Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-025-04553-0 OPEN

Check for updates

Construction and validation of an observational instrument to assess infant executive functions through playing

Elena Escolano-Pérez[1]✉ & Fernando Martín-Bozas[2]

The development of executive functions in childhood constitutes a key aspect for the subsequent development and learning of children. Early assessment of these functions can be crucial for detecting difficulties and implementing optimizing educational strategies. However, the assessment of children's executive functions still poses a challenge for teachers. To overcome this challenge, play is the primary tool for learning and assessment in early childhood education. The importance of play in early childhood education as a learning driver is present and validated from educational legislation to international scientific literature. Therefore, considering the strengths of play and systematic observation, the latter being the most appropriate methodology for evaluating child behaviour, the aim of this work was to construct an observation instrument to assess children's executive functions through play. The construction of the observation instrument was carried out following the steps established by observational methodology. Through an iterative process, preliminary sessions, contributions from the literature, and meetings among researchers were conducted to carry out the construction process. For its validation through Generalizability Theory, observation sessions were conducted with a sample of 17 children aged 4-5 years in early childhood education. This observation instrument, called IEEFECH (Instrument for Evaluation of Executive Functions in Early Childhood Education), was validated through the use of Generalizability Theory. The results supported the validity and reliability of the instrument. Therefore, IEEFECH constitutes an assessment tool for children's executive functions based on play that will allow teachers and other specialists to obtain objective, valid, and reliable information from which to design and implement educational strategies that respond to the needs of each child. In addition, this observation instrument will be useful using play as the central axis of evaluation, given its relevance as a method of learning, exploration, and assessment in early childhood education.

[1] Department of Psychology and Sociology, Faculty of Education, University of Zaragoza, C/Pedro Cerbuna, 12, 50009 Zaragoza, Spain. [2] Department of Educational Sciences, Faculty of Social and Human Sciences, University of Zaragoza, C/Atarazanas, 4, 44003 Teruel, Spain. ✉email: eescola@unizar.es

Introduction

The first six years of life lay the foundation for subsequent development and learning (Coelho et al. 2020; Rosas et al. 2019). Within the international context, and according to the International Standard Classification of Education (ISCED), these early years correspond to ISCED level 0: early childhood education. The purpose of early childhood education is to ensure the comprehensive and harmonious development of each student in all areas (physical, emotional, sexual, social, cognitive, and artistic), preparing them to be competent citizens in an ever-changing world. Early childhood education marks the beginning of the acquisition process of key competencies for lifelong learning; that is, the initiation of acquiring the skills necessary for students to progress successfully throughout their educational journey and address the main global and local challenges of a changing world. Key competencies constitute the ultimate goal of education and training at all educational stages, aiming to prepare students for future life in society once they leave the educational system (European Union, 2019). Among these key competencies is the ability to learn how to learn, which is closely linked to executive functions (EF) (Sala et al. 2020).

EF are cognitive and affective processes necessary for developing lifelong learning skills and for everyday functioning, as they allow us to coordinate our thoughts, actions, and emotions to achieve our goals, being especially necessary in novel, complex, changing situations that generate uncertainty and/or involve conflict (Diamond, 2020; Zelazo & Carlson, 2020). There are many models attempting to determine how many and which processes comprise EF. Although there are discrepancies among them, there is also some consensus in considering the following aspects as components of EF: the ability to initiate behaviors, inhibition, selection of relevant goals for the task, planning, organization of means, monitoring, flexibility, working memory, and regulation of responses of both emotional and behavioral nature (Diamond, 2013; Diamond, 2020).

A multitude of studies have demonstrated that EF are not only relevant for academic success but also for numerous areas of life: physical and mental health, work productivity, social competency, etc. (Coelho et al. 2020; Scerif et al. 2023). It has also been evidenced that the level of EF reached in early childhood is crucial for later development through middle childhood, adolescence, and adult life (Robson et al. 2020). Well-developed EF are a "kit" for life success (Diamond, 2014). Therefore, the acquisition of EF in the early years of life is critical for optimizing development. These issues justify that, while EF are of great interest in all educational stages—because, as mentioned earlier, they are linked to the key competence for lifelong learning of learning to learn—their approach is especially relevant in early childhood education (Diamond, 2016).

Although EF development is protracted, from birth to early adulthood, during the years corresponding to early childhood education, and especially between the ages of 3 and 6, EF show a significant peak of development (Poowanna et al. 2022). These changes in EF occur simultaneously with numerous brain changes (especially in the dorsolateral prefrontal cortex, but also in other interconnected areas) that underlie EF development. Thus, EF development occurs in parallel with the maturation of the dorsolateral prefrontal cortex—the main area underlying EF—and the refinement and specialization of its connections with other brain areas with which it is interconnected (Thompson & Steinbeis, 2020). However, it is well-known that brain development, and therefore EF development, is not innate, nor does it occur automatically. Brain and EF development depend on the dynamic interaction between environmental and genetic influences (Miguel et al. 2023; Veraksa et al. 2023). Therefore, optimal EF development requires adequate environmental input, especially since epigenetics emphasizes the importance of environmental influences on cognitive outcomes, as changes in the environment can lead to changes in heritability estimates of different skills such as EF (Freis et al. 2022). It is well-known that in the early years of life, the brain shows greater plasticity, meaning the brain is particularly sensitive to environmental influences—both positive and negative—making these early years a period of special malleability for brain and EF development (Miguel et al. 2023).

Early childhood education attendance is considered one of the most effective environmental inputs for promoting the development of children's human capital and specifically, to promote EF development (Del Boca et al. 2023). The early childhood education curriculum framework is based on play, and more precisely, on "learning through play" (Parker et al. 2022). Play is an intrinsic activity for the child, it occurs spontaneously and not only provides pleasure and enjoyment but also significant benefits for development and learning. Young children naturally explore and learn many skills and knowledge through play. Children learn to make sense of the world around them through play. When they play, children are developing their cognitive, physical, and communication skills (Mardell et al. 2019; Gibb et al. 2021). Again, this is supported by neuroscience findings showing that the prefrontal cortex of the brain is refined by play (Parker et al. 2022).

Thereby, play is an essential element to enhance brain and learning development, and consequently, also for the development of EF (Gibb et al. 2021; Traverso et al. 2019; Veraksa et al. 2023). Consequently, it is an indispensable tool for the teaching-learning process of children (Zosh et al. 2018).

Among the different typologies of existing play, it has been demonstrated that role play, games with rules, and building blocks are the ones that most enhance EF development (Lillard, 2015; Veraksa et al. 2023; Yang et al. 2022). Characteristics such as intense engagement and excitement are common aspects to all these types of play that could determine their potential for EF development (Veraksa et al. 2023).

As previously indicated, play is a natural, fun, and meaningful way to learn. Consequently, systematic observation of a child's behavior during play makes it possible to obtain relevant data to describe, explain, and understand fundamental aspects of the child's development and learning, such as the development of EF (Pyle et al. 2020; Yogman et al. 2018). In this sense, both early childhood education curriculum and the scientific literature indicate that systematic observation (characterized by the study of spontaneous behavior that occurs naturally in everyday contexts) is the optimum tool for learning and development assessment of early children's students (Peterson & Elan, 2020). Observing a child's actions through play allows an educator or other professional to accumulate a record of the child's development; to get information about what he knows and what he can do. It allows understanding the child's strengths, challenges, and needs. With this information, educators can begin to plan appropriate curriculum and effective individualized instruction for each child. Assessment using systematic observation is a critical part of a high-quality early childhood program. Systematic observation brings together the inherent objectivity and rigor necessary in educational evaluation, with the flexibility needed to capture the many and often complex changes that occur in a child's behavior in real-life settings (Belasko et al. 2023; Belza et al. 2020).

However, data from different research studies suggest that assessment of learning and development in early childhood education using systematic observation is still a challenge. The number of early childhood education teachers with specialized knowledge and methodological training on systematic observation is limited (Allen & Kelly, 2015). This leads them to frequently conduct assessments based on their subjective perceptions, which may overlook difficulties, limitations, or special educational needs of the students. As a result of this inadequate assessment, there could be children whose development is compromised by not receiving the stimulation, attention, and instruction they truly require. Early childhood teachers require more information about systematic observation (Cueto et al. 2017). More assessment instruments that provide ecological, accurate, and reliable information are also demanded by the scientific literature to adequately assess early children (McCoy, 2019; Silva et al. 2022; Souissi et al. 2022).

The scarcity of ecological measurement instruments that allow capturing valid and reliable information about childhood development, in general, and EF in particular, can be attributed to the rigorous work involved in constructing such instruments. Systematic observation lacks standard evaluation instruments, as this would limit the diversity of human behaviour and the contexts in which it occurs. Therefore, the construction of ad hoc instruments is required to not only capture the variability of human behaviour but also the diversity of contexts in which it occurs and may influence it. However, these ad hoc instruments must enable valid and reliable evaluation of the study object. This requires conducting pre-data analysis controls to ensure the quality of the recorded data, in order to account for possible errors or biases in these data, which could lead to erroneous conclusions about child development and learning, with the serious implications that entails. These necessary pre-data analysis controls include intraobserver reliability (ensuring that data obtained by an observer can generally be applied to an infinite number of intraobserver moments); inter-observer reliability (ensuring that data obtained by two or more observers can generally be applied to an infinite number of observers), and the validity of the constructed ad hoc instrument (ensuring that data obtained with the categories that compose the instrument cannot be generalized to an infinite number of other different categories). All of these processes to control the quality of observed data can be analysed using Generalizability Theory (GT), a multivariate structure developed by Cronbach et al. (1972) and adapted to observational designs by Mitchell, (1979). The aim of GT is to distinguish real variability from error variability in any measurement. GT provides information on sources of error that affect behavioural measurement. Therefore, GT is a theory of the multifaceted errors of behavioural measurement (Blanco-Villaseñor, 2001).

In relation to all of the above, the aim of this study was to construct and validate, using GT, an observational instrument aimed at assessing the EF employed by 4-5-year-old children while they engage in a playful activity.

Methods

Setting. The research took place in a public early childhood education center located in a city in the northeast of Spain (specifically, in the autonomous community of Aragon). The study was conducted in a classroom belonging to the final year of early childhood education.

The daily educational programming included a period of free play, during which children could engage in activities according to their interests using the materials available at each of the small learning stations called learning corners located inside the classroom. Each week, every child was required to play in all the corners.

Learning corners represent one of the current alternatives for organizing early childhood education classrooms, as stipulated by educational legislation in Spain for this stage. In each learning corner, numerous proposals are offered for engaging in activities of various types belonging to different educational domains. These corners are delimited spaces in the classroom where children work individually or in small groups simultaneously on different learning activities. Depending on the type of activity, the child's action may require direction from the teacher or assistance from them, while in other activities, children may act autonomously. Working through corners allows for addressing the differences, interests, and learning rhythms of each child. Ultimately, this type of classroom organization promotes a diversity of learning options for students, integrating all educational domains in an integrated, playful, and enriching manner. The corners serve various educational purposes related to the constructivist conception of learning (Conde Vélez et al. 2019).

In this study, the focus was on the children's activities while playing a game proposed in the logical-mathematical learning corner of the classroom. This game, called Camelot Jr., is explained later.

Design. An observational design was employed. Systematic observation is considered the optimal methodology for assessing childhood learning and development, particularly when studying children in a natural setting such as school. Specifically, following Anguera et al. (2008), the observational design was nomothetic, point, and multidimensional. It was nomothetic because multiple observation units were studied (precisely, 17 children were individually observed); it was point because a single session per participant was observed to assess their EF; and it was multidimensional because various domains of EF (working memory, inhibition, planning, and mental flexibility) were analyzed, in accordance with various theoretical models (Diamond, 2013; Miyake et al. 2000; Zelazo & Carlson, 2020), as well as other child and adult actions of interest for EF development, such as the assistance provided by the adult to children (Duncan & Tarulli, 2003). All these relevant aspects constituted the ad hoc observation instrument, which is elaborated in detail in a subsequent section. The observation conducted adhered to scientific criteria. It was active (as the aim was predetermined), non-participatory (the observers did not interact with the children), and direct (allowing complete perceptibility of behaviors). It was carried out through direct observation of recorded footage (Anguera et al. 2018). The guidelines outlined in the Guidelines for Reporting Evaluations based on Observational Methodology (GREOM) (Portell et al. 2015) and the Methodological Quality Checklist for Studies based on Observational Methodology (MQCOM) (Chacón-Moscoso et al. 2019) were followed.

Participants. The sample was composed by 17 children: 10 were female (58.8%) and 7 were male (41.2%). Their age range ranged from 4 to 5 years old ($M = 4.74$; $SD = 0.26$). Specifically, some students were 4 years old because the assessment was conducted at the beginning of the academic year, prior to the natural year change. The socioeconomic level of the participants ranged from low to medium-high. The inclusion criteria for the sample were: 1) belonging to the last academic year of early childhood education; 2) having informed consent from parents/legal guardians authorizing participation in the study; 3) having an adequate level of spoken Spanish; 4) not being students with special educational needs, i.e., not presenting disabilities, disorders in language and communication development, attention or learning difficulties, exceptional abilities, lack of language proficiency, or personal circumstances that hinder regular attendance at the educational center (Official Gazette of Aragon, 2017). The participant data were treated following the Declaration of Helsinki and the Organic Law 3/2018, of December 5, on the Protection of Personal Data and guarantee of digital rights. Additionally, this research was approved by the corresponding Ethics Committee

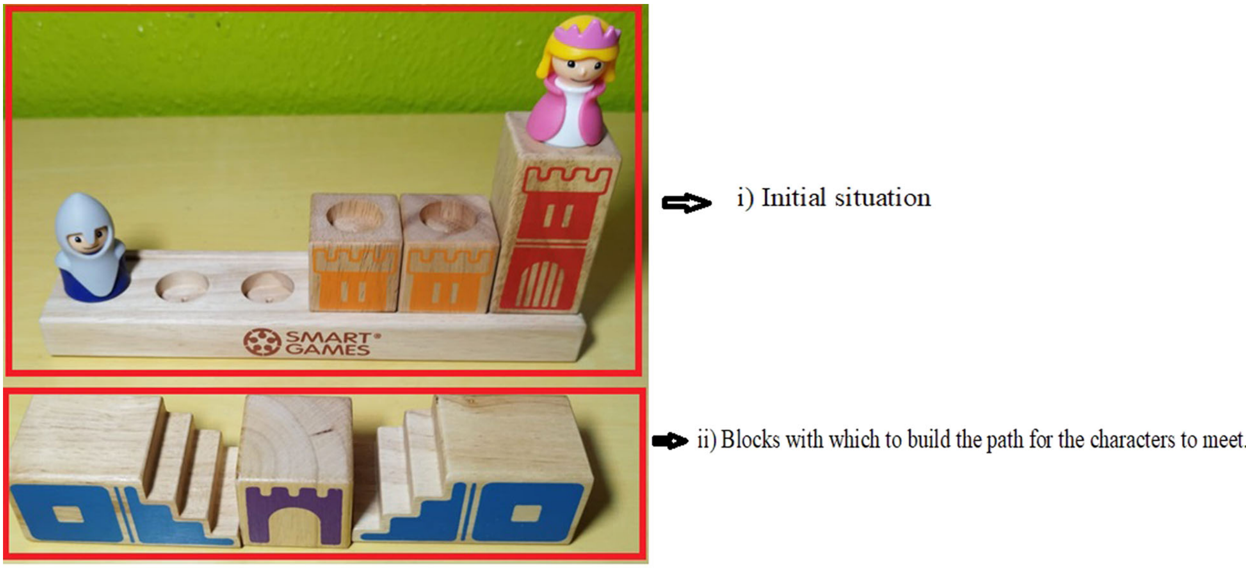

Fig. 1 Example of a challenge from the Camelot Jr. game 1) Starting situation. 2) Blocks with which to build the path for the characters to meet.

(favorable opinion RAT 2022-484, from the Ethics Research Committee of the Autonomous Community of Aragon, Spain).

Instruments

**Game that promotes the activation of EF*. To enhance the activation of EF in children, the Camelot Jr game was used, manufactured and marketed by the company Smart Games. It is a single-player game consisting of 48 challenges of progressive difficulty, a quality that makes the game suitable for players from 4 years old to 99 years old, according to the manufacturer's own brand information. The objective of the game is, in each challenge, to build a path with wooden blocks that allows the two characters of the game (a knight and a princess), located each in different places of a castle, to meet. For each challenge, both the initial position (i.e., the type of castle where the characters appear and the location of each character within it) and the blocks to use for the construction of the path vary. Both aspects are determined in the game's own instructions. Figure 1 illustrates, as an example, a challenge from the game (the initial situation of this challenge and the blocks with which to build the path are specified for better understanding).

In this study, the researcher provided the child with the initial position of each challenge and the blocks to use to construct the path. These blocks were offered to all children in the same orientation (a feature also determined in the game instructions). Thus, in each challenge, the child had to place these blocks in a specific location and orientation to build a path that could be traversed by the characters in order to meet. Additionally, this path construction had to be done following rules established—both in writing and graphically—by the game itself. Figure 2 outlines the rules to follow in constructing the path. Taking into account all these aspects, the Camelot Jr game can be considered both a game of rules and a wooden blocks game. As indicated in Introduction section, both types of games are related to EF development (Lillard, 2015; Veraksa et al. 2023; Yang et al. 2022). Specifically, to solve each challenge of the Camelot Jr game, the child had to plan where to place the pieces and in what orientation, remember the rules, and use them to solve the challenge (working memory); inhibit hasty responses during execution; monitor their execution and detect their own errors, as well as generate new alternatives if their initial solution was incorrect (cognitive flexibility). In conclusion, the game required the activation of different EF.

* *Recording instrument for activity*. To record the activity of each child while playing the Camelot Jr game (and thus subsequently be able to carry out a comprehensive observation and evaluation of their EF), the recording camera of a Huawei P30 Lite mobile phone was used.

**Observation instrument*. To evaluate the EF that the child put into action in each Camelot JR game challenge, an observation instrument was created ad hoc, as required by the observational methodology itself (Anguera, 2001). Since the construction and validation of the observation instrument was the objective of this work, its construction process and resulting characteristics are detailed in the Procedure and Results sections.

* *Recording instrument for observational data*. To carry out the coded recording of observational data (i.e., to code the child's actions indicative of his/her EF), the free software LINCE PLUS was used (Soto et al. 2019).

**Analysis instrument*. To calculate the validity and reliability of the constructed observation instrument using the GT, the freely accessible software SAGT v1.0 (Hernández-Mendo et al. 2016) was used.

Procedure

**Ethical Permissions*. The research team held informative meetings with the management team of the educational center and with the families of potential participants, obtaining informed consent from both. In addition, the favorable opinion of the corresponding ethics committee was obtained (Ethics Research Committee of the Autonomous Community of Aragon, Spain; reference RAT 2022-484).

**Recordings of the play sessions*. Before starting the recordings of the play sessions, the researcher immersed themselves in the classroom for one week, accompanying the teacher and children during this time. The aim was to familiarize the children with their presence and therefore, avoid reactivity bias (Anguera, 2001), i.e., to prevent the presence of the researcher from affecting the children's usual behavior, so that it would lose spontaneity. Only the natural and spontaneous behavior of the children constitutes an adequate indicator of their real level of development and learning (Anguera

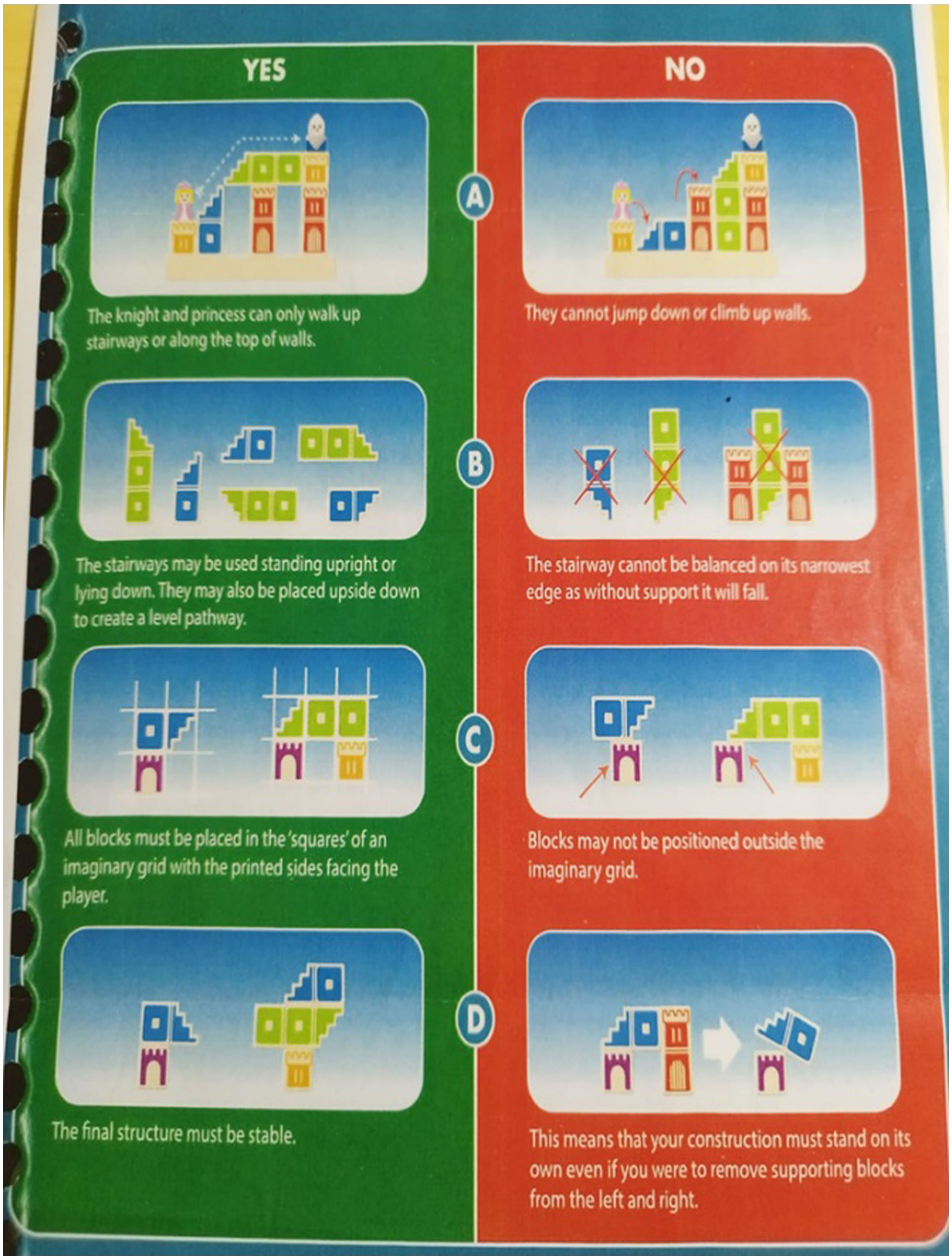

Fig. 2 Rules to comply with in the build of the path for the characters to meet.

et al. 2018). After this period of habituation by the children to the presence of the researcher, the play sessions began.

The play sessions took place in the participants' own classroom, during school hours, and more specifically, at the usual times allocated for free play in the different corners of the classroom. The play session for each child was conducted when they freely went to the logical-mathematical corner, to which they had to go at least once a week, as mentioned previously.

For each child, the play session with Camelot Jr proceeded as follows. In the logical-mathematical corner, the researcher was present, who presented and explained the game to the child. Specifically, they explained the objectives of the game and its rules, using the graphic examples included in the game material itself (see Fig. 2). These rules involved the following aspects: (1) the characters could not jump to reach their destination, they could only climb up and down stairs, and walk on flat ground; (2) ladder-shaped pieces could be placed vertically and horizontally; (3) pieces had to be aligned with the board so that they did not go beyond it; (4) the placement of pieces had to be stable and they could not fall; (5) all the pieces provided by the adult had to be used in solving the challenge. (6) Additionally, the pieces forming part of the initial challenge situation could not be moved.

Once it was ensured that the child understood the rules of the game through example tests and questions about the rules, the game would begin. In each challenge or level, the researcher presented the challenge in the corresponding initial position and provided the child with the pieces needed to solve the challenge (both the initial situation and the pieces to be used following the instructions given by the game for each challenge).

Before the child began their action, that is, prior to the child's execution of the challenge, the researcher asked the child to think about how they would solve the challenge (i.e., the adult promoted planning in the child). Thus, the child was given unlimited time to think about how they would execute the challenge. At the same time, once the child had thought about it, they were asked to explain what the execution process would be. Once they had thought about it, the child had to verbalize it. Only then could the child begin to place pieces to solve the challenge. During the execution of the challenge, if certain situations arose (such as the child making a mistake when placing a piece in violation of a challenge instruction), the adult could offer graduated assistance to guide the child towards successful resolution of the challenge. When the child indicated that they had completed the challenge, the researcher asked the child some questions aimed at enhancing their self-assessment. (These questions are found in Annex 1). In this way, in each challenge, three main moments or phases could be differentiated—planning, execution, and evaluation—phases closely linked to different EF.

After the child answered these self-assessment questions (regardless of the quality of their answers), the adult presented another challenge to the child, following the same procedure just explained. In this way, challenges were presented to each participant. The end of the game, and therefore, the session for each child, occurred when during the execution of two challenges (consecutive or not), the child had needed assistance from the adult (regardless of the amount of assistance required in the execution of each challenge). There was no time limit for solving each challenge or the total playtime for each child, although the average duration of the play sessions was 16 minutes.

Each and every child was recorded with a camera while playing to subsequently be able to evaluate their actions indicating their levels of EF through observation. In order to avoid children's reactivity bias, the video camera was positioned in front of the child but away from them, making it possible to record children's activity without their presence being able to alter it.

**Construction of the observation instrument*. The process of constructing the observation instrument followed a continuous development and revision process until achieving its final version, following the contributions of Anguera, (2001) and Anguera et al. (2020). These contributions have been taken into account in the use of observational methodology in recent research (Fuentes-Moreno et al. 2020).

First, preliminary sessions were conducted through narrative records which, subsequently, through various semi-systematization processes, the last of which was a list of distinctive features, allowed for the discernment of the different actions carried out by the studied children. Some of these distinctive features were, for example: repetition of action without paying attention, adult demonstrates help action, anticipation of mistake, placing pieces in different locations, positions, and orientations…

Subsequently, a first provisional system of categories was developed. This system was adapted mainly through the confluence between our empirical records and the theoretical framework offered by the scientific literature on EF (Diamond, 2020; Zelazo & Carlson, 2020). Planning is understood as the ability to provide cognitive control and organize behaviors to achieve the desired goal (Deng et al. 2022). Inhibition involves the capacity to deliberately control or override our automatic or dominant behaviors and thoughts (Escolano-Pérez & Bravo, 2017). Regarding working memory, it focuses on the capacity to store and manipulate information temporarily (Byom et al. 2021). Finally, cognitive flexibility entails the ability to mentally shift between different tasks or novel sets (Escolano-Pérez & Bestué, 2021).

Also, it has been taken into account observational empirical works focused on early cognition and, specifically, on early infant EF (Blanco-Villaseñor & Escolano-Pérez, 2017; Escolano-Pérez, 2020; Escolano-Pérez & Acero-Ferrero, 2022; Escolano-Pérez et al. 2019; Escolano-Pérez & Sastre, 2010). In this way, an amalgamation was made between: a) theoretically defined and studied concepts in the specific scientific literature about early EF development; b) categories that were part of observation instruments constructed and used in other empirical works on the construction of child activity and executive functioning, and c) our empirical records that allowed us to know other types of actions not previously defined by other authors.

Through several pilot sessions, the first version of the observation instrument was refined, leading to multiple intermediate versions (up to 12 versions, in which different criteria and categories were adjusted, ensuring mutual exclusivity between the categories nested in the same criteria and refining their definitions, examples, and counterexamples) until the observation instrument considered suitable (version 13) was constituted for its use, that is, for carrying out the encoded recording of the play sessions recorded on video.

* *Encoded recording*. In order to perform the recording of observational data using the LINCE PLUS software program, the video recordings and corresponding observation instrument were imported into the program. Once both elements were imported, LINCE PLUS allowed the video recordings to be viewed, with the observation instrument simultaneously displayed on the screen. Whenever a child's action corresponding to one or several categories of interest (and therefore, corresponding to one or several categories of the observation instrument) was detected, the video playback was paused and these categories were recorded. To do this, the corresponding categories of the observation instrument were clicked with the mouse. This process allowed for the encoded recording of the child's actions.

In this study, this coding process was initially carried out by one observer, who recorded all the play sessions. After 2 weeks, this same observer re-coded all the play sessions again. Additionally, and independently, another observer coded all the play sessions. Both observers were experts in infant development, EF, and systematic observation.

* *Data analysis: validation of the observation instrument*. To calculate the validity and reliability (intra- and inter-observer) of the ad hoc constructed observation instrument, the principles of GT were followed. Previous studies have addressed the validation of observation instruments (Romero-Jara et al. 2023) in different fields, but validations of instruments based on GT are less frequent. More specifically, for the calculation of the instrument's validity, a crossed design of two facets—Observer (O) and Categories (C)—was used, where Observer was the differentiation facet and Categories were the instrumentalization facet (O/C). Within the use of GT, and more specifically, for calculating intra-observer reliability, a crossed design of two facets (Categories and Time) was used, considering Categories as a differentiation facet and Time as an instrumentation or generalization facet (Categories/Time = C/T). For calculating inter-observer reliability, another crossed design of two facets was used, with Categories as the differentiation facet and Observers as the instrumentation facet (Categories/Observers = C/O).

As mentioned earlier, analyses corresponding to GT were conducted using the SAGT v1.0 software program.

Results

The ad hoc observational instrument designed to evaluate early EF through play was a mixed observational tool, comprising a field format integrated with category systems. Consistent with the methodological requirements of this type of observation instrument (Anguera, 2001; Anguera et al. 2018), categories nested within each criterion of the field format were exhaustive and mutually exclusive. It consisted of 17 criteria and 68 categories. The 17 criteria primarily represented different executive components in children (working memory, inhibition, planning, and mental flexibility), as well as the assistance provided by the adult to guide the child in their playful activity. Each criterion was specified by a minimum of two categories and a maximum of seven categories, resulting in a total of 64 categories, as previously mentioned. A summary of this observational instrument built ad hoc—and which was called Instrument for Evaluation of Executive Functions in Early Childhood Education (IEEFECH)—can be consulted in Supplementary material.

The 17 criteria were as follows:

1. Challenge Planning: the child takes time to think and explain where they will place the pieces.
2. Adherence to Planning: the initial placement corresponds to what was explained in the planning.
3. Correct Piece Placement: the child places the piece in the appropriate location and position to solve the challenge.
4. Mistake: either failure to comply with game instructions by the child or placing the piece in a location that does not solve the challenge.
5. Mistake Prediction and Correction: when facing a potential error, the child anticipates and avoids committing it.
6. Flexibility: generating alternative courses of action when faced with a situation requiring new actions where previous ones are ineffective.
7. Challenge Resolution: completing the challenge, regardless of the outcome, or restarting if necessary.
8. Resolution Evaluation: the child assesses the outcome upon finishing the challenge.
9. Consistency in Resolution Evaluation: focusing on the accurate evaluation of the challenge resolution by the child.
10. Resolution Justification: the child explains the reasons behind their evaluation of the resolution.
11. Complexity Evaluation: the child's verbal and/or non-verbal opinion on the challenge difficulty.
12. Challenge: the game's difficulty presented to the child that they must solve.
13. Piece: refers to the game piece involved in the action.
14. Response to Adult Assistance: the extent to which the child considers the assistance provided by the adult immediately after it is given.
15. Gradual Assistance: the type of support the child receives, either by adult decision or child request.
16. Instruction: the rules the child must follow during each challenge.
17. External Behaviors: actions of the child incompatible with challenge or game progress, temporarily or permanently leading to their suspension or interruption for reasons other than those considered in the game instructions.

These criteria could be coded in at least one of the three main phases established for each challenge: A) Planning Phase: before the child starts the action with the game pieces, i.e., when the child must think and verbally express how they will solve the challenge, with possible codes being Challenge Planning and Adherence to Planning); B) Execution Phase: while the child is solving the challenge (manipulating the pieces), with possible codes being Correct Piece Placement; Mistake; Mistake Prediction and Correction; Flexibility; Challenge Resolution; C) Evaluation Phase: once the challenge is solved, when the child must answer the adult's questions about the quality of their execution, with possible codes being Resolution Evaluation; Consistency in Resolution Evaluation; Resolution Justification; Complexity Evaluation. Additionally, several criteria could be coded in more than one of these three phases. These criteria, susceptible to coding in multiple phases, were: Challenge; Piece; Response to Adult Assistance; Gradual Assistance; Instruction; External Behaviors.

Table 1 Intra-observer reliability for the C/T design.

	Sum of Squares	df	Mean Squares	Standard Error	%
Time (T)	0.07	1	0.070	0.039	0.00
Categories (C)	23613.367	63	374.815	32.897	92.73
TC	890.430	126	14.134	2.479	7.27
Reliability Index	0.96				
Generalizability Index	0.98				

Table 2 Inter-observer reliability for the C/O design.

	Sum of Squares	df	Mean Squares	Standard Error	%
Observers (O)	0.07	1	0.070	0.001	0.00
Categories (C)	40658.49	63	645.373	56.603	99.97
OC	5.43	126	0.086	0.015	0.03
Reliability Index	1.00				
Generalizability Index	1.00				

Validity of the observation instrument. Within the methodological framework of the GT, an instrument is considered valid when variability corresponding to the category facet is very high and guarantees, therefore, the observational instrument's capacity for discrimination, which translates to a generalisability coefficient equal or near to 0, as is the case with this research work (coefficient G relative and absolute = 0.00) (Barbero et al. 2023).

The variance associated with the Observers facet has been 100% for all participants and 0% for both the Categories facet and the interaction facet Observer x Categories. The generalizability coefficients for this design structure have been zero (0.00) for all participants, ensuring the validity of the instrument, indicating that the homogeneity of the categories is optimal (Blanco-Villaseñor et al. 2014).

Intra-observer reliability. As shown in Table 1, the determination of sources of variance revealed that a significant portion of the variability (92.73%) was related to the Categories facet, with null or minimal variability in the Time facet (0%) as well as in the Time x Categories interaction (7.27%). Results from the generalizability analysis indicated high reliability in result generalization (0.98).

Inter-observer reliability. As presented in Table 2, the determination of variance sources showed that almost all variability (99.97%) was related to the Categories facet, being null or virtually null in the other facet, Observers (0%), as well as in the interaction of both: Observers x Categories (0.03%). Results from the generalizability analysis indicated high reliability in result generalization (1.00).

Discussion

The aim of this study was to construct and validate (through the use of GT) an observation instrument that would allow for the valid and reliable assessment of early EF exhibited by 4-5-year-old preschool children while solving a playful task. The constructed instrument, in its final version, consists of 17 criteria nested within 68 categories, enabling the evaluation of children's EF, distinguishing at which moment or phase of the playful task (Planning Phase, Execution Phase, Evaluation Phase) each of them may occur. The main EF assessed by this observation instrument include planning, inhibition, working memory, and cognitive flexibility. Additionally, it highlights the possibility of obtaining information about other actions of interest for the utilization of EF and the resolution of the playful task, such as the graded assistance provided by the adult to the child to solve the task.

The results of the generalizability coefficients were favourable, indicating that the validity and reliability, both intra- and inter-observer, of the instrument calculated through the use of GT, were excellent. Therefore, it is confirmed that the proposed instrument meets the methodological standards required in terms of reliability, precision, and validity. Consequently, the use of this observation instrument allows for obtaining objective, valid, and reliable information about various executive components (primarily planning, inhibition, working memory, and cognitive flexibility) while preschool children (4-5 years old) play. These results, therefore, once again highlight how play and systematic observation are resources of great relevance and usefulness in early childhood education. In this regard, we consider it necessary to advocate for the use of the most appropriate resources at each educational stage, and therefore, to promote the use of play as a means of learning in early childhood. Unfortunately, early childhood education is gradually evolving towards learning goals and methods traditionally associated with the Primary school curriculum, which has negative consequences for the optimal development and learning of students (Jahreie, 2023).

The findings of this research, namely the constructed observation instrument, constitute a significant contribution to the field of early childhood education, as it allows teachers to assess the development and learning of their students with the sufficient quality standards required, thereby overcoming the limitations that, as they themselves admit, often afflict the assessments they conduct on their students (Cuetos et al. 2017). Having valid and reliable instruments is a first step towards being able to detect the needs of each child, and from there, to design educational strategies that adequately address those needs and are thus effective. Identifying children's needs and supporting them in enhancing their abilities is a characteristic that defines quality early childhood education programs (Darling-Hammond et al. 2020). In this regard, consequently, this observation instrument contributes to improving the quality of early childhood education, and therefore, contributes to the achievement of Sustainable Development Goal 4: Quality Education. Specifically, it aids in addressing goal 4.2, which focuses on ensuring quality early education services and tools for proper development during the preschool stage.

Despite the contributions and implications of this study that have already been discussed, an additional point of significance is the incorporation of GT. This theory constitutes a powerful and useful tool for assessment research, yet it is scarcely employed within the educational assessment domain. However, it must be acknowledged that a notable limitation of this study is its small sample size. In the future, it would be interesting to increase the number of participants. However, it must not be forgotten that systematic observation is an intensive methodology, not an extensive one. The interest lies in the exhaustive description of the natural behavior of a small number of participants, not in how representative it is of a larger universe (Anguera et al. 2018; Belza et al. 2019; Portell et al. 2015). Consequently, it involves working with a small number of participants but collecting a large amount of data with high accuracy. This high-precision data is of great relevance for being able to respond to children's needs and contribute to the improvement of their development.

Data availability

Dataset is available from the corresponding author on reasonable request.

Received: 28 March 2024; Accepted: 11 February 2025;
Published online: 23 October 2025

References

Allen LR, Kelly BB (2015) Transforming the Workforce for Children Birth Through Age 8: A Unifying Foundation. National Academies Press (US), Washington

Anguera MT (2001) Cómo apresar las competencias del bebé mediante una aplicación de la metodología observacional. Contextos educ 4:13–34

Anguera MT, Blanco-Villaseñor A, Losada JL et al. (2020) Integración de elementos cualitativos y cuantitativos en metodología observacional. Ámbitos Rev int Comun 49:49–70. https://doi.org/10.12795/Ambitos.2020.i49.04

Anguera MT, Blanco-Villaseñor A, Losada JL et al. (2018) Pautas para elaborar trabajos que utilizan la metodología observacional. Anu Psico 48:9–17. 10.1344/%25x

Anguera MT, Chacón S, Blanco A (coords.) Evaluación de programas sociales y sanitarios. Síntesis (2008)

Barbero JR, Lapresa D, Arana J et al. (2023) An observational analysis of kicker–goalkeeper interaction in penalties between national football teams in international competitions. Int J Perform Anal Sport. 196-212. https://doi.org/10.1080/24748668.2023.2214969

Belasko M, Herrán E, Anguera MT (2023) Observing dressing in a collaborative or trusting environment: The right to participate in one's own care at the Emmi Pikler nursery school in Budapest. Early Years: Int J Res Dev 43(4-5):778–793. https://doi.org/10.1080/09575146.2021.2016652

Belza H, Herrán E, Anguera MT (2020) Early childhood, breakfast, and related tools: analysis of adults' function as mediators. EUR J Psicólogo Educativo 35:495–527. https://doi.org/10.1007/s10212-019-00438-4

Belza H, Herrán E, Anguera MT (2019) Early childhood education and cultural learning: systematic observation of the behaviour of a caregiver at the Emmi Pikler nursery school during breakfast. J Study Educ Dev 42(1):128–178. https://doi.org/10.1080/02103702.2018.1553268

Blanco-Villaseñor A (2001) Generalizabilidad de observaciones uni y multifaceta: Estimadores LS y ML. MCCO 3:161–193

Blanco-Villaseñor A, Escolano-Pérez E (2017) Análisis de datos observacionales mediante la Teoría de la Generalizabilidad y la utilización del Modelo Lineal General y Mixto: Un estudio empírico de desarrollo infantil. An. Psicol 33(3):450–460. https://doi.org/10.6018/analesps.33.3.271021

Blanco-Villaseñor A, Castellano J, Hernández-Mendo A et al. (2014) Aplicación de la TG en el deporte para el estudio de la fiabilidad, validez y estimación de la muestra. RPADEF 23(1):131–137

Byom L, Whaln M, Turkstra L (2021) Working memory for emotions in adolescents and young adults with traumatic brain injury. Brain Impairment 23(3):296–310. https://doi.org/10.1017/BrImp.2021.20

Chacón-Moscoso S, Anguera MT, Sanduvete-Chaves S et al. (2019) Checklist de calidad metodológica para estudios basados en metodología observacional. Psicothema 31(4):458–464. https://reunido.uniovi.es/index.php/PST/article/view/16988

Coelho LA, Amatto AN, Gonzalez CL et al. (2020) Building executive function in pre-school children through play: a program. Int J Play 9:128–142. https://doi.org/10.1080/21594937.2020.1720127

Conde Vélez S, Delgado-García M, García Prieto FJ (2019) Diseño y validación de un instrumento para analizar el trabajo por rincones en las aulas de Educación Infantil. Estudios Sobre Educación 26:53–83. https://doi.org/10.15581/004.36.53-83

Cronbach LJ, Gleser GC, Nanda H et al (1972) The dependability of behavioral measurements: Theory of generalizability for scores and profiles. Wiley, New York

Cueto S, Prieto JA, Nistal P et al. (2017) Teachers' Perceptions of Preschool Children's Psychomotor Development in Spain. Percept Mot Skills 124:725–739. https://doi.org/10.1177/0031512517705534

Darling-Hammond L, Flook L, Cook-Harvey C et al. (2020) Implications for educational practice of the science of learning and development. Appl Dev Sci 24:97–140. https://psycnet.apa.org/doi/10.1080/10888691.2018.1537791

Del Boca D, Monfardini C, See S (2023) Early Childhood Education Attendance and Students' Later Outcomes in Europe. B E J Economic Anal Policy 23(4):1081–1136. https://doi.org/10.1515/bejeap-2022-0260

Deng M, Cai D, Zhou X et al. (2022) Executive Function and Planning Features of Students With Different Types of Learning Difficulties in Chinese Junior Middle School. LDQ 45(2):134–143. https://doi.org/10.1177/0731948720929006

Diamond A (2013) Executive Functions. Annu Rev Psychol 64:135–168

Diamond A (2014) Want to Optimize Executive Functions and Academic Outcomes? Simple, Just Nourish the Human Spirit. Minn Symp Child Psychol Ser 37:205–232

Diamond A (2016) Why improving and assessing executive functions early in life is critical. In JA Griffin, P McCardle, LS Freund (Eds.), Executive function in preschool-age children: Integrating measurement, neurodevelopment, and translational research. Am Psychol Assoc. 11–43. https://doi.org/10.1037/14797-002

Diamond A (2020) Executive functions. Handb Clin Neurol 173:225–240. https://doi.org/10.1016/B978-0-444-64150-2.00020-4

Duncan RM, Tarulli D (2003) Play as the leading activity of the preschool period: insights from Vygotsky, leont'ev, and bakhtin. Early Educ Dev 14:271–292

Escolano-Pérez E, Acero-Ferrero M (2022) Evaluating in the Real-World Educational Intervention to Improve Interference Control in Children with Autism Spectrum Disorder. Children 26:1294. https://doi.org/10.3390/children9091294

Escolano-Pérez E (2020) Intra- and Inter-Group Differences in the Cognitive Skills of Toddler Twins with Birth Weight Discordance: The Need to Enhance Their Future from Early Education. Sustainability 12(24):10529. https://doi.org/10.3390/su122410529

Escolano-Pérez E, Bestué M (2021) Academic Achievement in Spanish Secondary School Students: The Inter-Related Role of Executive Functions, Physical Activity and Gender. IJERPH 18(4):1816. https://doi.org/10.3390/ijerph18041816

Escolano-Pérez E, Bravo MA (2017) Procesos cognitivos y afectivos implicados en la resolución de problemas: Desarrollo e intervención. Miscelánea Comillas 75:41–69

Escolano-Pérez E, Sastre-Riba S (2010) Early infant cognitive assessment: Validity of and instrument. Behav Res Methods 42(3):759–767. https://doi.org/10.3758/BRM.42.3.759

Escolano-Pérez E, Herrero-Nivela ML, Anguera MT (2019) Preschool Metacognitive Skill Assessment in Order to Promote Educational Sensitive Response From Mixed-Methods Approach: Complementarity of Data Analysis. Front psychol 10:1298. https://doi.org/10.3389/fpsyg.2019.01298

European Union Key competences for lifelong learning. Luxembourg (2019)

Freis SM, Morrison CL, Lessem JM et al. (2022) Genetic and environmental influences on executive functions and intelligence in middle childhood. Dev Sci 25(1):e13150. https://doi.org/10.1111/desc.13150

Fuentes-Moreno C, Sabariego-Puig M, Ambros-Pallarés A (2020) Developing social and civic competence in secondary education through the implementation and evaluation of teaching units and educational environments. Humanit Soc Sci Commun 7(39). https://doi.org/10.1057/s41599-020-0530-4

Gibb R, Coelho L, Van Rootselaar NA et al. (2021) Promoting Executive Function Skills in Preschoolers Using a Play-Based Program. Front Psychol 12:720225. https://doi.org/10.3389/fpsyg.2021.720225

Hernández-Mendo A, Blanco-Villaseñor Á, Pastrana JL et al. (2016) Sagt: aplicación informática para análisis de generalizabilidad. RIPED 11(1):77–89

Jahreie J (2023) Early childhood education and care teachers' perceptions of school readiness: A research review. Teach Teach Educ 135:104353. https://doi.org/10.1016/j.tate.2023.104353

Lillard AS (2015) The Development of Play. Handb Child Psychol Developmental Sci 11:1–44. https://doi.org/10.1002/9781118963418

Mardell B, Solis SL, Bray O (2019) The state of play in school: Defining and promoting playful learning in formal education settings. Int J Play 8:232–236. https://doi.org/10.1080/21594937.2019.1684157

McCoy DC (2019) Measuring young children's executive function and self-regulation in classrooms and other real-world settings. Clin Child Fam Psychol Rev 22:63–74. https://doi.org/10.1007/s10567-019-00285-1

Miguel PM, Meaney MJ, Silveira PP (2023) New Research Perspectives on the Interplay Between Genes and Environment on Executive Function Development. Biol Psychiatry 94(2):131–141. https://doi.org/10.1016/j.biopsych.2023.01.008

Mitchell SK (1979) Interobserver agreement, reliability, and generalizability of data collected in observational studies. Psychol Bull 86:376–390. https://doi.org/10.1037/0033-2909.86.2.376

Miyake A, Friedman NP, Emerson MJ et al. (2000) The unity and diversity of executive functions and their contributions to complex "frontal lobe" tasks: A latent variable analysis. Cogn Psychol 41(1):49–100

Official Gazette of Aragon Decreto 188/2017 de 28 de noviembre del Gobierno de Aragón, por el que se regula la respuesta educativa inclusiva y la convivencia en las comunidades educativas de la Comunidad Autónoma de Aragón [Decree 188/2017 of November 28 of the Government of Aragon, which regulates the inclusive educational response and coexistence in the educational communities of the Autonomous Community of Aragon]. Boletín Oficial de Aragón December 18 240:36450-36465. (2017) https://www.boa.aragon.es/cgibin/EBOA/BRSCGI?CMD=VEROBJ&MLKOB=994406845050

Parker R, Thomsen BS, Berry A (2022) Learning Through Play at School – A Framework for Policy and Practice. Front Educ 7:751801. https://doi.org/10.3389/feduc.2022.751801

Peterson G, & Elam E Observation and Assessment in Early Childhood Education. https://childdevelopment.org/docs/default-source/pdfs/observation-and-assessment-english2-8-20.pdf (2020)

Poowanna B, Sarnkong R, Wangsitthidet S et al The Development Executive Functions for Early Childhood in 21st Century. J Educ Learn 11(4). https://doi.org/10.5539/jel.v11n4p193 (2022)

Portell M, Anguera MT, Chacón-Moscoso S et al. (2015) Guidelines for reporting evaluations based on observational methodology. Psicothema 27(3):283–289. https://doi.org/10.7334/psicothema2014.276

Pyle A, DeLuca C, Danniels E et al. (2020) A Model for Assessment in Play-Based Kindergarten Education. Am Educ Res J 57(6):2251–2292. https://doi.org/10.3102/0002831220908800

Robson DA, Allen MS, Howard SJ (2020) Self-regulation in childhood as a predictor of future outcomes: a meta-analytic review. Psychol Bull 146:324–354. https://doi.org/10.1037/bul0000227

Romero-Jara E, Solanellas F, Muñoz J et al (2023) Connecting with fans in the digital age: an exploratory and comparative analysis of social media management in top football clubs. Humanit Soc Sci Commun 10(858). https://doi.org/10.1057/s41599-023-02357-8

Rosas R, Espinoza V, Porflitt F et al. (2019) Executive Functions Can Be Improved in Preschoolers Through Systematic Playing in Educational Settings: Evidence from a Longitudinal Study. Front Psychol 10:2024. https://doi.org/10.3389/fpsyg.2019.02024

Sala A, Punie Y, Garkov V et al. LifeComp: The European Framework for Personal, Social and Learning to Learn Key Competence. Publications Office of the European Union, Luxembourg. https://doi.org/10.2760/302967 (2020)

Scerif G, Blakey E, Gattas S et al. (2023) Making the Executive 'Function' for the Foundations of Mathematics: The Need for Explicit Theories of Change for Early Interventions. Educ Psychol Rev 35:110. https://doi.org/10.1007/s10648-023-09824-3

Silva C, Sousa-Gomes V, Fávero M et al. (2022) Assessment of preschool-age executive functions: A systematic review. Clin Psychol Psychother 29(4):1374–1391. https://doi.org/10.1002/cpp.2718

Soto A, Camerino O, Iglesias X et al. (2019) LINCE PLUS: Research Software for Behaviour Video Analysis. Apunts Educ Fís 137:149–153. https://doi.org/10.5672/apunts.2014-0983.es.(2019/3).137.11

Souissi S, Chamari K, Bellaj T (2022) Assessment of executive functions in school-aged children: A narrative review. Front Psychol 13:991669. https://doi.org/10.3389/fpsyg.2022.991699

Traverso L, Viterbori P, Usai MC (2019) Effectiveness of an executive function training in Italian preschool educational services and far transfer effects to pre- academic skills. Front Psychol 10:2053. https://doi.org/10.3389/fpsyg.2019.02053

Thompson A, Steinbeis N (2020) Sensitive periods in executive function development. Curr Opin Behav Sci 36:98–105. https://doi.org/10.1016/j.cobeha.2020.08.001

Veraksa AN, Veresov NN, Sukhikh VL et al (2023) Play to Foster Children's Executive Function Skills: Exploring Short-and Long-Term Effects of Digital and Traditional Types of Play. IJEC. https://doi.org/10.1007/s13158-023-00377-8 (2023)

Yang W, Ng DTK, Gao H (2022) Robot programming versus block play in early childhood education: Effects on computational thinking, sequencing ability, and self-regulation. Br J Educ Technol 53:1817–1841. https://doi.org/10.1111/bjet.13215

Yogman M, Garner A, Hutchinson J et al. (2018) The power of play: A pediatric role in enhancing development in young children. Pediatrics 142(3):20182058. https://doi.org/10.1542/peds.2018-2058

Zelazo PD, Carlson SM (2020) The neurodevelopment of executive function skills: Implications for academic achievement gaps. Psychol Neurosci 13(3):273–298. https://doi.org/10.1037/pne0000208

Zosh JM, Hirsh-Pasek K, Hopkins EJ et al. (2018) Accessing the inaccessible: redefining play as a spectrum. Front Psychol 9:1124. https://doi.org/10.3389/fpsyg.2018.01124

Author contributions

Corresponding author: Elena Escolano-Pérez. She was involved in conceptual and methodological structure, literature review, data analysis, interpretation of results, discussion, writing—original draft, writing—review and editing, and acquisition of the financial support. Fernando Martín-Bozas was involved in literature review, investigation, data coding, data analysis, interpretation of results, discussion, writing—original draft, and acquisition of the financial support. Both authors contributed to revising the manuscript and provided final approval of the version to be published.

Funding

This study was supported by the Aragon Government Research Group, Grupo de Investigación de Referencia "Educación y Diversidad" (EDI) (Grant number S49_23R). Further, Elena Escolano-Pérez gratefully acknowledges the support of Spanish Government (PGC2018-098742-B-C31 (2019-2021) and the Department of Psychology and Sociology of the University of Zaragoza. Fernando Martín-Bozas also thanks the Aragon Government for the research training contract obtained through competitive call for the period 2021-2025.

Competing interests

The authors declare no competing interests.

Ethical approval

The study followed the principles of the Declaration of Helsinki and the Organic Law 3/2018, of December 5, on the Protection of Personal Data and guarantee of digital rights (2018, Official State Gazette no. 294, of December 6). The study was endorsed from the management team of the participants' school. Although the study did not involve human experiments, ethical approval for this study was obtained from the Research Ethics Committee of the Autonomous Community of Aragon, Spain (reference RAT 2022-133; 30/11/2022).

Informed consent

Although this study did not involve human experiments, informed consent was obtained from the participants' parents, who agreed to their children's participation and the use of their data for research purposes. Consent was collected in December 2022, following the provision of all relevant information about the study. The personal data collected will be used exclusively for this research, applying rigorous techniques to ensure anonymity, including the assignment of random codes to conceal personal identities throughout the research process.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-025-04553-0.

Correspondence and requests for materials should be addressed to Elena Escolano-Pérez.