up some books, get some ideas, and start making things...”* (IIP\_LIKTE\_IK15. “10”\_32). In a similar vein, instead of starting a new project or challenge, they sometimes take on the role of helping their classmates: *“Two students were making a race track with pieces of wood, but it fell to the floor and broke into pieces. A classmate was watching them the whole time. The teacher suggested that she help, and the girl started to help her two classmates”* (IIP\_ELO\_48).

Once the children decide on a project, each school uses various techniques to help the kids make a detailed plan of their chosen venture: *“In the science workshop, before starting a project, all the students have to fill in a form (...). On the form, they have to answer the following questions: “What do I want to know? What do I know? What do I need? What activities will I do? What will be the final product? What will I learn?”* (IIP\_ELO\_16). In these cases, one teacher comments that when the children start a new project, the teacher’s help is essential: *“You see, they start a project, but without planning it, you know? So you have to help them plan the steps they need to take”* (IIP\_OIRTE\_IR2\_6).

It also bears mentioning that projects sometimes do not turn out as expected. In these cases, the children acknowledge that by experimenting and working on the same project, they manage to improve it and complete it: *“Well, in the language class I did a project on Joxe Miguel Barandiarán, and it didn’t turn out so good at first, but then I improved it, and it turned out better. And in the end, the others learned from it, and I learned too.”* (IIP\_EIKTE\_IK43. “11”\_36).

All the schools provide a space for children to share and present their work: *“After completing a project on some of the countries of Europe, one*

*student agreed to give a presentation on her work to the entire school. All the pre and primary school students listened to her presentation. The teacher also asked her several questions, which allowed her to share everything she had learned” (IIP\_OLO\_2). In this way, the kids become role models and a source of learning for their classmates: “It’s really true, because listening to the presentations helps the other kids learn. Like if I do a project on birds and a classmate doesn’t know much about birds, she can come to my presentation and learn about them...” (IIP\_EIKTE\_IK39. “11”\_30).*

### **Foundations and pedagogical contributions to sociocultural learning**

Teachers emphasise the importance of giving children the opportunity to experiment, which allows them to understand and learn different aspects of various subjects: *“In the end, it’s about doing and doing. They have to try. They have to experiment by planning and doing. But doing doesn’t always mean producing. Sometimes it’s observing, sometimes it’s writing, and other times it’s experimentation. But they have to try (...) because in the end, it’s not about internalising the content. It’s also about acquiring certain skills, and skills are acquired by doing. There’s just no other way” (IIP\_LIRTE\_IR12\_2). Another teacher emphasised the need for children to touch, feel, and experience the work they do: “It should be hands-on, don’t you think? There’s an old saying that goes, ‘I’ll learn everything I experience’. That’s really what it’s all about...” (IIP\_EIRTE\_IR35\_3-4).*

Children are also aware of this type of learning, and they emphasise the value of practical experience: *“You know, you learn things by doing them” (IIP\_ZIKTE\_IK33. “8” 5-6). Moreover, the children explain how they gain diverse knowledge through practice: “In the kitchen, if we don’t measure things right, it doesn’t turn out well, so we learn that we have to measure things correctly for something to come out right. It also happens in carpentry, and in the creativity workshop too” (IIP\_OIKTE\_IK8. “11”\_2). Likewise, children value being able to devote time and effort to one project: “So, in a project you’re always doing the same thing over and over, so it stays in your mind, and you learn it” (IIP\_EIKTE\_IK39. “11”\_9).*

Along the same lines, several teachers emphasise the importance of offering attractive and fun activities, highlighting the role of playfulness in the learning process: *“All of this is sometimes related to play, or to the students doing projects they like to do, you know?” (IIP\_LIRTE\_IR12\_4). Another teacher expands on this idea: “Even though any activity can be turned into game, it is certainly possible to learn from it, especially in childhood; I truly believe children learn through playing, and there’s really no other way at that age...” (IIP\_EIRTE\_IR37\_3-4).*

In response to this approach, one teacher makes the following comment: *“In my opinion, they don’t know exactly where the line is drawn. Sometimes, what we consider work might be play for them, and what we consider play may be work for them”* (IIP\_LIRTE\_IR13\_5). The vast majority of children believe it is possible to learn and play at the same time, as they are aware of the educational value of participating in activities that are attractive and fun. One child simply says, *“We learn by playing”* (IIP\_LIKTE\_IK15. “10”\_1). In a similar vein, it is important to note that kids are aware of the interdisciplinary value of each learning experience: *“You know, maths is in everything, and everything is inside everything else. When you build a house, you have to measure things, and then... you have to add and measure correctly, and all of that is maths...”* (IIP\_OIKTE\_IK7. “12”\_5).

Likewise, some children start to identify the subjects they find interesting, and they make decisions about them from a very early age: *“Some of the children are only 3 years old, yet they have started to choose a certain workshop because they like it; the library, for example... even though they know the pre-school teacher will be in the expression workshop, they go to the library, and there, at only 3 years old, they make a choice...”* (A\_OIRTE\_IR3\_8). In this regard, the children associate the opportunity to learn something interesting with the effort they make before and during the learning process: *“You know, if you’re working on a subject you don’t really like, you don’t put as much effort into it (...) but if you like it... you get excited about doing it”* (A\_OIKTE\_IK7. “12”\_4). Similarly, several children say they usually organise themselves in advance in order to finish all the work they have planned to do, as they cannot always go to all the workshops: *“It depends on when they have the workshops... like the artists’ workshop is only open on Mondays and Tuesdays, so I have to decide if I will go on Monday or Tuesday, because if I don’t go on one of those day, I won’t be able to go”* (A\_EIKTE\_IK43. “11”\_6).

Thus, children from all four schools agree that they enjoy having the option to choose where to go, what work to do, and with whom to do it. They believe this makes their projects easier, as it doesn’t feel like a task that is strictly defined or directed: *“You do things you like, about subjects you like”* (A\_LIKTE\_IK17. “11”\_10). Along the same lines, children say they feel more relaxed working this way, and the teachers concur. As one teacher points out, *“They’re working on something they really like, and that’s a big relief for them”* (A\_OIRTE\_IR2\_3).

Another teacher comments that this style of working allows both children and teachers to discover each person’s strengths: *“That’s absolutely right. And what’s more, it becomes very clear in the workshops. Every child has their strengths, you know?”* (A\_OIRTE\_IR4\_5). Another teacher

expands on the idea and asserts that this type of environment reduces comparison: *“My work isn’t better than yours. I’m not smarter than you. I’m not more important than you... because you have your project, and I have mine. And when I present my work, it has value, and when you present your project, it also has value”* (A\_LIRTE\_IR12\_6).

## Discussion and conclusions

This study is based on the participation of children and the teaching staff at four rural schools. The aim is to expand the limited knowledge of various educational experiences in rural settings, the final purpose of which is to encourage reflection and promote improvement in the educational system as a whole (Rodríguez, 2022). As affirmed by Tonucci (1996, p. 90), “Rural schools could become a useful laboratory for educational reform, which might even be extrapolated to city schools”. Therefore, it is imperative to optimise the opportunities inherent to critical, ethnographic research, as affirmed by Denzin (2017), who argues that citizens have a duty to promote a global transformation that fosters a full, inclusive, and participatory democracy.

It should also be noted that the findings included in this article are based on the experiences and opinions of children aged 6 to 12, which have considerable weight, as they are the ones who have lived through the entire gamut of innovative educational strategies that take place in these schools. This clearly shows that if we want to propose pedagogical experiences that are consistent with the real situation and needs of children, it is crucial to listen to their opinions and learn from their experiences (Tonucci, 1995, 2001, 2012, 2017). In this regard, the conclusions of this study confirm that by placing children at the centre of education, pedagogical efforts can have a powerful impact. As reflected in the pedagogical proposals of Decroly, Freinet, Malaguzzi, Montessori, and Tonucci, such efforts must start with listening (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Hoyuelos, 2021; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017). According to Freire (2006, p. 71), “Children must be guaranteed the right to learn how to decide, which can only be done by deciding”. As shown in the present study, the learning approach at *Eskola Txikiak* is focused specifically on this aspect, as each project begins with a process of active listening. One teacher puts it this way: *“Without a doubt, the important thing is that they have the right to decide at all times”* (A\_LIRTE\_IR12\_2).

Obviously, not all children feel, think, and act in the same way, nor do they have the same needs and concerns (Decroly, 2006; Freinet, 1979; Hall et al., 2014; Hoyuelos, 2021; Malaguzzi, 1996; Montessori, 1982; Tonucci, 1995, 2001, 2012, 2017). As pointed out by Robinson (2009), we

all have skills and abilities that we can develop; aptitudes that need to be identified and further enhanced. This study shows that in the school environment, it is feasible to accept the diversity intrinsic to human nature. Moreover, by providing sociocultural learning, each individual can become aware of these aptitudes and develop in a holistic way (Karrera et al., 2021; Tonucci, 2001, 2012, 2017).

Given the value of inquisitiveness in childhood learning (Decroly, 2006; Freinet, 1979; Ortiz and Cervantes, 2016; Ouyeder et al., 2016; Tonucci, 1995), schools must foster diverse contexts and experiences in order to satisfy and preserve children's inherent curiosity (Tonucci, 2001, 2012). The sociocultural learning strategies presented in this research have proven to be useful educational experiences for achieving this goal, as all the projects are based on the children's own interests and concerns. Likewise, based on the ideas of Gómez et al. (2020), it can be concluded that by carrying out the creative projects presented in this study, the children interact with a diversity of people, including kids and teachers of all ages. They also engage with various experiences, materials and objects, through which they gain knowledge and understanding of the world around them. Likewise, the spectrum of learning available to each child is broadened, since any project can be approached in a multitude of ways. As one child puts it, *"You don't always have to do what the teacher tells you to do. You can think of something else and do it"* (A\_LIKTE\_IK22. "9"\_12). This contributes to diversity, since providing different experiences and pedagogical resources allows the learning to be adapted to the needs of each child, thereby ensuring that everyone understands and learns through this process (Tonucci, 2001, 2012, 2017).

On the path to well-being, it is clear that if children have more educational experiences, they will have increased opportunities to find those that suit them best, after which they will start to experiment and learn throughout the journey (Freinet, 1976, 1979; Tonucci, 2001, 2012, 2017). Thus, it can be affirmed that these experiences contribute to children's self-knowledge. Moreover, this has been pointed out by various leaders of the New School pedagogical movement, who have asserted that through experimentation children become aware of their aptitudes, thereby allowing them to develop more fully (Al, et al., 2012; Freinet, 1976, 1979; Karrera et al., 2021; Tonucci, 2001, 2012, 2017). At the same time, members of the educational community recognise the inherent potential of each individual, which values the uniqueness of their attributes and avoids comparisons in diverse situations.

Furthermore, it has been seen that motivation and effort in each project is enhanced by trusting children to make their own choices, as they

are allowed to choose from a variety of creative scenarios and a wide range of projects (Freinet, 1976, 1979; Tonucci, 2001, 2012, 2017). When children perform a task related to their interests, they experience a sense of inspiration and well-being. As a result, they see themselves as skilled and capable (Arguiñano, 2019), which enables them to feel confident in achieving their goals and completing the task at hand more easily and productively (Yañez, 2016).

Regarding the sources of inspiration used by children when carrying out a project, it has been concluded that the age diversity promoted by these schools has proven to be an invaluable pedagogical asset that significantly improves learning (Boix, 2011; Santos 2011; Tonucci, 2012). Children work and interact with teachers and kids of all ages, so in many cases their classmates become a source of creative stimulation. Consistent with the ideas of Freinet (1976, 1979) and Tonucci (2001, 2012, 2017, 2022), the authors of the present study conclude that this aspect also benefits the teacher's work, because when children help each other, the teacher does not have to constantly attend to the needs of every child.

With regard to the ideas brought from home, as the children have the opportunity to initiate projects based on reflections that arise outside school hours, the learning process is integrated into everyday life. Education is no longer seen as something separate, which can only be accessed during the school day with specific teaching materials (Freinet, 1979). Moreover, it can be seen how these experiences blur the boundaries between different disciplines, thereby promoting a broader view of reality and strengthening the ability to establish connections between diverse subjects (Gúzman et al., 2019). This dynamic can be associated with the global perspective on knowledge acquisition set forth by Decroly (2006), since the projects that arise from the link between various workshops honour the child's concept of the world, thereby helping children understand the interconnection between various subjects and knowledge. Therefore, the authors conclude that through the learning that arises from these experiences, children develop their imagination and creative thinking skills (Zuloeta, et al., 2021), which are essential qualities for facing the increasingly uncertain and globalised world in which we live (Pérez, 2019; Vergara, 2016).

With regard to the pedagogical value of the sociocultural learning that emerges from these scenarios, experimentation in these environments takes on great importance (Lave and Packer, 2011). In fact, by carrying out practical experiments, children acquire various skills independently, which allows them to contemplate and understand the different learning processes underlying each project. This is evidenced by the following

comment made by one child: “*Well, one of my classmates didn’t know how to do the project at first, but now she is learning how to do it, and she is learning something she couldn’t do before*” (IIP\_LIKTE\_IK22. “9”\_2). In this environment, children develop greater self-awareness and a deeper commitment to their learning. They understand that although projects do not always turn out according to the original idea, their persistence and effort allow them to improve their creations (Tonucci, 2001), while at the same time enabling them to learn from their mistakes and find new solutions: “*It’s all about doing things better. At first, you try different things, and then you start to be more serious and take better measurements, and by the third time, it’s planned really well*” (IIP\_OIKTE\_IK7 “8”\_4). The moment when the children present and explain their work is also very important. In fact, it has been confirmed that at this time, not only does the person who carried out the project learn, but their classmates learn as well, by listening and reflecting on the presentation. In short, this type of school encourages cooperation and mutual learning due to the interaction that takes place between people who live and work in the same space (Abós, 2015; Boix, 2011; Santos, 2011; Trueba, 1989; Tonucci, 1995, 2001, 2012).

If we consider that emotion directly influences learning by either increasing or decreasing our awareness, attention, and memory (Immordino-Yang & Damasio, 2007), this study has confirmed the vital need to create learning situations that are enjoyable for children. From this point of view, it is crucial to highlight the value of play in the learning process (Rabadan, 2021; Trueba, 1989; Tonucci, 2012, 2017). In these settings, the line between learning and play becomes blurred, which facilitates the acquisition of knowledge. The option of learning through real experiences and actively participating in the educational process contributes to learning that is not only effective, but also fun, engaging, and motivating (Karrera et al., 2021). Thus, although the children often see learning as play, they are also aware of the educational value behind these experiences. As one child points out, “*We learn, but in a different way*” (IIP\_OIKTE\_IK7. “12”\_1).

Therefore, in the current global context, there are numerous advantages to placing children at the centre of the learning process and making them the protagonists of school life (Rico-Gómez and Ponce, 2022). In this regard, sociocultural learning is a highly useful strategy. In fact, by having experiences that are consistent with their personalities, and by developing personal skills such as effort, perseverance, and organisation, children take an active role in building their own personal project (Pérez and Ahedo, 2020). As a final note, in the sublime words of Rabadan

(2021, p. 66), “Preparing for life means knowing where, and how, we want to position ourselves in the world. And without a doubt, our experiences are fundamental; finding, knowing, and strengthening our own talents should be paramount”.

### **Limitations**

As this study is based on small educational communities, the number and diversity of participants is limited, which may have influenced the results obtained. Likewise, regarding participant diversity, it would be useful to include people who have just arrived at these schools, or who do not fit the working model of these schools; in other words, teachers who are not accustomed to carrying out on this new type of teaching role. By adding this factor, it would be possible to understand the concerns of those who work in these schools, and to identify aspects that might make it difficult to adapt to these innovative, pedagogical methods. In doing so, it would be possible to propose solutions and strategies to address these problems. Finally, this research does not propose a single course of action for all schools. Instead, in line with Malaguzzi (1996), it recommends using the foundational concepts of this educational system as inspiration for other schools, which could adapt this approach to their own particular setting. In this way, the *Eskola Txikiak* method could be useful for schools with similar characteristics, such as those influenced by specific geographical locations (Hall et al., 2014), as well as those with very different circumstances (Tonucci, 1996).

### **Future lines of research**

Authors Abós and Boix (2017) point out that alternative assessment approaches can be developed in education, combining age diversity with hands-on teaching practices. In this regard, as it have been addressed the participating schools organise and develop their educational practices, spaces, and schedules in a completely different way, it would be interesting to delve deeper into how children’s learning at these schools is evaluated and monitored. During this research, although it has been observed that these schools use evaluation strategies that are innovative, useful, relevant, and interesting, this has not been explored in depth. Therefore, in future lines of research, it would be highly recommended to study how the issue of assessment is developed in these unique environments of sociocultural learning. Finally, in addition to adding value to what has been outlined in this study, these recommendations for future research could be instrumental in providing useful models for other schools in the coming years.

## Bibliographic references

### References

- Åberg-Bengtsson, L. (2009). The smaller the better? A review of research on small rural schools in Sweden. *International Journal of Educational Research*, 48(2), 100-108. <https://doi.org/10.1016/j.ijer.2009.02.007>
- Abós, P. (2015). El Modelo de Escuela Rural ¿Es un Modelo Transferible a Otro Tipo de Escuela?. *Educação & Realidade*, 40(3), 667-684. <https://doi.org/10.1590/2175-623645781>
- Abós, P. (2020). La escuela ubicada en territorios rurales: una escuela diferente, un reto pedagógico. *Aula*, 26(29), 41-52.
- Aguirregoitia-Güenechea, M. (2024). Antecedents de les escoles rurals a Euskadi (Eskola Txikiak) des de principis del segle XX fins a l'actualitat. *Temps d'Educació*, 66, 177-195. <http://dx.doi.org/10.1344/TempsE1ducacio2024.66.11>
- Arguiñano, A. (2019). *Hezkuntza berrikuntza: Antzuola Herri Eskolako bilakaera berrikuntza pedagogikoaren bidean* [Tesis doctoral UPV/EHU] ADDI. <https://addi.ehu.es/handle/10810/42035>
- Basasoro, M. (2021). *Ikasleen parte hartzea eskola demokratikoak eraikitze-ko giltzarri. Irakasleen begirada berritu batetik abiatutako Antzuola Herri Eskolako hezkuntza proiektua*. [Tesis doctoral, UPV/EHU]. ADDI.
- Bauman, Z. (2003). *Modernidad líquida*. Fondo de Cultura Económica.
- Brown, L.M, y Flaumenhaft, J. (2019). Student-empowered curricular change. *Phi Delta Kappan*, 100 (6), 13-19. <https://doi.org/10.1177/0031721719834023>
- Boix, R. (2011). ¿Qué queda de la escuela rural? Algunas reflexiones sobre la realidad pedagógica del aula multigrado. *Profesorado. Revista de Currículum y Formación de Profesorado*, 15(2), 13-23. <https://dialhnet.unirioja.es/servlet/articulo?codigo=4089448>
- Cascio, J. (27 de febrero 2023). *Facing the Age of Chaos*. Medium. <https://medium.com/@cascio/facing-the-age-of-chaos-b00687b1f51d>
- Comisión Europea (2016). *EntreComp: The Entrepreneurship Competence Framework*. <http://publications.jrc.ec.europa.eu/repository/bitstream/JRC101581/lfna27939enn.pdf>
- Decroly, O. (2006). *La función de globalización y la enseñanza y otros ensayos*. Biblioteca Nueva.
- De la Vega, L. (2020). Investigación sobre enseñanza y desarrollo profesional docente en escuelas rurales: una revisión. *REXE*, 20(43), 307-325.
- Díaz, F. (2012). *Las claves para educar en tiempos de crisis. Cómo transformar las dificultades en estímulo para una mejor educación*. Toromítico.

- Denzin, N. K. (2018). *The qualitative manifesto: A call to arms*. Routledge. <https://doi.org/10.4324/9780429449987>
- Eskola Txikiak (26 de noviembre, 2024). *Sare Izaera. Printzipio eta baloreak*. Eskola Txikiak. <https://www.eskolatxikiak.eus/>
- Eustat (2021). *Renta familiar media de la C. A. de Euskadi por ámbitos territoriales, según tipo de renta (euros). 2021*. Recuperado de la base de datos de Eustat. [https://www.eustat.eus/elementos/ele0005700/renta-familiar-media-de-la-ca-de-euskadi-por-ambitos-territoriales-segun-tipo-de-renta-euros/tbl0005791\\_c.html](https://www.eustat.eus/elementos/ele0005700/renta-familiar-media-de-la-ca-de-euskadi-por-ambitos-territoriales-segun-tipo-de-renta-euros/tbl0005791_c.html)
- Eustat (2022). *Estadística municipal de habitantes*. [https://www.eustat.eus/elementos/not0020444\\_c.pdf](https://www.eustat.eus/elementos/not0020444_c.pdf)
- Eustat (2024). *Udalmap. Departamento de Economía, Trabajo y Empleo*. <https://www.euskadi.eus/indicadores-municipales/web01-a2ogaeko/es/>
- Fielding, M. (2004). Transformative approaches to student voice: Theoretical underpinnings, recalcitrant realities. *British Educational Research Journal*, 2, 295–311.
- Freinet, C. (1979). *Las invariantes pedagógicas*. Morata (2º ed.).
- Freire, P. (1970). *Pedagogía del Oprimido*. Siglo XXI.
- Freire, P. (1997). *Pedagogía de la autonomía. Saberes necesarios para la práctica educativa*. Siglo XXI.
- Gobierno Vasco (2022). *Caracterización de las zonas rurales de Euskadi*. Departamento de Alimentación, Desarrollo Rural, Agricultura y Pesca. [https://www.euskadi.eus/contenidos/informacion/caracterizacion\\_zonas\\_rurales/es\\_def/adjuntos/Caracterizacion-zonas-rurales.pdf](https://www.euskadi.eus/contenidos/informacion/caracterizacion_zonas_rurales/es_def/adjuntos/Caracterizacion-zonas-rurales.pdf)
- Gómez, L. V., Valverde, K., y Villón, A. M. (2020). El aprendizaje dialógico en los procesos de interacción social del contexto educativo. *Dilemas contemporáneos: Educación, Política y Valores*. 1, 1-13. <https://doi.org/10.46377/dilemas.v35i1.2222>
- Guamán, V. J., y Espinoza, E. E. (2022). Aprendizaje basado en problemas para el proceso de enseñanza aprendizaje. *Revista Universidad y Sociedad*, 14(2), 124-131. <http://scielo.sld.cu/pdf/rus/v14n2/2218-3620-rus-14-02-124.pdf>
- Gúzman, I., Settati, A., y Marín, R. (2019). Transdisciplinariedad y la transversalidad: una experiencia para religar la práctica educativa. *Cultura Educación Y Sociedad*, 10(2), 73–84. <https://doi.org/10.17981/cultedusoc.10.2.2019.06>
- Hall, K., Horgan, M., y Ridgway, A. (2014). *Loris Malaguzzi and the Reggio Emilia Experience*. Bloomsbury.

- Hammersley, M. (1999). "Introduction." In M. Hammersley (Ed.), *Researching School Experience: Ethnographic Studies of Teaching and Learning*, (pp. 1-14). Routledge.
- Hammersley, M. (2017). What is ethnography? Can it survive? Should it?. *Ethnography and Education*. 10.1080/17457823.2017.129845.
- Immordino-Yang, M. H., y Damasio, A. (2007), We Feel, Therefore We Learn: The Relevance of Affective and Social Neuroscience to Education. *Mind, Brain, and Education*, 1, 3-10. <https://doi.org/10.1111/j.1751-228X.2007.00004.x>
- Karrera, I. (2008). El método cualitativo como herramienta crítica válida en la resolución de conflictos escolares [Tesis doctoral, UPV/EHU]. ADDI. <https://ekoizpen-zientifikoa.ehu.es/documentos/60cd611e-63d7a5247e8d17b7?lang=eu>
- Karrera, I., y Arguiñano, A. (2018). Escuelas inclusivas en el País Vasco: organización escolar y currículum emergentes y flexibles. perspectivas y desafíos. En H. Arancibia, P. Castillo y J. Saldaña (eds). *Innovación Educativa: perspectivas y desafíos*, (139-174). Universidad de Valparaíso.
- Karrera, I., Arguiñano, A., Basasoro, M., Castillo, P. (2020) Innovative pedagogical experiences at the Basque Country inclusive school. *British Journal of Educational Studies*, 68, 753 - 770. <https://doi.org/10.1080/00071005.2020.171903>
- Karrera, I., Basasoro, M. & Arguiñano, A. (2021) Pedagogical keys to inclusion and democracy in a Basque Country school: gaining knowledge from the educational project at Antzuola Herri Eskola. *Educational Review*, 75(3). <https://doi.org/10.1080/00131911.2021.1923460>
- Tonucci, F. (1995). *El niño y la ciencia. Con ojos de maestro*. Troquel.
- Tonucci, F. (1996). Un modelo para el cambio. *Cuadernos de Pedagogía*, 247, 48-51.
- Tonucci, F. (2001). ¿Cómo introducir la investigación escolar?. *Revista Investigación en la Escuela*, 43, 39-50. <http://hdl.handle.net/11441/60305>
- Tonucci, F. (2012). La escuela que queremos: "Aprender con besitos". *Revista Investigación en la Escuela*, 77, 13-24. <http://hdl.handle.net/11441/60090>
- Tonucci, F. (2017). A modo de introducción...: La diversidad como valor en una escuela que cambia. *Aula Abierta*, 46(2), 9-12 9-12. <https://reunido.uniovi.es/index.php/AA/article/view/11980>
- Malaguzzi, L. (1996). *Los cien lenguajes del niño*. Reggio Children. [https://www.educarjuntos.com.ar/wp-content/imagenes/los\\_cien\\_lenguajes\\_del\\_nino.pdf](https://www.educarjuntos.com.ar/wp-content/imagenes/los_cien_lenguajes_del_nino.pdf)

- Martínez, J., y Rogero, J. (2021). El entorno y la innovación educativa. *REICE. Revista Iberoamericana sobre Calidad, Eficacia y Cambio en Educación*, 19(4), 71-81. <https://doi.org/10.15366/reice2021.19.4.004>
- Matías, E., y Vigo, B. (2020). El valor del lugar en las relaciones de inclusión y exclusión en un colegio rural agrupado. Un estudio etnográfico. *Márgenes, Revista de Educación de la Universidad de Málaga*, 1(2), 90-106. <https://doi.org/10.24310/mgnmar.v1i2.8457>
- Maturana, G. A., y Garzón, C. (2015). La etnografía en el ámbito educativo: una alternativa metodológica de investigación al servicio docente. *Revista de Educación y Desarrollo Social*, 9 (2), 192-205. <https://dialnet.unirioja.es/servlet/articulo?codigo=5386223>
- McLaren, P. (2005). *La vida en las escuelas. Una introducción a la pedagogía crítica en los fundamentos de la educación*. Siglo XXI.
- Mena, L. (2018). *El desarrollo de la autonomía en la infancia. Programa de aplicación en el aula*. [Tesis doctoral, UPV/EHU]. ADDI. [https://addi.ehu.es/bitstream/handle/10810/36145/TESIS\\_MENA\\_LUCIA\\_INES.pdf](https://addi.ehu.es/bitstream/handle/10810/36145/TESIS_MENA_LUCIA_INES.pdf)
- Montessori, M. (1982). *El niño: el secreto de la infancia*. Diana.
- Lave, J., y Packer, M. (2011). Hacia una ontología social del aprendizaje. *Revista de Estudios Sociales*, 40, 12-22. <https://doi.org/10.7440/res40.2011.02>
- Little, A. W. (2001). Multigrade teaching: towards an international research and policy agenda. *International Journal of Educational Development*, 21(6), 481-497. [http://dx.doi.org/10.1016/S0738-0593\(01\)00011-6](http://dx.doi.org/10.1016/S0738-0593(01)00011-6)
- Oquendo, S. (2019). Estrategia para el desarrollo de la competencia investigativa en estudiantes de básica primaria. *Encuentros*, 17(2), 95-107. <https://www.redalyc.org/journal/4766/476661510009/html/>
- Ortiz, G., y Cervantes, M. L. (2016). La formación científica en los primeros años de escolaridad. *Panorama*, 9(17), 10-23. <https://doi.org/10.15765/pnrm.v9i17.788>
- Oudeyer, P.-Y., Gottlieb, J., y Lopes, M. (2016). Intrinsic motivation, curiosity, and learning: Theory and applications in educational technologies. *Prog Brain Res*, 229, 257-284. <https://pubmed.ncbi.nlm.nih.gov/27926442/>
- Pérez, Á. I. (2019). Ser docente en tiempos de incertidumbre y perplejidad. *Márgenes, Revista de Educación de la Universidad de Málaga*, 3-17, <https://revistas.uma.es/index.php/mgn/article/view/6497>
- Pérez, J., y Ahedo, J. (2020). La educación personalizada según García Hoz. *Revista complutense de educación*, 31(2), 153-161.
- Rabadán, K. (2021). *Bidean*. Txalaparta.

- Rico-Gómez, M. L., y Ponce, A. I. (2022). El docente del siglo XXI: perspectivas según el rol formativo y profesional. *Revista mexicana de investigación educativa*, 27(92), 77-101. <https://www.scielo.org.mx/pdf/rmie/v27n92/1405-6666-rmie-27-92-77.pdf>
- Robinson, K. (2009). *El elemento. Descubrir tu pasión lo cambia todo*. Grijalbo.
- Rodríguez, M. E. (2022). La etnografía crítica: un transmétodo rizomático transcomplejo en la transmodernidad. *Perspectivas Metodológicas*, 22, 21-21. <https://doi.org/10.18294/pm.2022.3992>
- Santos, L. (2011). Aulas multigrado y circulación de los saberes: especificidades didácticas de la escuela rural. *Revista de currículum y formación de profesorado*, 15(2), 71-91).
- Santos-Gelvasio, A. (2022). Inclusión y atención a la diversidad en el aula rural multigrado: un estudio de caso. *Revista Caribeña de Investigación Educativa*, 6(2), 15-3. <https://doi.org/10.32541/recie.2022.v6i2.pp15-34>
- Urbano-Carazo, M. I. (2022). Factores Determinantes de la Gestión de Conocimiento y su incidencia en el Desempeño Organizacional en entornos BANI: Una revisión teórica. *Revista Científica Anfibios*, 5(1), 11-25. <https://doi.org/10.37979/afb.2022v5n1.99>
- Unesco (2015). *Educación 2030. Declaración de Incheón y marco de acción para la realización del Objetivo de Desarrollo Sostenible n° 4*. [https://unesdoc.unesco.org/ark:/48223/pf0000245656\\_spa](https://unesdoc.unesco.org/ark:/48223/pf0000245656_spa)
- Unicef (2015). *Convención sobre los derechos del niño*. <https://www.unicef.es/publicacion/convencion-sobre-los-derechos-del-nino>
- Vallejos, R., y Finol de Franco, M. (2009). La triangulación como procedimiento de análisis para investigaciones educativas. *REDHECS: Revista electrónica de Humanidades, Educación y Comunicación Social*. 4(7). 117-133. <https://dialnet.unirioja.es/servlet/articulo?codigo=3063110>
- Vergara, J. J. (con Pérez Gómez, A. I.). (2016). *Aprendo porque quiero*. Ediciones SM.
- Vigo-Arrazola, B., y Beach, D. (2022). Las escuelas rurales ante las políticas de mercado: Un análisis meta-etnográfico. *Archivos Analíticos de Políticas Educativas*, 30(173). <https://doi.org/10.14507/epaa.30.6962>
- Vigo, B., y Soriano, J. (2015). Family involvement in creative teaching practices for all in small rural schools. *Ethnography and Education*, 10(3), 325-339. <https://doi.org/10.1080/17457823.2015.1050044>
- Vygotsky, L. (1978). *El desarrollo de los procesos psicológicos superiores*. Austral

- Yañez, P. (2016). El proceso de aprendizaje: fases y elementos fundamentales. *Revista San Gregorio*, 11, 70-81. <https://dialnet.unirioja.es/servlet/articulo?codigo=5585727>
- Zuloeta, E. J., Rojas, N. R., y Caramutti, V. (2021). La creatividad en estudiantes educación inicial: una revisión bibliográfica. *Revista Conrado*, 17(82), 260-267. <https://conrado.ucf.edu.cu/index.php/conrado/article/view/1957>

# Comparing Visual Thinking and Flipped Classroom in International Relations Teaching

## Comparación entre el pensamiento visual y el aula invertida en la enseñanza de las relaciones internacionales

<https://doi.org/10.4438/1988-592X-RE-2025-412-739>

**Andrea Betti**

<https://orcid.org/0000-0002-4629-4572>

*Universidad Pontificia Comillas*

**Pablo Biderbost**

<https://orcid.org/0000-0002-4086-3658>

*Universidad de Salamanca*

**Esther Vaquero**

<https://orcid.org/0000-0001-5196-3462>

*Universidad Pontificia Comillas*

### Abstract

Developments in teaching innovation have recently seen a great push towards formats that can increase and improve students' active learning. A wide array of teaching formats based on this concept has been implemented across various university disciplines, from natural to social sciences. Most empirical analyses have so far focused on comparing the effects of traditional and active teaching formats on students' academic performance. This has improved scientific knowledge about the benefits and drawbacks of several teaching formats. Nevertheless, contrasting traditional lecturing formats with activity-based approaches to active learning risks yielding simplistic and unhelpful dichotomies between "traditional" and "active" instruction. In this article, we follow a different strategy. By using non-parametric tests, we compare two different teaching formats based on two different active learning strategies, such as flipped classroom and visual thinking. Students' performances in the two groups, composed of undergrad-

uate students enrolled in a dual degree in International Relations and Global Communication, are compared in terms of academic results, that is their grades, and a set of soft skills, such as learning perception, self-efficacy, and teamwork. The study detects a relevant effect in terms of grades, with students through the flipped format outperforming students taught through the visual thinking format. On the other hand, the study detects only one relevant effect in the soft skill of learning perception, with students in the flipped format scoring higher than students in the visual thinking format. While this does not necessarily indicate that FC formats are more effective than VT formats, this difference in students' grades and learning perception can signal that different teaching formats can be differently but equally useful depending on the teaching goal or need.

**Keywords:** Active learning, International Relations, Visual Learning, Films, Documentaries, Flipped Classroom, Higher Education.

### Resumen

Los avances en la innovación docente han experimentado recientemente un notable impulso hacia formatos capaces de incrementar y mejorar el aprendizaje activo del estudiantado. Una amplia variedad de formatos de enseñanza basados en este concepto se ha implementado en diversas disciplinas universitarias, desde las ciencias naturales hasta las ciencias sociales. La mayoría de los análisis empíricos se han centrado hasta ahora en comparar los efectos de los formatos de enseñanza tradicionales y activos sobre el rendimiento académico del alumnado. Esto ha permitido ampliar el conocimiento científico sobre los beneficios y limitaciones de los distintos formatos de enseñanza.

No obstante, contrastar los formatos tradicionales de clase magistral con enfoques basados en actividades orientadas al aprendizaje activo conlleva el riesgo de generar dicotomías simplistas y poco útiles entre la instrucción «tradicional» y la «activa». En este artículo seguimos una estrategia distinta. Mediante el uso de pruebas no paramétricas, comparamos dos formatos de enseñanza basados en dos estrategias diferentes de aprendizaje activo: el aula invertida (*flipped classroom*) y el pensamiento visual (*visual thinking*).

Se comparan los resultados de los dos grupos, integrados por estudiantes de grado matriculados en un doble programa en Relaciones Internacionales y Comunicación Global, tanto en términos de rendimiento académico —es decir, sus calificaciones— como en un conjunto de habilidades blandas, tales como la percepción del aprendizaje, la autoeficacia y el trabajo en equipo. El estudio detecta un efecto relevante en las calificaciones, ya que el estudiantado que siguió el formato de aula invertida obtiene resultados superiores a quienes fueron enseñados mediante el formato de pensamiento visual. Por otro lado, el estudio identifica únicamente un efecto significativo en la habilidad blanda de percepción del aprendizaje, con puntuaciones más altas entre el estudiantado del formato de aula invertida en comparación con el del formato de pensamiento visual.

Si bien esto no indica necesariamente que los formatos de aula invertida sean más eficaces que los de pensamiento visual, la diferencia observada en las califi-

caciones y en la percepción del aprendizaje puede señalar que distintos formatos de enseñanza pueden resultar, de manera diversa pero igualmente útil, adecuados según el objetivo o la necesidad docente.

*Palabras clave:* Aprendizaje activo, Relaciones Internacionales, Pensamiento Visual, Películas, Documentales, Clase Invertida, Educación superior.

## Introduction

In the last several decades, university education has seen a significant push towards teaching formats that can enhance students' "active learning." Instead of traditional formats, in which the professor gives frontal lectures and students take notes that they will use after class to try memorizing the class contents (Bligh 1998), teaching and learning are increasingly required to become collaborative, "problem-based", digitally-rich, and capable of fostering critical thinking and teamwork (Collins and Halverson, 2010). This is supposed to improve not only students' lower-order- skills, such as the "transfer" and "memorization" of "basic knowledge," (Omelicheva and Avdeyeva, 2008) but also their "higher-order skills", such as analysing, synthesizing, and evaluating (Anderson and Krathwohl, 2001).

Many teachers have answered this call by implementing active learning strategies in their classes, in both social and natural sciences (Strelan et al., 2020; Galindo-Domínguez, 2021). As a result, it has become more common to design studies aimed at empirically understanding and assessing the effectiveness of such strategies on students' learning (Hussain et al., 2023; Naing et al., 2023). First examples of these studies tended to compare traditional and active teaching formats (Cheng et al., 2018; Talbert and Bergmann, 2017; Bergmann and Sams, 2012;). These comparisons have significantly augmented the collective knowledge about their possible benefits and drawbacks. Nevertheless, merely opposing traditional and active teaching formats carries the risk of oversimplifying an important debate, exaggerating the benefits of active learning strategies, and leading to the insufficient conclusion that active is better than traditional learning, often disregarded as passive learning.

For this reason, following an increasing trend (Betti et al., 2022; Jensen et al., 2015; Lai & Hwang, 2016), this article investigates the effectiveness of active learning in the field of International Relations (IR) teaching, by comparing two teaching formats both based on active learning strategies. Two groups of students were taught the same IR class. One was taught through a Flipped Classroom (FC) format, while the other through a

Visual Thinking (VT) format based on films and documentaries. Their performances were compared in terms of grades and a set of soft skills, such as teamwork, self-efficacy, and learning perception.

On the one hand, our results identified a relevant effect in terms of grades, with students taught through the FC format outperforming students taught through the VT format. On the other hand, the only soft skill in which the study detected a relevant effect was learning perception, with FC students scoring higher than VT students.

### **Flipped Classroom and Visual Thinking**

The FC is based on the idea that “that which is traditionally done in class is now done at home, and that which is traditionally done as homework is now completed in class” (Bergmann and Sams, 2012, p. 13). Traditional in-class lectures, based on the professor’s frontal teaching of contents, are replaced by materials, usually pre-recorded video-classes, that students are required to study before class. Class time is, thus, spent entirely on solving students’ doubts about the contents and administering practical exercises, such as problem-based discussions, presentations, or group work, designed to apply what students have learned through the video-lectures. “With content provision moved outside the class”, class time can be more efficiently used to perform activities that promise to foster students’...problem solving and teamwork abilities (Jenkins, 2015, p. 607). The overall goal is to turn the classroom “into a dynamic, interactive learning environment”, in which teachers assist students in their own learning process, fostering a more creative engagement with the subject (Berge and Nederveld, 2015, p. 163).

VT was originally defined as “an inquiry-based method”, aimed at developing “students’ thinking skills by looking at and discussing art” (Tisham et al., 1999, p. 1). The main idea is that through the observation of images, learning can become more interactive and entertaining, fostering students’ capacity to understand abstract and difficult concepts through a visual experience. Far from being limited to artistic images, VT teaching formats can be based on a wide array of visual materials, such as films, documentaries (Van Munster and Silvest, 2015; Heck, 2017), comics (Schmid, 2020), or videogames (Valeriano & Habel, 2016), integrating traditional teaching instruments, such as readings and frontal lectures.

### **Previous studies**

Studies that measured the effects of active learning techniques on students’ learning usually operationalize academic achievement as grades

obtained in the assignments and exams. However, several studies also aimed to assess the effects of active learning on soft skills, understood as “interpersonal qualities...and personal attributes that one possesses” (Robles, 2012, p. 453). Teamwork is one of the most studied soft skills. It is usually referred to as the capacity to “direct and coordinate the activities of other team members”, “anticipate other team members’ needs”, and “apply” and “adjust strategies” (Salas et al., 2005, pp. 558-559). Learning perception is defined as the “cognitive effort required during learning” (Deslauriers et al., 2019: 19251), usually operationalized as students’ satisfaction or lack of satisfaction with teaching formats. Finally, self-efficacy is often defined as “students’ beliefs about whether they are able to show certain learning behaviour” (Baars and Wijnia, 2018, p. 127) based on their convictions about their own capacity to “produce designated levels of performance” (Bandura, 1994, p. 2).

Unlike other disciplines, especially natural sciences (Naing et al., 2023), there has been relatively less interest for active learning techniques in Political Science (PS) and International Relations (IR), especially in terms of empirical analyses on the effectiveness of such techniques on students’ learning. As to FC, there are some PS and IR studies that compared traditional teaching formats, centred around the professor and the exposition of contents during class time, with inverted formats, based on the substitution of traditional in-class teaching with pre-recorded video-lectures to be watched before class, so that class time can be devoted to exercises and applications. Some detected an improvement in students’ grades (Touchton, 2015), while others did not find any significant difference between a traditional and a flipped format in terms of students’ academic achievement (Lambach et al., 2017). Similar diverging results can be found in those studies that compared traditional and flipped formats in terms of students’ soft skills. While some identified an improvement in the teamwork skills of students taught through a FC (Cit., p. 563), others detected a worsening of students’ teamwork abilities in a FC (Jenkins, 2015, p. 610). As to learning perception, available studies again diverge. While some observed that students fundamentally split between those who prefer a flipped format and those who prefer a traditional one (Lambach et al., 2017), others detected a students’ preference for “mixed class sessions”, rather than “having all flipped-class sessions” (Jenkins, 2015, p. 610). Finally, to our knowledge, there are no PS or IR studies available on the relation between FC and students’ self-efficacy.

Nevertheless, following an impulse coming from other disciplines, several PS and IR studies found it more useful to test the effectiveness of

the FC format against “a control model that uses active learning” (Jensen et al., 2015, p. 11). This reduces the risk of examining teaching formats that are too different to be compared and increases the opportunities to “parse out the effects and pinpoint a specific causal factor” (Cit., p. 2). Some identified an improvement in students’ grades who were taught through a FC, in comparison to students that received online synchronous lectures (Whitman Cobb, 2016). In a similar way, some found that students’ grades can improve with a FC, if it is designed in combination with other forms of remote teaching, both offline and online (Van der Zwan and Afonso, 2019). In terms of soft skills, one of the very few studies that we could find that compared the flipped format with other active learning strategies, detected a simultaneous improvement in academic performance and learning perception, with students expressing satisfaction with the FC (Whitman Cobb, 2016).

Unlike other disciplines (Yen et al., 2018; Lin et al., 2019), in which there is a clearer consensus in favour of the benefits of the flipped teaching format on both students’ academic achievement and soft skills, there is less observed agreement in the fields of PS and IR. Different studies have reached dissimilar and partially conflicting conclusions. This adds to the impression of a discipline in which interest and curiosity for teaching innovation are still marred with prudence and scepticism. The first goal of this study is, thus, to investigate whether a teaching format based on FC can improve students’ achievement in terms of grades and soft skills.

These tendencies are even more visible in the study of the effectiveness of VT teaching formats on Social Science subjects. In disciplines, such as History, the use of visual materials, including films, has been common (Rose, 2016). Even though most available studies have been largely anecdotal and not based on systematic empirical analyses, some historians found that using films in History classes positively impacted on students’ understanding of complex situations (Wagner, 2018). However, those same historians also reported on the difficulty in using visual materials, such as films, in a History class, either due to the difficulty for students to discern facts from visual and artistic interpretations (Stoddard, 2012) or due to a lack of visual media analytical skills on the teachers’ part (Wagner, 2018).

Even more “weariness” (Dean, 2019, p. 257) can be observed in the use of VT formats in PS and IR. For some, the scepticism about the “feeling and affective dynamics that underpin everyday forms of political participation and engagement” would foster a “squeamishness” of these disciplines about using visual materials (Cit.). Possibly because of this,

not so many PS and IR teachers have been willing to use visual materials in their classes, although this tendency seems to be changing. Recent studies, for example, detected a students' preference for classes taught through image-rich slides rather than text-heavy slides. Students tend to feel more engaged with the class contents. Moreover, they would associate a larger cognitive overload with text-heavy slides, as they would generate a constant dilemma between reading the slides and listening to the teacher. Images would, thus, improve their satisfaction and learning perception about the class. However, the same study also warns about the abuse of images during the explanation of contents, not only because their content can be at times controversial but also because they are not always sufficient to communicate complex concepts or data. The risk of oversimplification could reduce the students' capacity to critically read and analyze information presented in specialized books and articles (Roberts, 2017).

Several other studies have compared traditional and VT formats based on the use of films. Most of them have identified positive effects on students' soft skills, such as creative thinking (Valeriano, 2013; Kuzma & Haney, 2001; Weber, 2001) and analytical skills (Lieberfeld, 2007). Films can increase students' capacity for analysis and help them connect and contextualize historical facts and characters (Gokcek & Howard, 2013). In a different way, others have warned about the possible downsides of using films for teaching purposes, for example when they oversimplify reality through inaccurate representations of facts (Kuzma & Haney, 2002, p. 93) or when they contain stereotypical and biased representations of specific cultures (Giglio, 2002). Even though most of these studies are not based on systematic empirical analyses, some have also tried to observe the possible effects of films on students' academic achievement and learning perception. Some identified benefits in using films, such as an improvement in students' capacity to grasp and retain concepts (Sunderland et al., 2009), thanks to the fact that films would help "make abstract theories and concepts more understandable" (Gokcek & Howard, 2013, p. 441; Kiasatpour, 1999), entertaining (Kuzma & Haney, 2001, p. 47), or engaging (Gokcek & Howard, 2013; Iretzberger, 2021).

Nevertheless, other observations noted that students' learning perception might not necessarily benefit from using films. At times, films can increase the feeling of a bigger workload (Kiasatpour, 1999, p. 85) or even confusion, especially because students can find it hard to distinguish between facts and fiction (Lieberfeld, 2007, p. 573). Even when films can occasionally improve students' understanding of complex processes, such as the ones related to decisionmaking, their comprehension of "the

theoretical concepts and how they could be applied” might not be enhanced (Inoue & Krain, 2014, p. 20). One of the few empirical analyses based on quantitative methods identified contradictory results. Films can increase students’ comprehension, engagement, and grades, but they can equally treat issues in a simplified way for the sake of artistic representation, thus reducing students’ capacity to see a clear relation between academic content and fictional representations (Swimelar, 2013, p. 24). All these analyses were based on comparisons between traditional and VT teaching formats. The only exception that we could find was a study that compared two different teaching formats both built around active learning strategies, one based on films and the other based on games, detecting that the latter had a more beneficial effect on students’ grades and learning perception than the former (Brandle, 2020).

In conclusion, unlike other subjects in which one can find a clearer consensus on the benefits of using films for students’ achievement and satisfaction, studies from PS and IR offer a wider range of results. These divergences could be partially motivated by the still relative lack of studies. Few of them are based on systematic and empirical analyses of the effects of using films on students’ grades and soft skills. Moreover, most of them are based on anecdotal comparisons between traditional lectures and the use of films. This leads to the need to test the possible benefits and drawbacks of using VT formats against other active learning formats in a more systematic way. PS and IR have been less prone to using VT formats in the classroom, making the study of their possible potential for teaching purposes still relatively underexplored. The second goal of this study is, thus, to investigate whether a teaching format based on VT can improve students’ achievement in terms of grades and soft skills.

For this reason, in this study, we propose to analyse two different teaching formats both based on active learning strategies, one designed in a flipped way and the other around the use of visual materials, such as films and documentaries. This aims to add empirical data on the effectiveness of two important active learning strategies for students’ learning, contributing to a larger debate, still partially underdeveloped in PS and IR, on what teaching formats can be implemented to improve students’ performance and experience.

## **Research Design and Methods**

Our goal is to discover whether teaching strategies based on the FC and VT can lead to an improvement in the students’ academic performance in terms of both hard skills, that is their grades, and soft skills, such as teamwork, self-efficacy, and learning perception.

This quasi-experimental study was administered in two different sections of an undergraduate second-year mandatory core class, called Comparative Political Systems, as part of a Dual Degree in International Relations and Global Communication. One group was administered half of the class – 7 out of 14 weeks of class – through an FC methodology, while the other group was administered the same half of class through a VT methodology. Both teaching formats were based on the same contents, taught by the same professor, and administered for the second seven weeks of the class.

For the FC format, the professor first recorded a set of video-lectures, with the support of the Teaching Innovation Unit at the University where the study took place. Videos included explanations of the contents of the class and specific sections highlighting the key concepts. Students were required to watch them carefully before coming to class. Based on those video-lectures, students were, then, required to perform a variety of activities in the classroom, under the supervision of the professor. For example, the professor asked questions to stimulate individual or group reflection about how to define and link specific political concepts and categories. In other cases, students were required to read and then make comments about specific readings selected by the professor and related to the video-lectures. Finally, students were asked to work individually or in group to present to the professor and the rest of their classmates the findings of small research tasks to be previously performed at home. The goal of these activities was to promote individual and collective reflection on complex issues, foster cooperative learning, and increase the possibility for the professor to provide instant and public feedback.

For the VT format, students had to watch a variety of visual contents before class time, such as documentaries and films, selected by the professor and related to some of the key concepts of the class. Visual contents were assigned on a weekly basis. After each assignment, students were required to go to class and participate in activities designed by the professor and aimed at reinforcing the knowledge and understanding of the main concepts of the class. These activities included class debates and short presentations based on small research tasks about the contents of the assigned visual materials. Moreover, by working in group outside the classroom, students were asked to produce brief visual contents, such as recorded videos or selections of photographs, through which they delved into specific concepts and phenomena related to the contents of the class. This way, students did not merely act as passive consumers of visual content. Films and documentaries became instruments to be analysed, evaluated, and used as starting points for a deeper reflection based

on small research assignments. This allowed them to experiment with skills, such as analyzing, comparing, contrasting, and synthesizing. In addition, through the assignments that involved their own production of visual contents, students could train themselves to visually communicate the results and findings of their learning process. Thus, visual contents were used both as teaching resources and students' own contributions.

The first seven weeks of the class were administered through a semi-traditional format based on a combination of frontal lecturing and active learning techniques, such as presentations, teamwork activities, and debates. Two hours per week were dedicated to frontal lecturing, while the other two were dedicated to students performing a variety of class activities based on their active participation, such as presentations of small research tasks, debates, or teamwork projects.

As Table 1 shows, we could not collect the data for all the students enrolled in the two groups, as some students did not answer any of the surveys.

**Table 1.** Description of the sample

	<b>Group 1: Flipped Classroom Methodology</b>	<b>Group 2: Visual Thinking Methodology</b>
Students enrolled	45	62
Sample	31 (69 %)	42 (68 %)

Both groups were analysed to find out whether there were significant differences in terms of their academic achievement prior to the start of the class that was used for the study. After observing that the series are normal, the t test of independent samples shows that there were no significant differences either in the average grades or in the variance among the two groups (See table 2). This allowed to detect that there were no significant differences between the two groups in terms of prior academic achievement. No significant difference was found either in terms of students' socio-economic level. Finally, all the students are Spanish nationals and were trained in the traditional Spanish educational system, which is based on traditional frontal lecturing. This way, we could establish that the two groups were comparable.

To answer the two following research questions, this study employs a methodology based on comparing the average results of each group of students, for both hard and soft skills. The goal is to discover whether the teaching format influences results. As explained below, the type of test used depends on whether the data series are normal or not.

## Results

### Research question 1 (RQ1): Academic achievement (hard skills)

Does a teaching strategy based on FC or VT improve students' academic performance measured in terms of grades?

This is the first goal of our study. To evaluate the impact of FC and VT in the students' academic results measured in terms of grades, we considered:

- the average of students' academic transcripts for each group.
- the academic results obtained in an exam that students took at the end of the first seven weeks administered through a semi-traditional teaching format.
- the results obtained in an exam that was administered at the end of the two teaching formats administered, respectively, through FC and VT.
- and the results obtained in the final exam administered at the very end of the class. This final exam was mostly based on questions aimed at testing students' capacity to retain the contents of the class.<sup>1</sup>

Due to the fact that some of these series were not normal<sup>2</sup>, we performed the non-parametric Krusal-Wallis test to compare the academic results obtained by the students of one group with those obtained by the students of the other group. Moreover, to analyze the evolution of students' grades throughout the duration of the study, we performed a Wilcoxon test. Results appear in Table 2.

<sup>1</sup> Data range from 0.00, being the lowest grade, to 10.00, being the highest grade. The minimum grade to pass an exam is 5.00.

<sup>2</sup> Normality test.

	Normality Test Hard Skills					
	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	p-value	Statistic	df	p-value
No Visual Thinking Midterm test	0,160	59	0,001**	0,880	59	0,000**
Visual Thinking Midterm test	0,117	59	0,045*	0,947	59	0,013
Visual Thinking Group Final Grade	0,112	59	0,064	0,960	59	0,051
Visual Thinking Group academic background	0,096	59	0,2	0,971	59	0,166
Flipped Classroom Midterm test	0,180	31	0,012	0,932	31	0,048*
Flipped Classroom Group academic background	0,096	31	0,2	0,969	31	0,496
Flipped Classroom Group Final Grade	0,108	31	0,2	0,942	31	0,094
No Flipped Classroom Midterm test	0,148	31	0,082	0,956	31	0,235

**Table 2.** Comparison of academic achievement in the FC and VT teaching formats in terms of hard skills (grades)

(1)	(2)	p-value	effect size
Visual Thinking Group Academic Background	Flipped Classroom Group Academic Background	0,072	0,328
No Flipped Classroom Mid-term Test	No Visual Thinking Mid-Term test	0,315	0,106
Flipped Classroom Mid-Term test	Visual Thinking Mid-Term test	0,000**	0,576
Flipped Classroom Mid-Term test	Flipped Classroom Group Final Grade	0,695	0,070
Visual Thinking Mid-Term test	Visual Thinking Group Final Grade	0,000**	0,784
Visual Thinking Group Final Grade	Flipped Classroom Group Final Grade	0,013*	0,262
Flipped Classroom Group Academic Background	Flipped Classroom Group Final Grade	0,005	0,505
Visual Thinking Group Academic Background	Visual Thinking Group Final Grade	0,361	0,119

As table 2 shows, our analysis detects a significant improvement in students' academic achievement measured in terms of grades when receiving the FC modality. Students' improvement is the result of the flipped teaching format used. This is demonstrated by the fact that their grades in the end of the first section exam (where the two groups of students received the class in the same semi-traditional format) are statistically similar, with a small effect size. This means that the improvement in their grades, when receiving the class in the flipped format, can be attributed to the teaching format.

Along these lines, our analysis detects a statistically significant difference in the exam grades that took place at the end of the two teaching formats and that were administered, respectively, through FC and VT. In this exam, students were evaluated for the very same contents, with the only difference being that such contents had been explained through FC to one group and VT to the other group. In this case too, students who received the class in the flipped format obtained better grades than students who received the class in the visual format. As can be observed, the effect size is relatively large. This means that not only is there a difference in grades, but also that the difference is significant.

Finally, this improvement can be observed also when analyzing the final grades of the class, which is the average of all the grades that they obtained throughout all the sections of the class. However, as can be observed, the improvement is relatively small.

Based on these results, it can be affirmed that the FC format was more effective than the VT format at improving students' grades.

**Research question 2 (RQ2): Soft skills**

Does a teaching strategy based on FC or VT improve students' academic performance in terms of soft skills?

This is the second goal of our study. To evaluate the impact of FC and VT on students' soft skills, operationalised in terms of a survey based on ten variables related to self-efficacy, six variables related to teamwork, and one variable related to learning perception (see table 3), we performed two tests in each group, one at the beginning and one at the end of the class to find out whether there was any change in each of the variables considered.

**Table 3.** Soft skills survey<sup>3</sup>

	<b>Self-efficacy</b>	<b>Completely disagree</b>	<b>Completely agree</b>
X1	I think I am going to get some excellent grades this year.		
X2	If I make an effort, I think I have enough capacity to achieve a good academic record.		
X3	I believe that I am able to understand even the most difficult topics in this course.		
X4	I think I have enough capacity to understand a subject, quickly and well.		
X5	I think I can pass the courses quite easily and even get good grades.		
X6	Although teachers are demanding and strict, I have great confidence in my own academic ability.		
X7	I think that I am prepared and well qualified to achieve academic success.		
X8	When they ask me to do projects or homework, I am sure that I will do them well.		
X9	I work effectively in any team, no matter who the teammates are.		

<sup>3</sup> The Institute of Science Education at the University where the study took place provided the questionnaire. It was based on established questions previously used by this Institute. A pretest was done with five students to validate its wording and comprehension.

<b>Self-efficacy</b>		<b>Completely disagree</b>	<b>Completely agree</b>
X10	Considering the difficulty of the degree, what I am learning, and my own abilities, I think I'll be fine when I finish (the degree).		
<b>Teamwork<sup>4</sup></b>		<b>Very negative perception</b>	<b>Very positive perception</b>
X11	Participation in teamwork sharing information, knowledge, and experiences.		
X12	Acceptance and compliance with the rules agreed upon in the group (deadlines, parts of the work, format, etc.).		
X13	Action to face team conflicts in this subject.		
X14	Commitment to the management and operation of the equipment		
X15	Management of meetings effectively.		
X16	Communication and cohesion within the group.		
<b>Learning perception</b>		<b>Few</b>	<b>A lot</b>
X17	Regardless of your results in the exams, think how much you will learn in this subject about Comparative Political Systems.		

After having observed this change, defined as “C” for each soft skill, we analysed whether such a change was the same in case of using the FC or VT. This way, we compared the change observed in each of the soft skills in the teaching format based on the FC with the changes observed in each of the soft skills in the teaching format based on VT. Due to the fact that not all series were normal<sup>5</sup>, in order to homogenise the analysis, we used non-parametric techniques of analyses that are less restrictive,

<sup>4</sup> Before the study, students had already attended a class of a mandatory diploma in “Development of Personal, Communication, and Professional Abilities, offered by the University where the study took place. This class was called “Teamwork.” This class provided them with an initial perception of their teamwork skill. Such starting perception was measured at the beginning of the study. We presupposed that this perception was superior to zero, due to the students’ previous attendance of the “Teamwork” class. Upon finalising the study, we again measured this perception, by considering whether it had improved or worsened compared to the initial one.

<sup>5</sup> Normality Test, whereas V indicates that change is computed within the VT format and F is computed within the FC format.

such as the median test, the Mann-Whitney U test, and the Kruskal-Wallis test. Results appear in table 4.

Normality Test Soft Skills						
	Kolmogorov-Smirnova			Shapiro-Wilk		
	Statistic	df	p-value	Statistic	df	p-value
C1V	0,323	31	0,000**	0,799	31	0,000**
C2V	0,257	31	0,000**	0,841	31	0,000**
C3V	0,203	31	0,002**	0,922	31	0,027**
C4V	0,303	31	0,000**	0,834	31	0,000**
C5V	0,302	31	0,000**	0,852	31	0,001**
C6V	0,286	31	0,000**	0,847	31	0,000**
C7V	0,284	31	0,000**	0,847	31	0,000**
C8V	0,277	31	0,000**	0,849	31	0,000**
C9V	0,264	31	0,000**	0,857	31	0,001**
C10V	0,228	31	0,000**	0,900	31	0,007**
C11V	0,238	31	0,000**	0,852	31	0,001**
C12V	0,281	31	0,000**	0,875	31	0,002**
C13V	0,321	31	0,000**	0,829	31	0,000**
C14V	0,291	31	0,000**	0,835	31	0,000**
C15V	0,308	31	0,000**	0,824	31	0,000**
C16V	0,259	31	0,000**	0,855	31	0,001**
C17V	0,221	31	0,001**	0,879	31	0,002**
C1F	0,370	31	0,000**	0,731	31	0,000**
C2F	0,323	31	0,000**	0,827	31	0,000**
C3F	0,297	31	0,000**	0,848	31	0,000**
C4F	0,371	31	0,000**	0,718	31	0,000**
C5F	0,266	31	0,000**	0,846	31	0,000**
C6F	0,227	31	0,000**	0,870	31	0,001**
C7F	0,261	31	0,000**	0,886	31	0,003**
C8F	0,184	31	0,009**	0,932	31	0,051
C9F	0,339	31	0,000**	0,742	31	0,000**
C10F	0,292	31	0,000**	0,880	31	0,002**
C11F	0,328	31	0,000**	0,765	31	0,000**
C12F	0,394	31	0,000**	0,729	31	0,000**
C13F	0,210	31	0,001**	0,937	31	0,067
C14F	0,310	31	0,000**	0,834	31	0,000**
C15F	0,278	31	0,000**	0,836	31	0,000**
C16F	0,258	31	0,000**	0,854	31	0,001**
C17F	0,268	31	0,000**	0,859	31	0,001**

**Table 4.** Comparison of changes in the soft skills between VT and FC teaching formats

Variable	Median Test	Mann-Whitney U Test	Kruskal-Wallis Test
C1	0,169	0,174	0,174
C2	0,538	0,329	0,329
C3	0,057	0,073	0,073
C4	0,849	0,618	0,618
C5	0,311	0,244	0,244
C6	0,653	0,735	0,735
C7	0,834	0,476	0,476
C8	0,744	0,866	0,866
C9	0,815	0,995	0,995
C10	0,412	0,249	0,249
C11	0,255	0,347	0,347
C12	0,466	0,861	0,861
C13	0,417	0,101	0,101
C14	0,849	0,981	0,981
C15	0,972	0,428	0,428
C16	0,925	0,699	0,699
C17	<,001***	0,002***	0,002***

As Table 4 shows, unlike with grades, our analysis does not detect any improvement in terms of students' soft skills either in the section administered through FC or in the one administered through VT. No statistically significant differences can be observed in either format. The only exception is the difference that can be observed in terms of learning perception, understood as how students expect that the teaching format can affect their grades. Those who received the class through the flipped format tend to perceive that they are going to obtain better grades, as compared to the perceptions of those who received the class in the visual format.

## Discussion

Our study corroborates previous ones that detected a positive impact of FC on both students' grades and learning perception (Whitman Cobb, 2016; Van der Zwan and Afonso, 2019). Moreover, by detecting a lack of a similar improvement among students who were administered the class

through VT, our study goes in a similar direction to those that expressed prudence about tailoring entire classes or sections of them around visual materials (Lieberfeld, 2007, p. 573; Inoue & Krain, 2014, p. 20). The FC format, with emphasis on the possibility to watch video-lectures at home and discuss doubts and unclear aspects of the class in the classroom with both the teacher and the classmates, proved helpful for students to consolidate their understanding of complex concepts and come up with ideas on how to apply them. The same cannot be said for the format based on VT.

Nevertheless, far from indicating a lack of usefulness in VT formats, this difference in students' grades and learning perception can indicate that dissimilar active teaching formats can be differently useful depending on the teaching goal or need. Several reasons can explain why the FC format proved more effective than VT for students' grades.

First, even though, as a teaching technique, FC is certainly more recent (Bergmann and Sams, 2012) than classroom projections of films and documentaries, that have been long used in various social sciences (Rose, 2016), the former rests on a larger tradition in terms of study about its design, implementation, and empirical assessment. This has produced large literature that aims to empirically test its strengths and limitations in terms of improving students' performance with both grades and soft skills. Even though it probably has a longer history, using films and documentaries has not led yet to the same production in terms of empirical analyses of its effectiveness for students' learning.

Second, the FC can be implemented only through a comprehensive re-organization of class time and space. This is due to its characteristics, based on removing frontal lecturing from the classroom to substitute it with video-lectures to be watched outside of class and with exercises and applications to be performed in the classroom under the teacher's supervision (Jenkins 2015, 607; Berge and Nederveld 2015, 163). This entails preparing contents that need to be put at students' disposal before class time and designing class exercises and activities based on the contents that students are supposed to have learned with the video-lectures. This way, class time can be valuably used to solve students' doubts and help them master applications of what they learned. In this regard, the FC can make use of or integrate a variety of teaching techniques, such as presentations, games, simulations, and visual materials. This tends to make the FC a comprehensive teaching methodology, and not just as a set of materials that can be used in a classroom. In a different way, while films and documentaries have been long used in classrooms, they are not yet recognised as an independent and self-sufficient teaching format. They

should probably be seen as background materials to complement and support teaching, not only in its traditional, frontal lecturing format but also in any other format that aims to make student learning more active and engaging, including a FC. When an entire class or a section of it is reorganized in terms of watching films and documentaries for students, as in our study, it might not necessarily improve students' understanding and knowledge (Kiasatpour, 1999, p. 85; Lieberfeld, 2007, p. 573; Inoue & Krain, 2014, p. 20; Swimelar, 2013, p. 24; Brandle, 2020). This conclusion seems to hold when assessing students' performance through both a relatively more objective measurement, that is their grades, and a more subjective one, that is their learning perception. Previous analyses based on VT showed that films and documentaries can hinder students' learning perception as they do not always get to transmit to them the impression that fictional representations are useful to master complex academic concepts (Betti et al. 2024; Lieberfeld, 2007, p. 573; Inoue & Krain, 2014, p. 20; Swimelar, 2013, p. 24; Brandle, 2020).

Several studies have warned to be careful when taking on students' perceptions about the effectiveness of teaching formats at a face value (Deslauriers et al. 2019). Sometimes, students can have a hard time evaluating what and how much they have learned of a subject. Their perceptions could be related to "a number of different factors", including "novelty" that are not necessarily relevant in pedagogical terms (McNally et al. 2017, p. 292). Nevertheless, when dealing with learning, students' perceptions about what they learn and what they expect to learn are important. Perceptions can be deceiving, especially when they are not accompanied by some objective measurement of what students effectively learned. However, this is not a sufficient reason not to include them in an empirical analysis about the impact of active learning on their performance. If students are the main actors of the learning process, their perceptions are a relevant source of data to understand what formats can be effective at improving learning. Improving learning perception can be, thus, quite useful in a teaching process as it can have positive consequences on students' motivation and engagement.

With all this, this study does not suggest that the FC is necessarily better than VT for students' academic performance. For example, neither VT nor the FC enhanced other important students' soft skills, such as self-efficacy and teamwork. This result both diverges from studies that observed an improvement in students' performance about these types of skills (Whitman Cobb, 2016) and converges with others that did not find any significant difference between using one teaching format or another in terms of students' soft skills (Jenkins 2015, p. 610).

These results, rather, lead us to the conclusion that, instead of seeing FC and VT, or any other active learning teaching format, as mutually exclusive or incompatible, future teaching approaches should be based on a complementary integration of different teaching strategies, techniques, and materials. Instead of making students' learning depend on a single teaching innovation, teachers should be ready to combine different techniques, depending on the goals and characteristics of each class (Jenkins 2015, p. 610; Betti et al. 2020; Betti et al., 2022). Different teaching formats can provide different advantages and benefits. This can foster distinct but equally meaningful learning processes. On the one hand, in an FC format, students can view the contents of lectures before class time. Based on those contents, they must perform exercises in class, under professor's supervision and in collaboration with classmates. This process can lead to a constructive and mutually reinforcing dialogue among students and between students and the professor, capable of generating deeper and enduring learning. On the other hand, in a VT format, students are required to study political and social phenomena through visual materials that they must analyse before class and through visual content that they must produce at home and present during class time. This also can create the conditions for an iterative dialogue among themselves and with the professor. Both formats can, thus, provide context for different but complementary learning processes, both potentially capable of producing meaningful learning.

Moreover, no active teaching format should aim to replace traditional elements of teaching, such as teachers' frontal lectures or students' notes. On the one hand, there are moments in a class in which the complexity of concepts might require frontal class lecturing. This can allow teachers to more precisely evaluate students' responses to explanations of specific concepts. On the other hand, when it is necessary to perform exercises to apply concepts or solve social and political problems, it can be more practical to ask students to study those concepts by watching a pre-recorded teacher's explanation at home and, then, come to class with hypotheses and ideas to be used in those exercises.

Without ignoring the difficulties to obtain this, ideally speaking, a classroom should be able to integrate different teaching strategies, techniques, and materials throughout the same year or semester (Betti et al. 2024; Betti et al. 2022). Depending on the necessity, this should include traditional frontal lecturing, pre-recorded video-lecturing, class exercises and activities, such as debates, presentations, and simulations, exams and assignments, and a combination of visual and non-visual materials, such as textbooks, notes, pictures, films, documentaries. In this regard,

we think that there cannot be a “one-size-fits-all” format, satisfactory for all teaching contexts and necessities.

## **Conclusions, Limitations and Future Research**

While the FC has been the object of a large literature aimed at empirically assessing its effects on students' learning, what we have so far about VT are mostly anecdotal descriptions of teaching implementations. The volume of empirical studies applied to the use of films and documentaries in PS or IR is still smaller than what one can find about the FC. More analyses are necessary to better understand the cognitive aspects of students' performance in which VT could be useful. This study aimed to contribute to this literature. Our tentative conclusion is that the FC should be considered a more complete and comprehensive teaching format that can help restructure class time and space in a way that maximizes the benefits of teaching and learning. VT through films and documentaries should be seen more as a teaching resource that can provide instructors with a large variety of visual materials to be used in class to complement the explanations of content or to manage evaluated activities, such as commentaries, debates, or small research projects, to be performed in class or at home. Upon reaching this conclusion, we are aware of the limitations of this study. It took place throughout a limited time span of only seven weeks. Moreover, it involved students with specific contextual characteristics, such as being enrolled in a private institution and coming from a similar upper-middle class socio-economic background. We are, thus, aware of the necessity to repeat the study with larger samples across longer time spans and to involve more randomly selected socio-economic backgrounds. However, we also think that this study provides a valuable starting point to comparatively assess the usefulness and effectiveness of different teaching formats based on active learning.

Future studies will need to provide more data to corroborate this conclusion. For example, they will need to further compare different teaching strategies to enhance our comprehension of the best aspects and practices that can be extracted by each of them, with the goal of productively integrating them into teaching. In addition, they should aim to understand the most effective solutions to evaluate student learning depending on the type of teaching format or technique applied. Finally, they should keep on devising research strategies to reliably include students' learning perceptions in the teaching process. This could entail the use of interviews, surveys, or focus groups. This would be useful to understand how different teaching formats can be adapted to the learning needs of different

students. Along these lines, it would be also important to examine the teachers' perceptions to understand how different teaching styles can suit different formats based on active learning. Due to the constant transformations that characterize the realms of teaching and learning, experimenting with applications and empirical assessments of classes based on FC, VT, and other potentially innovative teaching strategies remains the most effective way to enrich our understanding of improving students' learning.

As a conclusion, future studies should not only improve our empirical understanding of the effectiveness of teaching formats based on active learning, but also include in the debate their main pedagogical aspects. Future analyses should, for example, consider the importance of inclusivity for teaching. This entails evaluating whether teaching formats can guarantee equal access to technology, reduce the impact of socio-economic inequalities, and consider a diversity of learning needs. This is the only way to ensure that teaching formats can improve different students' profiles. This also requires universities to dedicate adequate financial and human resources to effectively implementing innovative teaching formats, so that teachers can acquire the competences to integrate them into their lectures. Along these lines, constantly examining teachers' and students' perceptions will be essential to identifying potential problems and devising solutions. For these goals to be achieved, universities will need to stimulate open and informed discussions based on partnerships and knowledge-based networks of teachers, educators, students, families, public and private institutions. Such networks will have to exchange the most appropriate teaching formats, multimedia resources, and digital tools capable of fostering students' learning.

## **Supplemental Materials**

Supplemental materials for the quantitative analysis, together with the full list of films and documentaries used for the study, are not included for space reasons but they can be made available upon request.

## **Ethical Standards**

The study received the authorization of the Teaching Innovation Evaluation Committee of the University where it took place. The Committee granted funds for the project "Implementation and assessment of Visual Thinking strategies in the Dual Degree in International Relations and Global Communications", for the academic period 2021-2023. Students agreed to participate in the study by filling in and signing a form. During the process,

they received full information about the research goals. Their personal data was not shared with any person or institution. Their participation was completely voluntary.

## References

- Anderson, L. W., & Krathwohl, D. (2001). *A Taxonomy for Learning, Teaching, and Assessing. A Revision of Bloom's Taxonomy of Educational Objectives*. New York: Longman.
- Baars, M., & Wijnia, L. (2018). The relation between task-specific motivational profiles of self-regulated learning skills. *Learning and Individual Differences*, (64), 125-137. <https://doi.org/10.1016/j.lindif.2018.05.007>.
- Bandura, A. (1994). Self-efficacy. In R. Vylayanur (ed.), *Encyclopaedia of Human Behaviour* (pp. 71-81). London: Academic Press.
- Berge, Z. L. and Nederveld, A. (2015). Flipped Learning in the Workplace. *Journal of Workplace Learning*, 27(2), 162-172. <https://doi.org/10.1108/JWL-06-2014-0044>.
- Bergmann, J. & Sams, A. (2012). *Flip Your Classroom: Reach Every Student in Every Class Every Day*. Eugene: ISTE.
- Betti, A., Biderbost, P. & García Domonte, A. (2020). Flipping the Classroom in Political Science: Student Achievement and Perceptions. *Revista de Ciencia Política*, 40(3), 589-615. <https://doi.org/10.4067/S0718-090X2020005000102>.
- (2022). Developing Students' "Soft Skills" through the Flipped Classroom: Evidence from an International Studies Class. *International Studies Perspectives*, 23(1), 1-24. <https://doi.org/10.1093/isp/ekab014>.
- Betti, A., Biderbost, P., Vaquero E. (2024). Is Visual Thinking Useful for Teaching International Relations? Evidence from an Implementation through Movies and Documentaries in an Undergraduate Class. In Seda Bayrakdar / P Seda Bayrakdar / Patrycja Chodnicka-Jaworska (eds.). *Current Dynamics in Business and Economics: Theoretical Approaches and Empirical Discoveries* (pp. 351-360). Berlin: Peter Lang.
- Bligh, D.A. (1998). *What's the Use of Lectures?* Bristol: Intellect.
- Brandle, S. M. (2020). Games, Movies, and Zombies: Making IR Fun for Everyone. *Journal of Political Science Education*, 16(4), 459-78. <https://doi.org/10.1080/15512169.2019.1568880>.
- Cheng, L., Ritzhaupt, A.D., & Antonenko, P. (2018). Effects of the flipped classroom instructional strategy on students' learning outcomes: a meta-analysis. *Education Technology Research Development*, <https://doi.org/10.1007/s11423-018-9633-7>.

- Collins, A., & Halverson, R. (2010). The Second Education Revolution: rethinking education in the age of technology. *Journal of Computer Assessed Learning*, 26, 18-27. <https://doi.org/10.1111/j.1365-2729.2009.00339.x>
- Dean, J. (2019). Sorted for Memes and Gifs: Visual Media and Everyday Digital Politics. *Political Studies Review*, 17(3), 255-66. <https://doi.org/10.1177/1478929918807483>.
- Deslauriers, L., Logan, S., McCarthy, K.M., Callaghan, K., & Kestin, G. (2019). "Measuring actual learning versus feeling of learning in response to being actively engaged in the classroom." *PNAS* 116(39): 19251-19257. <https://doi.org/10.1073/pnas.1821936116>.
- Galindo-Domínguez, H. (2021). Flipped Classroom in the Educational System: Trend or Effective Pedagogical Model Compared to Other Methodologies? *Educational Technology & Society*, 24 (3), 44-60.
- Giglio, E. (2002). Using Film to Teach Political Concepts. *European Political Science*, 1(2), 53-8. <https://doi.org/10.1057/eps.2002.10>.
- Gokcek, G., & Howard, A. (2013). Movies to the Rescue: Keeping the Cold War Relevant for Twenty-First Century Students. *Journal of Political Science Education*, 9(4), 436-52. <https://doi.org/10.1080/15512169.2013.835561>.
- Heck, A. (2017). Analyzing Docudramas in International Relations: Narratives in the Film *A Murderous Decision*. *International Studies Perspectives*, 18(4), 365-90. <https://doi.org/10.1093/isp/ekw012>.
- Hussain, M.I., Preetha, R., Naik, M.S., Panaskar, H.C., & Das, A.D. (2023). Assessing the Effectiveness of Flipped Classroom Strategy on Student Performance. *European Chemistry Bulletin*, 12, 2883-2896. <https://doi.org/10.1186/s12909-024-06105-7>.
- Inoue, C. Y. A. & Krain, M. (2014). One World, Two Classrooms, Thirteen Days: Film as an Active Teaching and Learning Tool in Cross-national Perspective. *Journal of Political Science Education*, 10(4), 424-442. <https://doi.org/10.1080/15512169.2014.947422>.
- Iretzberger, M. (2021). Film and its Unused Potential for Teaching International Relations: Exploring New Possibilities in a Classroom Experiment. *Journal of Political Science Education*, 17(1), 567-82. <https://doi.org/10.1080/15512169.2020.1854774>.
- Jenkins, S. (2015). "Flipping the Introductory American Politics Class: Student Perceptions of the Flipped Classroom." *PS: Political Science and Politics*, 48(4), 607-611. <https://doi.org/10.1017/S1049096515000840>.
- Jensen, J.L., Tyler, A.K., & Godoy, P. (2015). Improvements from a Flipped Classroom May Simply be the Fruits of Active Learning. *Life Sciences Education*, 14(1), 1-12. <https://doi.org/10.1187/cbe.14-08-0129>.

- Kiasatpour, S. M. (1999). The Internet and Films: Teaching Middle East Politics Interactively. *PS: Political Science and Politics*, 32(1), 83-9. <https://doi.org/10.2307/420755>.
- Kuzma, L. M., & Haney, P. J. (2001). And...Action! Using Films to Learn about Foreign Policy. *International Studies Perspectives*, 2(1), 33-50. <https://doi.org/10.1111/1528-3577.00036>.
- (2002). Comments by Kuzma and Haney: Using Film in the Classroom, *International Studies Perspectives*, 3, 92-4. <https://doi.org/10.1111/1528-3577.00081>.
- Lai, C.L., & Hwang, G.J. (2016). A Self-Regulated Flipped Classroom Approach to Improving Students' Learning Performance in a Mathematics Course. *Computers and Education*, 100, 126-140. <https://doi.org/10.1016/j.compedu.2016.05.006>.
- Lambach, D., Karger, C., & Goerres, A. (2017). Inverting the Large Lecture Class: Active Learning in an Introductory International Relations Course. *European Political Science*, 16(4), 553-569). <https://doi.org/10.1057/s41304-016-0078-3>.
- Lieberfeld, D. (2007). Teaching about War through Film and Literature. *PS: Political Science and Politics*, 40(3), 571-4. <https://doi.org/10.1017/S1049096507070837>.
- Lin, Y-N., Hsia, L-H., Sung, M-Y., & Hwang, G-H. (2019). Effects of integrating mobile technology-assisted peer assessment into flipped learning on students' dance skills and self-efficacy. *Interactive Learning Environments*, 27(8), 995-1010. <https://doi.org/10.1080/10494820.2018.1461115>.
- McNally, B., Chipperfield, J., Dorsett, P., Del Fabbro, L., Frommolt, V., Goetz, S., Lewohl, J., Molineux, M., Pearson, A., Reddan, G., Roiko, A., & Rung, A. (2017). Flipped Classroom Experiences: Student Preferences and Flip Strategy in a Higher Education Context. *Higher Education* (73), 281-298. <https://doi.org/10.1007/s10734-016-0014-z>.
- Naing C., Whittaker, M.A., Aung, H.H., Chellapan, D.K., & Riegelman, A. (2023). The effects of flipped classrooms to improve learning outcomes in undergraduate health professional education: A systematic review. *Campbell Systematic Reviews*, <https://doi.org/10.1002/cl2.1339>.
- Omelicheva, M. Y., & Avdeyeva, O. (2008). Teaching with Lecture or Debate? Testing the Effectiveness of Traditional versus Active Learning Methods of Instruction. *PS: Political Science and Politics* (July), 603-7. <https://doi.org/10.1017/S1049096508080815>.
- Roberts, D. (2017). The Message is in the Medium: Evaluating the Use of Visual Images to Provoke Engagement and Active Learning in Politics

- and International Relations Lectures. *Politics*, 38(2), 232-249. <https://doi.org/10.1177/0263395717717229>.
- Robles, M. M. (2012). Executive Perceptions of the Top 10 Soft Skills Needed in Today's Workplace. *Business Communication Quarterly*, 75(4), 453-465. <https://doi.org/10.1177/1080569912460400>.
- Rose, G. (2016). *Visual Methodologies: An Introduction to Researching with Visual Materials*, 4th Edition. London: Sage.
- Salas, E., Sims, D.E., & Burke, S.C. (2005). Is there a Big Five in Teamwork? *Small Group Research*, 36(5), 555-99. <https://doi.org/10.1177/1046496405277134>.
- Schmid, J. (2020). (Captain) America in crisis: popular digital culture and the negotiation of Americanness, *Cambridge Review of International Affairs*, 33(5), 690-712. <https://doi.org/10.1080/09557571.2019.1708864>.
- Stoddard, J. D. (2012). Film as a "thoughtful" medium for teaching history. *Learning, Media, and Technology*, 37(3), 271-88. <https://doi.org/10.1080/17439884.2011.572976>.
- Strelan, P., Osborn, A., & Palmer, E. (2020). The Flipped Classroom: A Meta Analysis of effects on student performance across disciplines and education levels. *Educational Research Review* 30. <https://doi.org/10.1016/j.edurev.2020.100314>.
- Sunderland, S., Rothermel, J.C. & Lusk, A. (2009). Making Movies Active: Lessons from Simulations. *PS: Political Science and Politics*, 42(3), 543-47. <https://doi.org/10.1017/S1049096509090878>.
- Swimelar, S. (2013). Visualizing International Relations: Assessing Student Learning through Film. *International Studies Perspectives*, 14(1), 14-38. <https://doi.org/10.1111/j.1528-3585.2012.00467.x>.
- Talbert, R., & Bergmann, J. (2017). *Flipped Learning: A Guide for Higher Education Faculty*. London: Routledge.
- Tisham, S., MacGillivray, D., & Palmer, P. (1999). *Investigating the Educational Impact and Potential of the Museum of Modern Art's Visual Thinking Curriculum*, Final Report, Harvard Project Zero.
- Touchton, M. (2015). Flipping the Classroom and Student Performance in Advanced Statistics: Evidence from a Quasi-Experiment. *Journal of Political Science Education*, 11(1), 28-44. <https://doi.org/10.1080/15512169.2014.985105>.
- Valeriano, B. (2013). Teaching Introduction to International Politics with Film. *Journal of Political Science Education*, 9(1), 52-72. <https://doi.org/10.1080/15512169.2013.747840>.
- Valeriano, B., & Habel, P. (2016). Who are the Enemies? The Visual Framing of Enemies in Digital Games. *International Studies Review*, 18. <https://doi.org/10.1093/isr/viv007>.

- Van der Zwan, N., & Afonso, A. (2019). Activating the Research Methods Curriculum: A Blended Flipped Classroom. *PS: Political Science and Politics*, 52(4), 749-753. <https://doi.org/10.1017/S1049096519000581>.
- Van Munster, R., & Sylvest, C. (2015). *Documenting World Politics: A Critical Companion to IR and non-fiction film*. London: Routledge.
- Wagner, D-A. (2018). Teachers' Use of Film in the History Classroom: A Survey of 19 High School Teachers in Norway. *Nordidactica*, [www.kau.se/nordidactica](http://www.kau.se/nordidactica).
- Weber, C. (2001). The Highs and Lows of Teaching IR Theory: Using Popular Films for Theoretical Critique. *International Studies Perspectives*, 2(3), 281-87. <https://doi.org/10.1111/1528-3577.00058>.
- Whitman Cobb, W. N. (2016). Turning the Classroom Upside Down: Experimenting with the Flipped Classroom in American Government. *Journal of Political Science Education*, 12(1), 1-14. <https://doi.org/10.1080/15512169.2015.1063437>.
- Yen, S-C., Lo, Y., Lee, A., Enriquez, J. (2018). Learning online, offline, and in-between: comparing student academic outcomes and course satisfaction in face-to-face, online, and blended teaching modalities. *Education and Information Technologies*, 23, 2141-2153. <https://doi.org/10.1007/s10639-018-9707-5>.





# **REVIEWS**



## Reviews

**Natallia Kabiak (2025). *Learn Russian through Contemporary Short Film*. London: Routledge. 134 pp. ISBN: 9781032739380**

In recent years, the integration of authentic media into language pedagogy has reshaped approaches to foreign language learning. *Learn Russian through Contemporary Short Film* (Routledge, 2025), authored by Natallia Kabiak, contributes to this evolving landscape by offering an innovative framework for teaching Russian as a foreign language (RFL) through the medium of contemporary short film. The textbook, designed for learners at the B1–C1 levels of the CEFR, combines linguistic rigor, cultural exploration, and media literacy within a coherent and adaptable structure.

Kabiak, a lecturer in Russian Studies at the University of Melbourne, has previously published textbooks that foreground authentic cultural materials. In this latest work, she expands on this approach by bringing twelve Russian and Kazakh short films (2014–2022) into the classroom. These films—ranging from *Собеседование* (The Interview) to *Варя* (Varya)—serve as pedagogical catalysts for developing both communicative competence and intercultural understanding. The result is a textbook that not only enhances linguistic proficiency but also stimulates empathy, critical thinking, and appreciation of contemporary Russian society.

The book is carefully organized into twelve self-contained modules, each centered on a single short film. Every module follows a consistent progression designed to scaffold language learning:

- Pre-viewing activities introduce contextual vocabulary, cultural concepts, and predictive discussion.
- Viewing tasks engage learners with audiovisual stimuli.
- Post-viewing exercises build receptive and productive skills through comprehension questions, grammar and word-building tasks, and structured discussions.
- Writing assignments and comprehension tests extend learning beyond viewing.
- Answer keys allow for independent study and assessment flexibility.

For instance, *The Interview* helps students explore professional discourse and moral dilemmas surrounding honesty and self-presentation, while *The Breath* opens ethical and emotional conversations about betrayal and forgiveness. Through these storylines, Kabiak fuses linguistic form, human emotion, and moral reflection, turning short film into a medium for intellectual and communicative growth.

Kabiak's pedagogical design is grounded in communicative, task-based, and sociocultural learning theories. Rather than teaching grammar as decontextualized rules, she embeds structures in meaningful communication. This inductive approach aligns with Schmidt's (1990) Noticing Hypothesis and Ellis's (2003) task-based framework, which advocate for discovery-driven learning.

The use of authentic short films also reflects Krashen's (1982) Input Hypothesis, providing comprehensible input enriched by emotional engagement. More broadly, Kabiak's classroom embodies Vygotsky's (1978) sociocultural theory, emphasizing learning as a socially mediated process supported by interaction and scaffolding.

Additionally, the book illustrates principles of multimodal learning (Kress & van Leeuwen, 2001): films combine visual, auditory, and linguistic codes, helping learners construct meaning across multiple channels. In this way, cinematic experience becomes not just supplementary material, but a central vehicle for linguistic and cultural acquisition.

A central strength of the textbook is its fusion of language with contemporary culture. Each short film reflects everyday realities in Russian and Kazakh contexts—employment struggles, family dynamics, environmental awareness, or personal transformation. The tasks accompanying these films foster the development of intercultural communicative competence (Byram, 1997) by encouraging learners to interpret, compare, and evaluate cultural perspectives through language.

The reflection-based exercises consistently move learners from comprehension toward critical evaluation and personal response. By inviting comparisons with their own societies, students not only practice Russian but also cultivate empathy and cultural literacy. In this sense, the book functions as both a linguistic and a humanistic project.

Kabiak's textbook stands out for several reasons:

- **Authenticity:** Use of unmodified cinematic texts exposes learners to natural speech, dialectal variety, and colloquial vocabulary.
- **Pedagogical balance:** Each module harmonizes all four skills—listening, speaking, reading, and writing.

- **Adaptability:** Modular design supports diverse teaching contexts, from university programs to independent study.
- **Self-directed learning:** The inclusion of answer keys empowers learners beyond the classroom.
- **Digital dimension:** Incorporation of real online viewer comments introduces learners to informal, digital Russian discourse.

Furthermore, by framing the viewing of short films as inquiry rather than entertainment, Kabiak repositions learners as active interpreters of meaning. This pedagogical stance aligns with the movement toward critical media literacy, where students learn to analyze how language and imagery interact to convey ideology and emotion.

Despite its numerous strengths, a few aspects merit consideration. Some learners—especially those at the lower end of the intermediate spectrum—might find the implicit grammar instruction challenging without additional support. Supplementary grammar notes or a digital companion could enhance accessibility.

Film access also depends on reliable internet connectivity and the continued availability of streaming links. A secure institutional platform or curated companion site would improve sustainability.

Lastly, the cultural density of the films, rich in local references and social nuance, may require instructor guidance. These challenges, however, are intrinsic to authentic material and underscore the need for thoughtful mediation rather than prescriptive control.

*Learn Russian through Contemporary Short Film* makes a compelling case for film-based instruction as a site of linguistic, cultural, and aesthetic education. It exemplifies the interdisciplinary turn in applied linguistics that integrates media studies, digital literacy, and intercultural theory.

Kabiak's model can inspire educators in other less commonly taught languages seeking innovative ways to engage students with real-world discourse. The book also offers fertile ground for future research in multimodal input processing, learner motivation, and media-based language acquisition.

Beyond its immediate pedagogical value, the textbook invites a reconceptualization of language education as an act of cultural participation—an engagement with art, empathy, and ethical understanding.

In sum, Natallia Kabiak's *Learn Russian through Contemporary Short Film* represents a forward-looking and methodologically sound contribution to Russian language pedagogy. By merging linguistic training with authentic cultural experience, the book revitalizes how language, film, and identity intersect in the classroom. Its modular structure, scholarly depth, and emotional resonance make it a valuable resource for teachers, learners, and researchers alike.

For programs in Russian Studies, Applied Linguistics, or Intercultural Communication, this textbook is not merely an instructional tool—it is a model of how education can humanize, modernize, and globalize language learning.

### **References**

- Byram, M. (1997). *Teaching and assessing intercultural communicative competence*. Multilingual Matters.
- Ellis, R. (2003). *Task-based language learning and teaching*. Oxford University Press.
- Krashen, S. (1982). *Principles and practice in second language acquisition*. Pergamon.
- Kress, G., & van Leeuwen, T. (2001). *Multimodal discourse: The modes and media of contemporary communication*. Arnold.
- Schmidt, R. (1990). The role of consciousness in second language learning. *Applied Linguistics*, 11(2), 129–158.
- Vygotsky, L. S. (1978). *Mind in society: The development of higher psychological processes*. Harvard University Press.

***Xiying Wang***

**Garrido Yserte, R. y Casanueva, H. (2025, Eds.). *Geopolítica de la Educación Superior: La Universidad como actor global*. Catarata. ISBN: 978-84-1067-419-6**

The field of Geopolitics has undergone a notable transformation in recent years. Traditionally limited to sociology or politics, its study in education was approached from two perspectives: Comparative Education (curriculum and instruction) and the History of Education. Until recently, approaches were scarce, with recent works in Spain such as those by the GIEEPE-URJC group in the Economics of Education standing out. Therefore, this more internationalized vision of Geopolitics in Higher Education makes this volume a valuable contribution from educational, sociological, and social perspectives. A glance at its contents, largely prepared by authors from the University of Alcalá, reveals a thematic variety relevant to the university of the future, the central objective of the book. It does not limit itself to describing the current state of the debate—between social protectionism and the neoliberal approach—but rather shows how the university can be a global actor capable of influencing the major changes of the 21st century. Although academically organized, each chapter has the practical idea that universities must assume challenges such as transfer, employability, and complexity in a world marked by technological, ecological, and social tensions that polarize education between the West and Asia. José Vicente Sanz's prologue emphasizes this moment of social and economic change for the university, with a special mention of the IAES Research Group as a facilitator of research for the transformation of knowledge and a driver of social action to change an entrenched pattern and seek new, more current dimensions. Afterwards, the introduction by Rubén Garrido Yserte and Héctor Casanueva moves us to reconsider what “university” means and its governance in a world full of uncertainties. I believe this introduction helps quite a lot because it explains that the university not only receives the impact of global changes but can also lead collective responses. The book is structured into three parts that address the relationship between higher education and global challenges. In the first, Concepción Olavarrieta analyzes threats such as the climate crisis and inequality, underlining that the university cannot remain static, while Héctor Casanueva presents the UN “Pact for the Future” as a multilateral response and highlights the role of the university as a bridge between policies and society. The second part focuses on the ethical and cooperative dimension: María del Carmen Patricia Morales reflects on intellectual and moral solidarity; Yuma Inzolia

proposes reimagining higher education to be inclusive and sustainable; Francisco Rojas Aravena and Emily Alfaro Rojas show how universities can cooperate for peace through academic diplomacy; and M<sup>a</sup> Jesús Such Devesa, M<sup>a</sup> Jesús Salado García, and Elena Mañas delve into the university's role regarding the SDGs, transforming them into concrete objectives. The third part addresses institutional transformation: Alfonso González Hermoso de Mendoza suggests that it is not enough to update programs, but that a cultural change is required; Xóchitl Arias González and Guillermina Benavides Rincón insist on training for complexity; Martina Bo provides the student perspective; María Teresa Gallo Rivera and Rubén Garrido Yserte discuss the university with purpose and the need to measure its social value; and the chapters by Óscar Montes Pineda, María Cristina Pineda de Carías, Yiem Ataucusi, and Garrido Yserte with Natalia Usach address digital equity, territory, and regional development, showing that the university must act globally and locally. Taken together, the work is not just a diagnosis, but a call to action: it posits that the university of the future must be flexible, committed, and aware of its ethical responsibility. Although some chapters are more conceptual, the book maintains coherence and offers a solid proposal for understanding how higher education is inserted into geopolitics and how it can be transformed to contribute to a more just and sustainable world.

In summary, this is a volume that comes to fill a gap that must be completed in the future. The geostrategy of education is an aspect called to have enormous relevance in the coming years and, therefore, this work is an excellent starting point for research that must advance in the already near future.

**Jesús García Laborda,**  
*Universidad de Alcalá*

**Gaukbar Niyetalina,**  
*Turan University*

**Patrick Parra Pennefather (2025). *Regenerating Learning: Transforming How You Learn with Generative AI*. 565 pp. ISBN: 979-8868810602**

Generative AI did not appear out of thin air; it marks a new phase in the decades-long evolution of machine learning. Tools such as ChatGPT are moving from isolated pilots to the center of everyday social and organizational life. Across the world, AI-driven integration is becoming routine—especially in well-resourced organizations and educational settings, where generative AI is being systematically embedded into work, teaching, and learning workflows. This shift, however, also brings ethical challenges and the pressure of an “efficiency above all” narrative. Generative AI can both reshape established ways of learning and working and amplify enduring social complexities, including bias and misinformation—making it feel less like a single innovation and more like an ever-intensifying alarm bell. Against this backdrop, *Regenerating Learning: Transforming How You Learn with Generative AI* engages directly with the rapid spread of generative AI. Drawing on rich cases and educational theory, it traces how human–AI interaction is reshaping learning mechanisms and revisits core questions: how we learn, what we learn, and what must be updated in this new technological landscape. With a stance of open curiosity and necessary critical discernment, the book clarifies common misconceptions, highlights iterative learning through practical use, and helps readers navigate opportunities and risks with sharper judgment. It offers actionable pathways for creative and knowledge workers while providing a broader framework for understanding learning transformation and capability renewal—balancing efficiency with quality and equity, and advancing a more inclusive, sustainable learning ecosystem for the future.

The book consists of twenty chapters. Chapter 1 stresses rewriting the narrative of learning with an open mindset, treating “learning from AI” as a starting point for transformation. Chapter 2 argues for reflecting on—and adjusting—existing teaching and learning models, using self-awareness to clarify how past experiences shape one’s learning. Chapter 3 frames generative AI as an opportunity to break entrenched pedagogical paradigms and rethink what “intelligence” means, while guiding readers to recognize their own learning patterns. Chapter 4 introduces “interrupted learning” in the workplace. Chapter 5 brings in the role of the “researcher,” emphasizing human-centered learning design. Chapter 6 focuses on action and practice, exploring generative AI’s potential to connect knowledge learning with real-world application. Chapter 7 highlights the need to build personal stand-

ards for evaluating AI outputs. Chapter 8 discusses how to adapt to AI: treating work as an ongoing process, embracing failure, breaking habits, and learning from AI-generated results. Chapter 9 centers on the “prototyping” process in human–AI interaction. Chapter 10 emphasizes iterative learning through the use of AI tools. Chapter 11 examines how to balance efficiency, ethics, and risk when using generative AI. Chapter 12 underscores the importance of understanding algorithms. Chapter 13 highlights continuous learning through ongoing interaction with AI systems. Chapter 14 introduces the greater sense of control enabled by customizable chatbots. Chapter 15 argues that AI literacy is only a starting point and that deeper, more specific learning is achieved through deliberate practice. Chapter 16 emphasizes the value of learning with others. Chapter 17 focuses on organizational learning and the reassessment of learning value as companies integrate AI. Chapter 18 stresses maintaining narrative control and creative authorship in AI-mediated content production. Chapter 19 analyzes evidence-free techno-hype and inflated claims. Chapter 20 returns to fundamental reflections on intelligence, creativity, and related core questions.

Overall, grounded in learning and educational theory, the book offers a nuanced analysis of how generative AI is reshaping workplace learning—how we learn, what we learn, when and where we learn, and why we learn. It moves beyond the traditional “school narrative,” situating learning within the broader realities of work and everyday life. Importantly, it is not merely a how-to guide for using tools; it advances a more transferable learning methodology. Through exploratory interaction with generative AI, the book invites readers to reflect on and reconfigure their learning habits, and—via the “iceberg model”—underscores that tool proficiency is only the visible surface, beneath which lie deeper opportunities for learning ecosystems and organizational transformation. Throughout, it promotes “curiosity with caution,” helping readers separate genuine value from technological hype and offering timely guidance for learners and knowledge workers worldwide. By emphasizing self-regulation and iterative learning while confronting practical concerns such as bias, hallucinations, privacy, and security, it equips readers with a sturdier framework for navigating the trade-offs between efficiency and risk.

However, although the book offers a fairly systematic, learning-theory-grounded account of how generative AI is reshaping key aspects of workplace learning, there remains room for further development in several respects. On the one hand, the book effectively combines theory with practice and draws on rich cases, yet most examples come from contexts with relatively strong technical and resource conditions. As a result, it offers limited coverage of variations across regions, industries, and populations

with different levels of access to resources. Future work could therefore conduct comparative studies in more diverse settings and further articulate institutional integration pathways within formal educational environments such as schools. On the other hand, while the book highlights the importance of organizational policies and support, the governance framework for translating ideas into implementation still warrants greater specificity. For instance, future research could map issues such as safety and privacy protection, data sovereignty, and cross-border compliance to different usage scenarios by risk tier, and develop reusable cases and templates to clarify requirements and make them easier to execute—thereby advancing the discussion toward more concrete, actionable guidance.

Undoubtedly, by bridging social context and educational theory, the book offers a systematic account of how generative AI is reshaping learning and work, the risks it introduces, and practical pathways for responding. It not only deepens our understanding of “learning” itself, but also links technological adoption to the transformation of learning ecosystems, providing a framework that is both theoretically grounded and practically useful for future research. Moreover, by confronting the structural imbalances that an “efficiency above all” narrative can generate, it encourages organizations and educational systems to build more durable consensus and governance around privacy, safety, and equity. It further offers methodological support for cross-cultural collaboration and the development of global competence—advancing workplace and learning transformation in the generative AI era and contributing to a more inclusive, well-governed, and sustainable learning ecosystem.

#### **Funding Information:**

- 2019 Department of Education of Yunnan Province Scientific Research Fund Project: Research on the Graduate Guidance Path in Foreign Language Disciplines in East China under the Background of “Double First-Class” Construction (Project No. 2019J0217).
- 2019 Foreign Language Teaching and Research Press National University Foreign Language Teaching and Research Project Translation Special Fund: Research on the Development and Influencing Factors of Self-Efficacy in Novice Translators (Project No. 2020YN0025).

**Zizhao Xiong,  
Qiuju Su,  
Yang Yao,**

*Facultad de Lenguas Extranjeras (Escuela Internacional),  
Universidad Forestal del Suroeste, Chín*

Revista de Educación es una publicación científica del Ministerio de Educación y Formación Profesional español. Fundada en 1940, y manteniendo el título de Revista de Educación desde 1952, es un testigo privilegiado de la evolución de la educación en las últimas décadas, así como un reconocido medio de difusión de los avances en la investigación y la innovación en este campo, tanto desde una perspectiva nacional como internacional. La revista es editada por la Subdirección General de Atención al Ciudadano, Documentación y Publicaciones, y actualmente está adscrita al Instituto Nacional de Evaluación Educativa de la Dirección General de Evaluación y Cooperación Territorial.



GOBIERNO  
DE ESPAÑA

MINISTERIO  
DE EDUCACIÓN, FORMACIÓN PROFESIONAL  
Y DEPORTES