



ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN
ASTROFISICA

Ciclo 38

Settore Concorsuale: 02/C1 - ASTRONOMIA, ASTROFISICA, FISICA DELLA TERRA E DEI PIANETI

Settore Scientifico Disciplinare: FIS/05 - ASTRONOMIA E ASTROFISICA

**ASTRONOMY FOR EDUCATION. EXPLORING THE POTENTIAL OF PUBLIC
ENGAGEMENT THROUGH GAME-BASED AND PLAYFUL LEARNING**

Presentata da: Rachele Toniolo

Coordinatore Dottorato

Andrea Miglio

Supervisore

Stefania Varano

Co-supervisor

Silvia Casu
Germano Bianchi
Jader Monari

Esame finale anno 2026

Contents

Contents	ii
Introduction	2
1 Theoretical framework: engagement, games and learning	4
1.1 Public Understanding of Science	5
1.2 The crisis of the deficit model	7
1.3 Public Engagement with Science	10
1.4 PES in schools: games, play and learning-by-doing	13
2 Methodologies: game-based and playful learning	17
2.1 The pedagogical framework of GBL and PL	19
2.2 The role of astrophysics	22
2.3 Evaluate the effectiveness	25
2.3.1 Nature of Science	27
2.3.2 Attitudes and self-efficacy in astronomy	31
2.4 The research questions	32
3 Case studies overview and rationale	34
3.1 PIXEL - Picture (of) the Universe	34
3.1.1 Game design process	37
3.1.2 Evaluation of engagement and scientific relevance	39
3.2 Cosmo Hunters	42
3.3 Design choices and implications for evaluation	46
4 PIXEL - Picture (of) the Universe: experiencing NOS	48
4.1 PIXEL embedded in a GBL program	48
4.1.1 Sample and research design	49
4.1.2 Data analysis	51
4.1.3 Results	53
4.1.4 Discussion	58
4.2 Short-term impact: a comparison with an out-of-school activity	59
4.2.1 Sample and research design	60
4.2.2 Data analysis	62
4.2.3 Results	63

4.2.4	Discussion	72
4.3	Conclusion	73
5	Cosmo Hunters: from game mechanics to learning outcomes	75
5.1	Sample and research design	76
5.2	Data analysis	80
5.3	Results	81
5.3.1	Engagement	81
5.3.2	Learning outcomes	81
5.4	Discussion	86
5.5	Conclusion	87
6	Conclusion	89
7	AMACA: the challenge of multisensoriality	92
	References	95
A	PIXEL supporting materials	107
A.1	Tables	107
A.2	Quantitative data analysis	118
B	Cosmo Hunters supporting materials	122
B.1	Debriefing	122
B.2	Frontal lecture	123
B.3	Data	123
C	AMACA supporting materials	126
C.1	The educational path	126

*To my beloved Nonno Valentino,
who through his example showed me that the right to education
is both a gift and a privilege never to be wasted.*

Introduction

Over the past decades, the relationship between science and society has become an increasingly central topic in both academic research and institutional debate. Concerns regarding declining interest in science, the spread of misinformation, and the growing complexity of socio-scientific issues have highlighted the need to rethink how science is communicated within society. In this context, Public Engagement with Science (PES) has emerged as a key framework for promoting democratic and participatory forms of interaction between scientific communities and the public.

Within this broader landscape, science education – and schools in particular – occupy a critical yet challenging position. On the one hand, schools play a fundamental role in shaping future citizens’ scientific literacy and their capacity to engage meaningfully with science-related issues. On the other hand, formal educational contexts are characterized by structural constraints, such as standardized lecture-oriented practices, which can limit the effectiveness of engagement-based approaches. Indeed, a substantial body of research has documented students’ declining interest in school science.

In response to these challenges, active and participatory methodologies have gained increasing attention within science education. Among them, Game-Based Learning (GBL) and Playful Learning (PL) have been widely proposed as promising approaches to foster meaningful learning supported by intrinsic motivation. The exploitation of playfulness and experimentation shifts the focus from passive knowledge transmission to learning-by-doing practices that place students at the center of the educational process.

In this context, this thesis aims to investigate the effectiveness of PES activities based on GBL and PL within the context of astrophysics education. Astrophysics is chosen as the disciplinary domain for several reasons. First, it is widely recognized for its strong inspirational potential and its capacity to attract learners’ interest across several age groups. At the same time, astrophysics is often communicated through amazing images that emphasize spectacular results while skipping the methodological processes and their human dimensions. This makes it a particularly fertile ground for exploring educational approaches that foreground the practices and dynamics that contribute to the knowledge production.

To address this issue, effectiveness is examined across multiple dimensions, including learning outcomes, understanding of the Nature of Science, and affective engagement

with it such as students' attitudes toward astronomy and their self-efficacy. This multidimensional perspective reflects the assumption that meaningful science education cannot be reduced to content acquisition alone, but must also consider how learners relate to science.

Methodologically, the research adopts a case study approach, focusing on two board games, PIXEL and Cosmo Hunters, designed for different educational aims and target groups. The evaluation and research strategy integrated quantitative and qualitative methods in order to capture both measurable outcomes and the underlying processes through which learning and engagement occur.

The contribution of this thesis is threefold. From a theoretical perspective, it contributes to the ongoing discussion on PES in formal education by explicitly linking game-based methodologies to the theoretical frameworks of Nature of Science and affective dimensions of engagement. From a methodological standpoint, it offers an example of how mixed-methods evaluation can be systematically integrated into the design and analysis of educational interventions. Finally, from a practical perspective, it provides insights that may inform educators and science communicators interested in developing and evaluating playful approaches to science education.

The thesis is structured as follows. Chapter 1 outlines the theoretical framework, situating the study at the intersection of science communication, public engagement, and research in education. Chapter 2 introduces the methodological foundations of game-based and playful learning and discusses the evaluation dimensions adopted in this work. Chapter 3 presents an overview of the selected case studies and their design rationale. Chapters 4 and 5 report and discuss the results of the empirical studies on PIXEL and Cosmo Hunters, respectively, while Chapter 6 synthesizes the main findings and reflects on their implications for research and practice in astrophysics education. Finally, Chapter 7 introduces the challenge of multisensoriality as the future development of this trajectory.

Chapter 1

Theoretical framework: engagement, games and learning

The theoretical framework of this thesis is situated at the intersection between science communication and formal school learning contexts. Before proceeding, it is necessary to clarify what is meant by science communication and how this field has evolved over time. Indeed, communicating science is not merely a matter of transferring information, but a practice that reflects how science and the scientific community relate with society.

From a theoretical point of view, the term *science communication* describes a wide range of activities, practices, and research efforts aimed at communicating science in two ways: to experts and non-experts. On the other hand, the first approach includes all communication practices internal to the scientific community, such as participation in specialized conferences or peer-reviewed publications in academic journals, which contribute to the development of scientific knowledge. The second category encompasses, for instance, science journalism, communication through social media and traditional media, science popularization at various levels, outreach, science festivals, and science centers – to name a few. In the following, based on the purpose of this thesis, the term science communication is used to refer to science communication to non-experts.

[Bryant \(2003\)](#) defines science communication as the set of processes through which scientific knowledge and culture are received by the wider community. [Burns et al. \(2003\)](#), on the other hand, offers a definition focused on outcomes: “is the use of appropriate skills, media, activities, and dialogue to produce one or more of the following personal responses to science: Awareness, Enjoyment, Interest, Opinion-forming, and Understanding”.

Over the past 50 years science communication has gained momentum, marked by strong dynamism and interdisciplinarity ([Guenther and Joubert, 2017](#)). In recent decades, it has developed into a field of study that has grown “at the intersection of science education, social studies of science, mass communication, museology, and several other longer-established academic and professional activities” ([Trench and Bucchi, 2010](#)). This success

is largely attributable to new research and innovation policies in European countries, which increasingly place at the center of the debate the need for a participatory and democratic model of science communication (Delaney and Tornasi, 2020).

Moreover, the growing attention of the scientific community toward science communication stems from a concern about the widening gap between science and society (Weingart et al., 2021), as made especially evident by public debates on socially impactful scientific issues, such as vaccine hesitancy during the COVID-19 pandemic or conflicting viewpoints with regards to the climate crisis. Similarly, the goals of science communication have evolved over time, varying according to the context and target audience. Based on interviews with scholars and educators in the field, Davies (2021) identifies six key roles that science communication can play in society: to ensure the accountability and legitimacy of publicly funded science; to serve practical purposes; to enhance democracy; to fulfill cultural functions; to achieve specific economic goals; and to act as promotion or marketing.

The following sections outline the milestones that have shaped the development of science communication. At times, it may seem like a semantic debate over the use of similar-sounding terms, but in reality, changes in vocabulary reflect deeper shifts in the relationship between science and society. It is important to stress that the perspective presented here is strongly Eurocentric, as many of the key developments occurred in the United Kingdom, one of the countries with the longest-standing tradition in the promotion and dissemination of scientific culture (Pitrelli, 2003).

1.1 Public Understanding of Science

The institutionalization of modern science communication took place in 1985. In that year, the UK Royal Society published a report, titled “Public Understanding of Science” (PUS) – also known as the Bodmer report after its chairman Sir Walter Bodmer – highlighting the deterioration in the relationship between science and public opinion (Figure 1.1). The report was released in a context marked by increasing public skepticism toward scientific institutions, driven by a series of technological controversies and environmental risks, such as debates on nuclear energy.

The document was drawn up by a working group composed of scientists, politicians, sociologists, and journalists, but notably lacked any representatives from the general public (Pitrelli, 2003). Its core assumption was that the gap between science and society was due primarily to a lack of knowledge among citizens, a perspective that strongly influenced objectives and tone of the document.

To address this growing divide, the report proposed several recommendations, addressing all stakeholders in science communication, including formal education, mass media, government, industry, and the scientific community. The main idea was to foster the de-

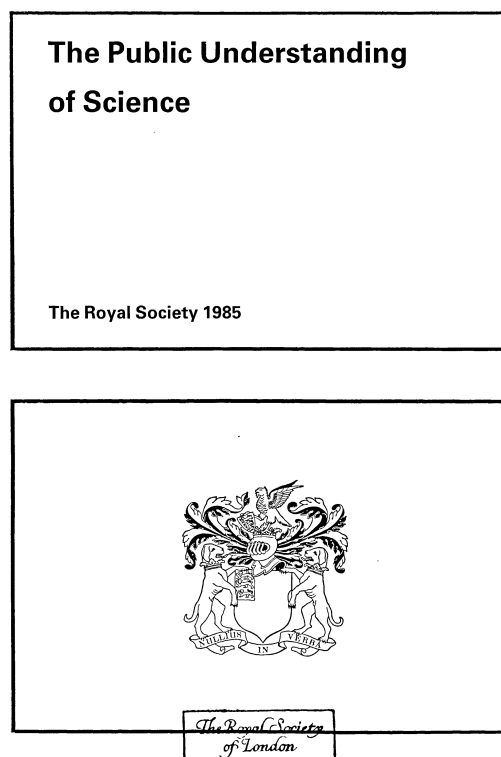


Figure 1.1 – Frontispiece of the Bodmer report.

velopment of the scientific literacy of citizens, i.e., their understanding of science, because “better public understanding of science can be a major element in promoting national prosperity, in raising the quality of public and private decision-making and in enriching the life of the individual” ([The Royal Society, 1985](#)). In this perspective, a deeper understanding of science was expected to naturally lead to its appreciation, which would in turn foster public trust and ultimately strengthen support for scientific progress – a virtuous cycle seen as essential for the nation’s prosperity ([Sandrelli, 2023](#)). This model was supported and grounded in a rational, linear view of communication, assuming that ignorance was the root of public misunderstanding and that the solution lay in simply providing more (and better) information.

One of the main outcomes of the Bodmer Report was the foundation of the UK Committee on the Public Understanding of Science (CoPUS), born within the collaboration of the Royal Society, the Royal Institution, and the British Science Association. It aimed to make scientific advances accessible to non-expert ([Bowater and Yeoman, 2012](#)) and providing funding for a wide range of initiatives ([Quin, 1991](#)). In doing so, the Committee played a key role in boosting the visibility of science journalism, science centers, festivals, and popular science books ([Stilgoe and Sykes, 2009](#)). Through these efforts, CoPUS helped

raise the visibility of science in everyday life, contributing to what some later referred to critically as the “PUS industry” (Pitrelli, 2003).

Underlying this institutional momentum was a specific vision of science communication. The PUS paradigm has its foundation on the communication model known as the *deficit model* (Wynne, 1991), i.e., a sociological construct that sees the public as having a “knowledge deficit” on scientific topics, made worse by inadequate and sensationalist media coverage. As a consequence, a feeling of irrational fears spread easily among the public, worsening the attitude through science (Bucchi and Trench, 2014). From this hypothesis arises the idea that scientists have to fill this gap by giving more information, to change attitudes towards science (Short, 2013).

Throughout the second half of the twentieth century, this top-down, one-way approach (represented in Figure 1.2) shaped much of institutional science communication. It portrayed the public as a passive, uniform audience that had to be educated by experts, i.e., assuming “public deficiency, but scientific sufficiency” (Gross, 1994). Over time, this vision came to be described as paternalistic (Hilgartner, 1990) – a model in which the authority of science and scientists was assumed to be superior, while citizens were regarded as ignorant and incapable of forming independent, valid opinions about science-related issues.

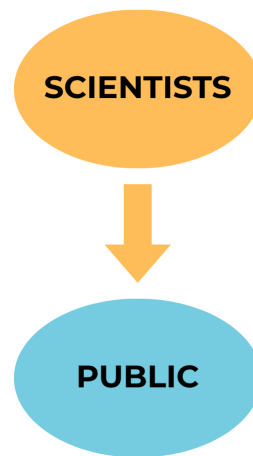


Figure 1.2 – Representation of top-down approach of the deficit model.

1.2 The crisis of the deficit model

The Bodmer report had a huge impact on science communication, not only in the UK but in all Europe. It legitimized popularizing science (Miller, 2001), inviting scientists to communicate their discipline. As stated in the Summary of the document:

Scientists must learn to communicate with the public, be willing to do so, and indeed consider it their duty to do so.

Following the Bodmer Report's call, UK policy evolved rapidly to institutionalize science communication as a national priority. An important milestone was the White Paper "Realising Our Potential" ([Her Majesty's Stationery Office, 1993](#)), which recognized the role of science in wealth creation and quality of life ([Bailey, 2024](#)). The main output of the White Paper was the mobilization of the UK's scientific community, with the institution of PUS schemes by Britain's research councils for its own communities ([Miller, 2001](#)).

However, the economic and academic efforts proved to be unsuccessful in enhancing PUS, as shown by two surveys commissioned by the CoPUS in 1988 and 1996. In the first one, [Durant et al. \(1989\)](#) assessed that 10% or less of those questioned were scientifically literate, where scientific literacy was measured through multiple-choice questions on various scientific topics regarded as fundamental. Beyond the question of what should or should not be considered essential knowledge in science, the results appeared to support the idea that the public had a "knowledge deficit", thereby reinforcing the dominant paradigm of one-way science communication. Ten years apart, for its tenth anniversary, CoPUS commissioned a follow-up survey, but it revealed little – if any – progress in public scientific literacy ([Miller, 2001](#)). These findings highlighted the limitations of the deficit model; the assumption that the mere increasing public's exposure to scientific information would lead to improved understanding and positive attitudes was starting to show signs of strain.

A critical turning point was represented by the crisis of BSE (Bovine Spongiform Encephalopathy), also known as "mad cow disease". The BSE was a neurodegenerative disease of cattle that spread in the UK in the 1990s. Initially downplayed by UK authorities, the revelation in 1996 that the disease could be transmitted to humans triggered a profound crisis of trust in scientific and governmental institutions. Its origin was traced back to the feed given to cattle, which included infected slaughterhouse waste. These aspects led to the mass media representation of BSE as a man-made disaster caused by profit-driven negligence and the violation of natural boundaries (herbivores forced to become carnivores). As [Sturloni \(2003\)](#) highlights, the BSE case brought to light deeply rooted societal fears and anxieties connected to a theme that permeates everyday life (the risk associated with food) undermining traditional top-down risk communication models used so far. Despite reassurances from experts, the public perceived the risk as unacceptable, not simply due to scientific uncertainty, but because it concerned the ethical, human, and cultural values of communities.

In all Europe, the BSE crisis exposed the limitations of the deficit model, which assumed that simply providing more information to citizens would improve their attitudes toward science. The superficiality was to ignore their lay knowledge, i.e., the situated and contextual knowledge that individuals develop through direct experience and interaction with their social environment, including ethical, environmental, and cultural aspects that hold meaning within the context in which the person operates ([Wynne, 1992](#)). Therefore,

scholars began to question the assumptions of the deficit model (Wynne, 1991), arguing that greater attention should be paid to people's values, experiences, and contextual reasoning, rather than just labeling them as "ignorant". In the words of Wynne (1995):

Public ignorance is often a rational response to the ways in which scientific knowledge is institutionally produced and communicated.

In Wynne's perspective, lay people actively negotiate meaning rather than passively receive and accept facts. Therefore, the PUS model has to acknowledge that understanding science involves not only cognitive comprehension, but also trust, relevance, and social context (Wynne, 1992). These perspectives gave rise to the conceptualization of the "context model", which shifted the focus to the lay, local knowledge, and personal interest of the public in the process of understanding science. In the words of Ziman (1992):

(...) scientific knowledge is not received impersonally as the product of disembodied expertise, but comes as part of life, among real people, with real interests, in a real world.

Science communication was no longer intended as a solely transfer of content from experts to the passive public, but as dialogue and mutual acknowledgment and negotiation of values, needs, and personal interests. Science communicators need to be "aware of the nature and existing knowledge of the intended audience" (Miller, 2001).

Alongside the development of the context model, another approach emerged – both as its evolution and as a model on its own: the dialogue model (Borchelt and Hudson, 2008). In this framework, the public is no longer seen as a homogeneous lump, but rather as a union of "consumers, activists, government [and other groups] all of whom had their own knowledge and stance on the issues" (Bowater and Yeoman, 2012), which deserve to be considered, in a two-way communication paradigm.

The shift from the deficit model to the dialogue model was led by the "Science and Society" report drawn up by the UK House of Lords Science and Technology Committee (House of Lords Select Committee on Science and Technology, 2000), which consistently emphasized the need for dialogue between science and the public (Pitrelli, 2003). Rather than lamenting public ignorance, it proposed a "3-D" model (dialogue, debate, and discussion) and formally distanced itself from the deficit model. As Bucchi and Trench (2014) mentioned, the shift reflected in funding programs and policy documents of different countries, which gradually shifted their focus from *communication* to *dialogue*, from *science and society* to *science in society*.

As a consequence of this substantial change of approach, CoPUS closed its doors in 2002 with a final statement:

We have reached the conclusion that the top-down approach which CoPUS currently exemplifies is no longer appropriate to the wider agenda that the science communication community is now addressing. (Short, 2013)

It's the end of an important phase of science communication that paved the way to the current line of thought: Public Engagement with Science (PES). Still, as [Miller \(2001\)](#) stated "the end of the deficit model does not mean there is no knowledge deficit". As the author pointed out, there has to be a gap in scientific knowledge between scientists and members of the general public, due to the educational and professional development that the former have undergone. However, there is a profound difference between treating the public as ignorant and dismissing lay knowledge, and acknowledging this gap with respect, recognizing that the public brings valuable forms of situated and experiential understanding. Rather than being viewed with condescension or haughtiness, this diversity of knowledge should be embraced as a resource for dialogue and mutual learning.

1.3 Public Engagement with Science

On October 4, 2002, the renowned American journal *Science* published a short article titled "From PUS to PEST" (Figure 1.3). In less than half a column, it summarized the shift in approach that would guide science communication in the decades to come. As already anticipated by the changes within the research community discussed in previous sections, the PUS paradigm was no longer considered appropriate. A more inclusive and participatory model was needed – one called Public Engagement in Science (and Technology). Indeed, as the article states, "it is no longer enough for science communicators simply to educate the public, (...) they must put their heads above the parapet on controversial issues" ([Holden, 2002](#)).

This marked the end of classical science popularization as traditionally conceived and of the deficit model. The scientific community began opening up to dialogue and participation, restoring to the public an active role in the construction and negotiation of meaning alongside the communicators ([Sandrelli, 2023](#)), as represented in Figure 1.4. The term *public* is replaced by its plural form, *publics*, in an effort to overcome the generalizations made up to that point ([Einsiedel, 2005](#)). In the mid-2000s, the approach underwent another change – not in name, but in preposition. The focus shifted from *engagement in* to *engagement with* science, to emphasize a clear departure from the PUS paradigm and to reaffirm the importance of democratizing science and promoting technologies that are responsive to the needs and values of the public ([Jones, 2011](#)).

From a semantic point of view, the term Public Engagement derives from politics, specifically from the concept of "participatory democracy". The Higher Education Funding Council for England defined it as "the involvement of specialists listening to, developing their understanding of, and interacting with, non-specialists" ([Short, 2013](#)). It's by definition a two-way process with the goal of generating mutual benefit ([NCCPE, 2023](#)) and mutual learning ([Nisbet and Markowitz, 2015](#)).

From PUS to PEST

Scientists in the United Kingdom have decided that the term "public understanding of science" has outlived its usefulness. In addition to making an unfortunate acronym, they say, the phrase has a condescending ring to it. So they've cooked up a more inclusive-sounding replacement: public engagement in science and technology (PEST).

Science minister David Sainsbury recently told the crowd at a science festival in Leicester that it is no longer enough for science communicators "simply to educate the public," according to *The London Financial Times*. Instead, they must get down in the trenches and interact with them—or, as Fiona Fox, director of the Science Media Centre, put it: "put their heads above the parapet on controversial issues" such as genetically modified foods.

Figure 1.3 – From PUS to PEST ([Holden, 2002](#))

Included within the scope of PES are citizen science activities, in which individuals participate as collectors or contributors of data, as well as open science initiatives, where researchers make data, protocols, publications, and other materials freely accessible online, "allowing the interested public to access not only *ready-made science* but also *science-in-the-making*" ([Bucchi and Trench, 2014](#)). The goals guiding PES activities can vary depending on the context in which they take place and the nature of the activities themselves. [Weingart et al. \(2021\)](#) conducted a content analysis of academic papers on the topic and identified five key motivations underlying PES initiatives. These include *democratization*, which involves engaging citizens to empower them to participate competently in society – thus promoting both the democratization of society and, more specifically, of science itself – and *education*, aimed at informing and educating the public by improving access to both general and specialized scientific knowledge. Other motivations concern *legitimation*, where engagement initiatives are designed to build public trust in science and support for science-related policies, and *innovation*, in which participation fosters creativity and

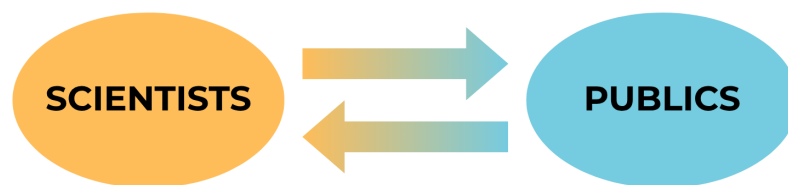


Figure 1.4 – Representation of the PES paradigm.

knowledge co-production by recognizing citizens as valuable sources of expertise whose local knowledge complements formal science. Finally, *inspiration* relates to fostering curiosity and enthusiasm for science, with the broader goal of sustaining a STEM¹-educated workforce.

In the same study, the authors also categorize the prominent criticisms or concerns of public engagement that emerge from the academic literature they analyzed. These include practical limitations, such as lack of time, resources, and manageable group sizes, and challenges related to evaluation, with few systematic assessments and difficulties in operationalizing appropriate methods. Another recurring issue is that scientists are often unprepared or insufficiently trained for public interaction, as engagement activities can conflict with traditional academic norms. A further concern is that the deficit model sometimes persists beneath the rhetoric of engagement, leading to superficial rather than genuine dialogue. Questions are also raised about intentions, since public views should meaningfully influence decision-making rather than serve as token gestures. Moreover, the diversity of audiences poses challenges, as there is no such thing as a single, homogeneous public. Finally, several factors discourage scientists from participating, including lack of incentives, recognition, and institutional support; engagement is often perceived as incompatible with academic advancement or as insufficiently valued within the scientific community. In addition, scholars such as [Rogers-Hayden and Pidgeon \(2007\)](#) reported a conceptual criticism of a return to a disguised deficit model:

It is a kind of deficit model again, a deficit model of public engagement – we’ll dialogue in order so that they understand us better and will agree with us, instead of we’ll sit them down and educate them.

The risk, therefore, is that the deficit model is only overcome in theory, while its underlying assumptions persist. To avoid the risk of falling into rhetoric, it is essential to develop forms of public engagement grounded in robust theoretical frameworks and supported by adequate evaluation methods – only then can engagement fulfill its democratic and transformative potential.

¹Science, Technology, Engineering, and Mathematics

1.4 PES in schools: games, play and learning-by-doing

When it comes to the specific context of formal education settings such as schools, science communication is referred to as *science education*, understood as the set of practices, theories, and educational policies aimed at the teaching and learning of science. It is evident that science education differs significantly from public engagement. The main difference lies in the formal context in which science education takes place, characterized by defined curricula and dedicated evaluation mechanisms. This setting does not foresee voluntary participation: students cannot choose whether or not to take part in activities offered within the school environment. This crucial aspect is an opportunity to shift the focus from *how* to effectively engage people, to *what* to deliver to students, in terms of skills, critical thinking, and scientific literacy – i.e., preparing future citizens to actively participate in complex sociol-scientific issues.

At the same time, science education practices have traditionally been more closely aligned with the PUS model rather than the PES approach, favoring a unidirectional transmission of knowledge from teacher to student – as exemplified by traditional lectures – rather than fostering active engagement. On the other hand, several studies have shown a low level of interest in the scientific content that students encounter during their school experience (Osborne and Collins, 2001; Schreiner and Sjøberg, 2004). In response to this situation, European and national policies call for a more engaging and motivating approach to science education, supported by the extensive literature that has long demonstrated the greater effectiveness of active learning and learning-by-doing compared to passive learning (Dewey, 1938; Bruner, 1977). These approaches stand in contrast to traditional frontal lectures, often described as an “information dump”, which positions the teacher as the sole source of knowledge and the student as a passive recipient. By contrast, learning-by-doing promotes a student-centered model in which learning occurs through active engagement, embodying the paradigm shift from PUS to PES in the educational context.

Focusing on learning-by-doing methodology, its theoretical foundations can be traced to the work of philosopher and educator John Dewey, who argued that education is not a matter of “telling” and “being told”, but rather an active and constructive process (Dewey, 1916). These ideas were further developed by scholars such as Jerome Bruner, who emphasized discovery learning, and Lev Vygotsky, who highlighted the social and collaborative dimensions of knowledge construction. Subsequent empirical research has consistently confirmed the benefits of experiential and active learning approaches in fostering deeper understanding.

In this context, play and games represent effective tools to promote active learning. However, before examining their educational potential, it is essential to clarify the conceptual distinction between them, two terms often used interchangeably in everyday language but rooted in different theoretical traditions.

Play refers to a spontaneous and open-ended activity that occurs across different contexts and times, permeating both individual and social life. It unfolds without clearly defined rules or constraints and represents more an attitude than a specific action. A paradigmatic example is building a sandcastle: the activity itself has no material purpose other than simple enjoyment, nor does it follow precise rules. *Game*, on the other hand, denotes a structured form of play governed by explicit rules, objectives, and boundaries.

This distinction has long been the subject of scholarly debate, with each theorist proposing a unique interpretation. In this thesis, reference is made to one of the earliest systematic attempts to define the notion of play, proposed by Caillois (2001). The author employs the term play in its broadest sense – encompassing both free and structured forms of activity – and identifies a set of characteristics that an activity must possess in order to be considered *playful*: free, separate, uncertain, unproductive, regulated, and make-believe.

Free indicates that play must not be subordinated to other activities, nor should it be mandatory. Imposing it often implies the end of the game itself (Nesti, 2017).

Separate emphasizes that play unfolds within clearly delimited boundaries of space and time. This aspect was also highlighted by Huizinga (1949), who introduced the notion of the *magic circle* – a conceptual space that circumscribes play and within which players are fully aware of their temporary detachment from everyday reality. While not entirely impermeable to ordinary life, this space is sufficiently distinct to enable experimentation free from risks (Nesti, 2017).

Uncertain underscores that the outcome of play cannot be predetermined, nor can the unfolding of events be entirely foreseen. This inherent indeterminacy opens space for innovation, creativity, and the initiative of the players.

Unproductive, i.e., “ending in a situation identical to that prevailing at the beginning of the game” (Caillois, 2001). This principle resonates with other scholars’ interpretations. For Suits (1978) play is “the voluntary attempt to overcome unnecessary obstacles”. According to Fagen in Bruner (1966), play entails a genuine “pleasure of inefficiency” which, combined with a relaxed environment and sustained excitement, makes it a space for experimentation that fosters learning.

Regulated refers to the presence of a system of norms that governs the playful experience with varying degrees of constraint. Building on this characteristic, Caillois established a continuum along which all playful activities can be placed. At one end lies *paidia*, representing the least structured and most spontaneous form of play, exemplified by the unrestrained running and improvisation of children. At the opposite end is *ludus*, characterized by a highly codified normative system, such as that of chess game, where the rules strictly determine goals and outcomes.

Make-believe means that play is accompanied by an awareness of inhabiting a reality different from the everyday one.

Returning to the theoretical framework concerning the use of play and games in edu-

cational contexts, scholars such as Bruner and Vygotsky laid the foundations for the perspective that views play as a tool for active learning and socio-cognitive development. According to [Vygotsky \(1978\)](#), play represents an ideal context for activating the zone of proximal development, the learning area in which an individual can achieve new competencies. Through the reproduction of real-life situations, play anticipates and stimulates social development, enabling children to acquire motivations, skills, and attitudes. Similarly, [Bruner \(1966\)](#) describes play as a privileged context for learning, facilitating the acquisition of skills by lowering levels of frustration and anxiety while simultaneously increasing engagement and attention. It is precisely thanks to the separate reality provided by the magic circle that learning is perceived as less threatening in relation to the risks of failure. As exemplified in Figure 1.5 (adapted from [Zarzycki, 2016](#)), errors regain their pedagogical value and become an opportunity for safe learning without risks for the player.



Figure 1.5 – Representation of the magic circle adapted by [Zarzycki \(2016\)](#).

The educational potential of play and games thus lies in their ability to operationalize the principles of learning-by-doing: knowledge is acquired through direct action and experimentation rather than through passive transmission. Play and games provide a free and safe space where ideas can be tested and reformulated.

As highlighted by [Nesti \(2017\)](#), studies on games and play are certainly not new, but over the past twenty years they have evolved into a fully-fledged discipline, *game studies*, in response to the need to understand and investigate the diffusion of video games in society. Research on games has thus acquired an unprecedented level of scientific legitimacy, placing at the center of inquiry its importance and role in society. Despite the different strands and research domains, the core focus remains the development of reflections on *gameplay*, understood as “the part of the game that absolutely requires the player’s participation” ([Mäyrä, 2008](#)), namely, the relationship between game and player when the latter actively engages with it.

Game studies are part of the broader domain of *game science*, a research area with

still blurred boundaries, characterized by a high degree of multidisciplinary, ranging from psychology, anthropology, and the humanities, to mathematics, the hard sciences and life sciences, as well as the social sciences and economics. While game studies focus more on the interaction between games and society, game science conceives games as artifacts – material or immaterial – that can be used either as tools to generate social impact through engagement, or as instruments to measure skills, competences, and abilities.

In the Italian landscape, the emerging need to consolidate these disciplines has materialized in the foundation of the *Game Science Research Center*² (GSRC, logo in Figure 1.6), an interuniversity research hub that aims at promoting, supporting and spreading the research in the field of Game Science and play. To date, more than 200 scholars have joined the Center, bringing together a wide range of professional backgrounds, including the social sciences and humanities, education and pedagogy, STEM disciplines, game design and game studies. Among the founding and partner research institutions is also the Italian National Institute for Astrophysics (INAF).

The mission of the GSRC is articulated around four main objectives. First, it promotes networking, by fostering coherence and relevance in research activities and by connecting scholars and projects from an interdisciplinary perspective. Second, it advances game science for society, encouraging initiatives within the so-called “third mission” of universities and public research bodies – such as public engagement, citizen science, and science communication – that use games as tools for participation and involvement. Third, it focuses on knowledge transfer, by developing collaborations with productive sectors that employ games either as design instruments or as products, with particular attention to entertainment, human resource selection, educational tools, strategic thinking, and workplace well-being. Finally, it works toward certification, by designing standard procedures to evaluate systems and rules used in playful activities – aimed at public engagement, education, or social change – with the ultimate goal of establishing quality certification protocols.



Figure 1.6 – The logo of Game Science Research center.

²See the website at gamescience.imtlucca.it

Chapter 2

Methodologies: game-based and playful learning

The application of games and play in educational contexts takes shape through two main methodologies: Game-Based Learning (GBL) and Playful Learning (PL).

GBL is a theoretical and experimental research area that investigates the possible applications of games for the development of innovative teaching methodologies ([Nesti, 2017](#)). The goal of GBL is to innovate teaching practice starting from an established and scientifically proven fact, namely that learning outcomes improve when the activity is associated with pleasure and enjoyment ([Pivec et al., 2003](#); [Pekrun et al., 2011](#)).

PL, on the other hand, emphasizes cultivating a playful attitude within the learning environment. It is characterized by the same principles underpinning GBL – curiosity, experimentation, creativity, and willingness to embrace failure as part of the learning process ([Whitton, 2018](#)). In this sense, GBL can be considered a more formal and goal-oriented subset of PL.

The application of these two methodologies is supported by several studies demonstrating the achievement of educational goals, i.e., the acquisition of factual knowledge through the use of games ([Setiawan and Phillipson, 2019](#); [Bayeck, 2020](#)). According to [Andreoletti et al. \(2023\)](#), this occurs through direct interaction with the components and mechanics of the game: the more the knowledge acquired during the activity is functional to the gameplay itself, the greater the learning effectiveness. In accordance with this perspective, a game that relies on quizzes and tests to progress is a valuable tool for exercising or assessing previously acquired knowledge, but it is not effective in fostering the development of new one. GBL has also been shown to be effective in fostering the development of 21st-century skills ([Qian and Clark, 2016](#)). These skills – that differ substantially from those traditionally prioritized within formal educational systems ([Squire, 2005](#)) – can be defined as a comprehensive set of knowledge, abilities, attitudes, values, and ethical dispositions aimed at enabling individuals to engage and think critically in an information-rich

and globalized society (Binkley et al., 2011).

If GBL is to provide an effective response to the growing demand for a more active and engaging science education, it is essential that the game mechanics themselves enable the emergence of knowledge and skills, in a *maieutic*¹ way, that is, by guiding learners to discover knowledge through structured questioning and experimentation rather than direct transmission. For instance, taking into account the tabletop game *Terraforming Mars*², players are not simply told that a breathable atmosphere requires a certain concentration of oxygen; instead, they must increase the oxygen track to at least 14% in order to make the planet habitable. The scientific knowledge is thus embedded in the mechanics of the game and acquired through play, rather than through rote memorization.

This becomes even more significant when the objective is not the acquisition of factual knowledge but rather the understanding of the methodologies and processes of science, making visible through gameplay mechanics the often-overlooked dimensions of scientific knowledge construction and restoring its complexity as a human endeavor. This perspective is further reinforced by the analogy between play and research (Toniolo et al., 2023). Indeed, the properties of playful activities described in Section 1.4 closely parallel those underlying the free experimentation that characterizes scientific inquiry. Researchers operate within a distinct reality marked by errors and failures, where outcomes are invariably uncertain. It is only through the intrinsic motivation accompanying the research process that scientists are able to enter the same state of flow that is typically associated with gameplay. Furthermore, the social interaction occurring during a board game session effectively simulates the collaboration and competition that characterize scientific research in its essence as a human enterprise. Therefore, developing a GBL activity does not simply mean embedding educational content through drill-and-practice methods similar to worksheets, or emphasizing rote memorization of facts (Squire, 2003). Rather, it entails designing a game in which the mechanics themselves support the acquisition of knowledge and skills through direct experience, employing elements of game design that are grounded in learning theories.

In the context of science education, some examples can be retraced in the literature. For instance, *Diamond: The Game* was developed to encourage secondary school students to explore scientific careers and experiences (Murray et al., 2022); Cardinot and Fairfield (2022) created a board game aimed at enhancing students' understanding of astrophysics concepts and their perceptions of scientists; *Catan: Global Warming*, an expansion of the

¹The maieutic method, attributed to Socrates and described by Plato, aims to bring out the knowledge already latent in the learner's mind through dialogue and critical reflection rather than transmitting information directly.

²*Terraforming Mars* is a strategic board game designed by Jacob Fryxelius, in which players collaborate and compete to make Mars habitable by raising global parameters such as oxygen, temperature, and ocean coverage. The game is widely recognized for its scientific grounding, as many of its mechanics are inspired by real planetary science and environmental processes.

well-known commercial game, integrates mechanics that stimulate dialogue around climate change (Illingworth et al., 2019).

2.1 The pedagogical framework of GBL and PL

From a pedagogical point of view, GBL and PL can also be interpreted as a methodological declination of constructivism, an important pedagogical theory developed by Piaget (1962). According to this view, learning is not the simple transmission of knowledge but an active process of personal construction through daily interaction with people and objects. A perspective that recalls the already mentioned learning-by-doing philosophy of Dewey (1938), i.e., the ludic environment is a fertile ground for experiential learning (de Freitas, 2006).

Another fundamental aspect of play that makes it a powerful educational tool is its ability to capture and sustain attention and intrinsic motivation, features that can be explained, among others, by Flow Theory and Self-Determination Theory. Intrinsic motivation is the internal drive to engage in an activity for its own sake, motivated by enjoyment, or a sense of personal challenge. Its opposite is the extrinsic motivation, which prevails when drive comes by external rewards or pressures. For example, the action of studying is driven by intrinsic motivation when it is pursued out of genuine interest in the topic. Conversely, when it is oriented toward obtaining a reward, such as achieving a good grade, it reflects extrinsic motivation.

The Flow Theory of Csikszentmihalyi et al. (2014) describes an optimal psychological state in which an individual becomes so deeply absorbed in an activity that obstacles are experienced as surmountable. In this state, time, fatigue, and all concerns other than the activity itself fade into the background, thereby fostering intrinsic motivation. Several criteria are required for flow to occur: there must be a balance between challenges and skills (as illustrated in Figure 2.1), clear goals, immediate feedback on performance, a sense of control over the situation, altered perception of time, total concentration, and loss of self-consciousness associated with a reduction of self-criticism. All of these are aspects that manifest naturally in playful action, as emphasized by Chen (2007).

Similarly, the Self-Determination Theory proposed by Ryan and Deci (2000) identifies three basic psychological needs that sustain and maintain intrinsic motivation: autonomy, competence, and relatedness (Figure 2.2). Autonomy refers to the perception of being the author of one's actions; competence concerns the perception of being effective in one's activities (self-efficacy); relatedness encompasses the feeling of being connected to and significant for others. According to Ryan et al. (2006), a well-designed ludic environment can satisfy these needs, making the experience highly engaging and conducive to motivation, well-being, and learning.

Both theories imply the same central principle: intrinsic motivation fosters a deeper

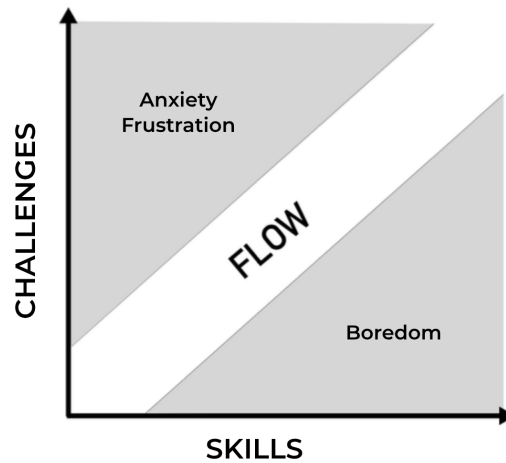


Figure 2.1 – The balance between challenge and skill is a fundamental condition for the emergence of flow. If the challenge exceeds skill, the result is anxiety and frustration; conversely, if skill exceeds challenge, the result is boredom.

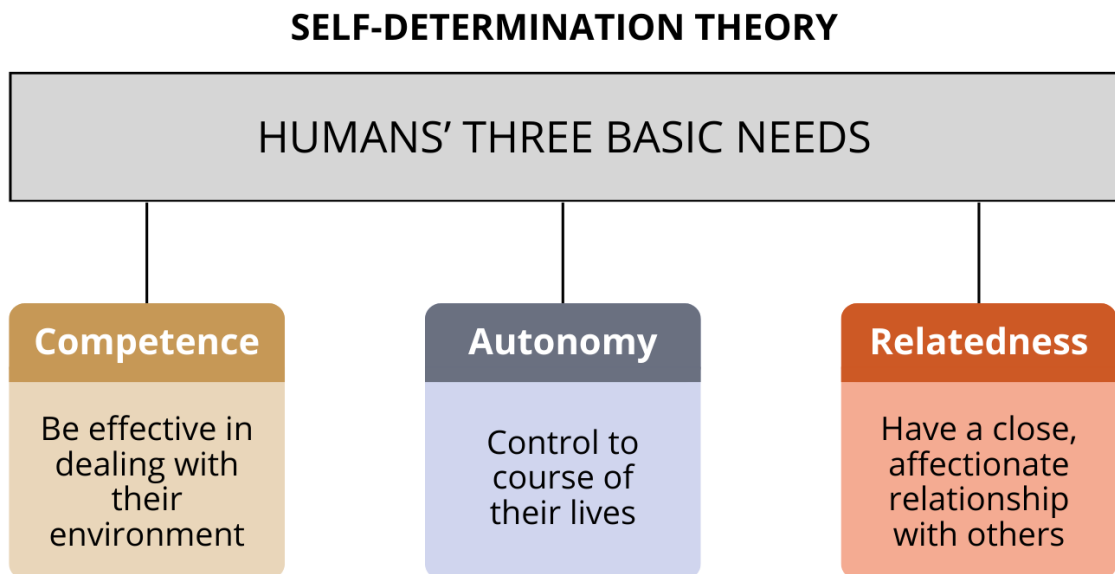


Figure 2.2 – Representation of the SDT.

elaboration of lived experiences, and consequently, more meaningful learning ([Andreoletti et al., 2023](#)). By contrast, when motivation is extrinsic, key principles of meaningful learning are lost, as learners focus on external rewards rather than on the process itself. As argued by [Pink \(2009\)](#), extrinsic motivation can provide an initial push, but in the long term it risks undermining sustained engagement, leading individuals to disengage or abandon the activity once rewards are no longer present. A clear example of this dynamic can be found in *gamification*, i.e., the use of game elements and design techniques in non-game contexts ([Deterding et al., 2011](#)). The primary goal of gamification is not learning but behavioral conditioning. In marketing, for instance, supermarket loyalty programs reward customers with game elements such as prizes, leader boards, or badges. In educational settings, gamification may appear as a multiplication table contest with rules and rewards, or a board game where students must answer questions correctly to roll the dice and move forward. In contrast, meaningful GBL aims to cultivate intrinsic motivation by leveraging the natural capacity of games to generate interest and engagement, thereby making them an integral part of the learning process.

However, it is important to bear in mind that within the educational context in which GBL and PL are applied, the distinction between intrinsic and extrinsic motivation is often blurred. Indeed, when a playful – or non-playful – activity is proposed in the school setting, students rarely have the option to decide whether to participate, and are therefore driven by extrinsic motivation. However, they can still decide whether and to what extent to engage with the activity ([Ligabue, 2020](#)), thereby opening up the possibility of becoming genuinely involved in the game.

Focusing on game, what has been discussed so far applies to both digital and analogue GBL, depending on whether video games or tabletop games are employed. Each modality presents strengths and limitations, as highlighted by [Andreoletti et al. \(2023\)](#). From a practical point of view, board games in educational contexts allow for easier modification and adaptation to specific settings, reduce the technological challenges posed by digital tools, and are more economically sustainable – an essential factor when designing activities for schools. From a methodological perspective, they offer clearer transparency in the relationships between the various elements of the game, whereas in digital games such dynamics are often hidden from the player. This characteristic is particularly relevant when games are used to simulate processes such as how scientific research works. On the other hand, tabletop games present a steeper learning curve than digital games, as players must master all rules before starting, whereas digital games introduce rules gradually. Finally, board games require greater social interaction and a sense of shared participation, what [Calleja \(2022\)](#) defines as “collective effervescence”, which promotes the joint construction of knowledge.

For these reasons, this thesis chooses to exploit GBL based on tabletop games. In an era of digital hyper connection, it is crucial to create moments of genuine social interaction

that are not mediated by screens. Board games thus provide valuable opportunities to engage in dialogue, learn to cope with both success and failure, experience collaboration and competition, and develop authentic social skills and relationships. All of this unfolds within a safe environment that allows for each individual’s freedom of expression while fostering mutual respect.

In conclusion, promoting PES through methodologies such as GBL or PL is far from a straightforward task. It implies possessing a deep understanding of the scientific domain intended to represent, so as to avoid oversimplifications or misleading analogies. At the same time, strong competencies in outreach are equally essential, since a purely technical approach risks alienating non-specialist audiences. Striking the right balance between scientific accuracy and effective communication is therefore crucial, and requires a multidisciplinary collaboration among professionals and fields of research, involving game designers, scientists, science communicators, and educators.

2.2 The role of astrophysics

Within this framework, astrophysics represents a fertile domain for the application of PES principles in educational contexts through GBL and PL.

First, astrophysics is perceived as a fascinating science that stimulates curiosity in everyone, making it easier to involve and engage people. Indeed, it has been described as a *Gateway Science* because of its capacity to “inspire both students and the general public, increasing interest in STEM and science literacy in the public” (Salimpour et al., 2021), thanks to its intrinsic multidisciplinary nature.

In parallel, astrophysics is one of the oldest sciences, carrying with it a highly diverse cultural and historical legacy that enables the introduction of elements of equity, diversity, and inclusion (Salimpour and Fitzgerald, 2024), thus bridging the gap between the sciences and the humanities. As an observational science, astrophysics has a unique, universally accessible laboratory – the sky – and various cultures around the world are innately connected to some aspect of it.

Moreover, in a centuries-old stereotypical portrayal, astronomers are often depicted as solitary individuals gazing at the stars from remote observatories. This representation, however, is far away from the reality of contemporary research, which is characterized by large international collaborations, multidisciplinary, and heterogeneous working groups. Precisely for this reason, it offers both a challenge and an excellent starting point for introducing the social and institutional dimensions of science, thereby helping to represent science as a truly human enterprise instead of only a cognitive effort.

Finally, astrophysics is commonly perceived as a purely visual science, limited to the acquisition and study of images of distant celestial objects. In reality, contemporary as-

astrophysics is inherently multisensory, as most of the Universe cannot be seen with human eyes. Telescopes capture radiation across the entire electromagnetic spectrum translating these invisible signals into images or models that make the cosmos perceptible. Astronomical images are therefore the result of a complex process of translational. From an educational standpoint, this opens to the possibility of designing multisensory and inclusive experiences that allow learners to explore the Universe through a variety of senses and expressive languages. This is particularly meaningful when considering branches of astrophysics that lack the possibility of representing their results through a stunning image. Indeed, astrophysics is commonly known to the public through the high-resolution images obtained with the latest generation of telescopes, such as the iconic image of Carina Nebula produced by the James Webb Space Telescope (Figure 2.3a).

By contrast, in radio astronomy, and in all other branches of astrophysics that observe and study wavelengths outside the visible band, the resulting images often appear less breathtaking. A classic example is the first-ever image of a black hole (Figure 2.3b), revealed in 2019 by the Event Horizon Telescope collaboration. Its release was greeted with great excitement and satisfaction by the scientific community, as it represented an important confirmation of Einstein's general theory of relativity and was undoubtedly a landmark scientific achievement – made possible through extensive international collaboration that required immense investments of time, energy, and resources from more than 60 institutions across nearly 20 countries. Yet, at first glance, the image may appear to non-expert eyes as almost blurred, falling short of the expectations of those who anticipated something as visually striking as other more famous visual representations of the Universe.

This disparity in representation provides a unique opportunity to shift the focus from viewing science and its processes in terms of results only (the Wow), to the human and technological endeavor underpinning them (the How) characterized as much by failures as by successes, by effort as by satisfaction, and by both collaboration and competition.

These principles underpin the mission of the Office of Astronomy for Education³ (OAE), established by the International Astronomical Union (IAU). The OAE coordinates a global network of national and regional offices to promote astrophysics as a vehicle for science education and as a means to foster cultural diversity, inclusion, and international collaboration. Its name reflects the core philosophy guiding the office's work: astrophysics is not conceived as a body of knowledge to be transmitted, but as a shared cultural and scientific endeavor through which broader educational and societal goals can be pursued. This vision aligns with the very essence of Public Engagement, reaffirming the idea that science education should serve society as much as it serves knowledge itself.

OAE consists of many Center and Nodes, each with different core thematic and interests. The OAE Center Italy (logo in Figure 2.4) is hosted by INAF, the National Institute for Astrophysics, and focuses on the development of person-centered learning framework.

³See the official website at astro4edu.org



(a) Image credit: [NASA et al. \(2022\)](#).



(b) Image credit: [EHT Collaboration \(2019\)](#).

Figure 2.3 – (a) Cosmic Cliffs in the Carina Nebula (NGC 3324), captured by the James Webb Space Telescope. (b) Image of the M87 black hole, elaborated from radio data obtained by the Event Horizon Telescope.



Figure 2.4 – Logo of OAE Italy.

2.3 Evaluate the effectiveness

The aim of this thesis is to address a long-standing question that practitioners and researchers face when dealing with innovative teaching methodologies: are they really effective? The answer to this question is far from straightforward, and it is certainly not the ambition of this research to provide a definitive response to such a complex issue. The concept of effectiveness, in fact, strongly depends on the domain to which it is applied, whether it concerns the learning of factual knowledge, the experimentation of processes, or the improvement of more affective dimensions such as attitude and self-efficacy.

Therefore, the research question is formulated as follows:

To what extent and through which dimensions can PES activities based on GBL be evaluated as effective in the context of astrophysics education?

To address this research question, the thesis considers two case studies based on tabletop games – PIXEL and Cosmo Hunters – both developed using a GBL approach. The two games were designed for distinct targets and with different educational objectives: PIXEL aims to foster the experience of scientific processes and dynamics, whereas Cosmo Hunters focuses on the understanding of key concepts related to astronomical observations across different wavelengths. Precisely because of these differences, the overarching research question has been contextualized for each case study, taking into account different evaluation dimensions. Before presenting these dimensions in detail, it is provided a brief overview of evaluation in the field of education.

As highlighted in [Weingart et al. \(2021\)](#) and previously discussed, evaluation is one of the most frequently mentioned critical issues in academic papers dealing with public engagement with science. According to [Bartlett \(2024\)](#), evaluation should not merely ask whether an initiative “works”, but rather what works, for whom, and under what circumstances. In this sense, evaluation becomes not only a means of measurement but a process of reflection and learning – both for participants and for practitioners.

Therefore, evaluation represents a methodological process that requires careful design and the adoption of multiple approaches reflecting the diversity of activities that can be developed in science communication and education. Broadly speaking, evaluation methods can be classified into two main categories: quantitative and qualitative. Each offers distinct advantages and limitations (summarized in Table 2.1), and the most robust evaluations often integrate both within a mixed-methods framework.

Quantitative evaluation relies on the collection of numerical data to measure changes in participants’ knowledge, skills, or attitudes. Instruments typically used in this approach include multiple-choice tests and Likert-scale questionnaires. The latter is a self-reporting instrument in which individuals respond to a series of statements by indicating the extent of their agreement using a numerical scale (e.g., from 1 = strongly disagree to 5 = strongly

Quantitative Methods	
Advantages	Disadvantages
Provide an overview of the full sample.	Rigid structure; participants cannot easily justify or explain responses.
Enable rational and objective statistical analysis.	Data collection occurs in an artificial, structured context.
Allow rapid data collection from large numbers of participants.	Designing valid quantitative instruments can be difficult.
Easy to replicate across studies.	Require a larger number of participants for statistical reliability.
Qualitative Methods	
Advantages	Disadvantages
Require fewer participants than quantitative methods.	Time-consuming and potentially expensive.
Yield understanding of <i>why</i> things occurred and the underlying processes.	Subject to bias; greater difficulty in proving reliability and validity.
Allow exposure to subtleties and complexities.	Hard to gain representative samples.
Flexible and adaptable throughout the data collection process.	Results are context-dependent and difficult to generalize.

Table 2.1 – Comparison between qualitative and quantitative evaluation methods adapted from [Bartlett \(2024\)](#).

agree). The advantage of these methods is that they provide clear metrics allowing for comparability, replicability, and generalization across contexts. However, such approaches tend to simplify complex learning processes and may overlook the nuances of participants' experiences. Moreover, quantitative instruments usually require large participant samples to ensure statistical reliability and validity.

Qualitative evaluation, by contrast, focuses on how and why learning or engagement occurs, exploring participants' perceptions through observation and interaction. Common tools include semi-structured interviews, focus groups, reflective journals, and thematic analyses of open-ended responses. The advantage of qualitative approaches is that they provide rich interpretative insights and contextual depth, allowing evaluators to uncover the mechanisms behind observable outcomes. However, they are more context-dependent and subjective, which makes generalization more difficult and findings potentially prone to bias.

In many cases, combining quantitative and qualitative evidence – known as a mixed-methods approach – is the most effective way to evaluate PES initiatives and education

activities. This combination enables evaluators to measure what changes (quantitative data) while also explaining how and why those changes occur (qualitative data).

Developing an evaluation strategy – of any type – is the first step toward designing effective and meaningful educational experiences. The closer the evaluation is integrated into the activity’s conception, rather than added retrospectively, the more effectively it witnesses learning and engagement. In this way, evaluation is not a detached measurement exercise but an integral component of thoughtful, evidence-based educational design.

In the real world, these considerations take shape in the persistent gap between practitioners and theorists in the field of Public Engagement, largely due to their different background: PES researchers often come from social science backgrounds, while practitioners typically have a STEM education (Bailey, 2024). This divergence highlights the need for a common ground where theoretical frameworks and practical experience can converge to create more coherent and effective approaches to science communication and education.

While the need for more active and engaging approaches to science education has been widely acknowledged, evaluating their effectiveness requires moving beyond engagement. In this thesis, GBL and PL are treated as methodologies that enable specific educational effects to emerge. For this reason, the evaluation focuses on constructs that capture how learners relate to science and scientific knowledge, considering the teaching and learning of Nature of Science, as well as students’ attitudes towards astronomy and their self-efficacy, as key dimensions through which the effectiveness of PES activities based on GBL can be meaningfully assessed. The following sections introduce the theoretical frameworks of these dimensions.

2.3.1 Nature of Science

In Section 2.1 it was extensively discussed how games, through their mechanics, can represent effective tools not only for the acquisition of factual knowledge, but also for the development of soft skills. Thanks to the safe environment they provide, games allow learners to experiment with situations and processes that would be more intimidating or difficult to experience in real-life contexts. Moreover, within the field of astrophysics education, GBL represents a particularly fertile ground for the application of PES principles. This means developing teaching methodologies that go beyond the mere transmission of scientific content, and instead promote learning and understanding of how such knowledge is constructed. After all, “the aim of teaching is not to make information accessible, but to teach students how to transform accessible information into usable knowledge” (Savia, 2016).

According to a long-standing strand of theoretical and pedagogical research, this can be achieved through the teaching and understanding of the Nature of Science (NOS), defined as the values and assumptions inherent to scientific knowledge (such as tentativeness,

parsimony, empirically based, amoral, etc.), including methodologies, aims and values, practices, social interactions, professional activities, etc. (Lederman and Zeidler, 1987). Including NOS teaching in science education first and foremost entails defining what science is and which of its foundational principles should be taught in school curricula. As underlined by Lederman (1992), research on this topic began more than a century ago and has, over the years, undergone ramifications and evolutions that led it into a central theme within the science education field.

The rationale for integrating NOS into school curricula lies in its potential to foster a form of scientific literacy that goes beyond the mere – and often memorized – acquisition of content. Understanding the Nature of Science equips learners with a range of competencies essential for both personal and civic life. From a broader perspective, it enhances democratic citizenship, enabling students to make sense of and make informed decisions on socio-scientific issues (Irzik and Nola, 2014). Through this lens, learners develop the ability to critically evaluate scientific claims presented in the media and to recognize the inherent uncertainties and limitations of scientific knowledge. As summarized by Allchin (2011), “students should develop an understanding of how science works with the goal of interpreting the reliability of scientific claims in personal and public decision making”. This means also allowing learners to understand the mechanisms through which scientific knowledge is produced and validated, moving beyond naïve and static conceptions of science such as the so-called myth of absolute truth, which assumes a fixed, recipe-like method underlying all scientific practice, or the laws-are-mature-theories fable, which portrays scientific knowledge as infallible and immutable (Lederman, 1992). Moreover, NOS education helps students to grasp science as a human enterprise (Irzik and Nola, 2014), counteracting the stereotypical image of the scientist as an isolated genius and instead emphasizing collaboration, diversity, and the collective nature of scientific inquiry. It also enables learners to appreciate science as an integral part of contemporary culture (Driver et al., 1996) and to recognize its broader value and relevance in today’s world (Irzik and Nola, 2014). Among all these transversal skills, understanding NOS contributes to improving the learning of scientific content itself, as it strengthens critical thinking skills and supports the ability to apply concepts in novel contexts (Driver et al., 1996).

Moreover, fostering learning and understanding of the Nature of Science makes it possible to highlight the cognitive and social processes that lead to the construction of scientific knowledge, without overlooking its inevitable limitations. A balanced and reflective approach allows to move beyond a “demystifying” narrative and to recognize that science, while subject to limits, biases, and revisions, possesses methodological and normative tools to self-correct, maintain the reliability of its claims, and distinguish itself from other forms of knowledge. As emphasized by Abd-El-Khalick and Lederman (2023), the COVID-19 pandemic has made this need particularly evident: revisions to recommendations regarding mask use and modes of transmission have illustrated the tentativeness of scientific knowl-

edge; controversies over vaccines have highlighted the importance of understanding the plurality of scientific methods. Therefore, a comprehensive and well-informed understanding of NOS fosters a balanced critical attitude toward science, encouraging greater engagement with it and strengthening public trust.

Despite numerous efforts in both research and educational policy, effectively teaching it in classroom settings remains a major challenge. This is evident from numerous studies on NOS perceptions among both students and teachers (see, for example, the reviews by [Lederman, 1992](#) and [Abd-El-Khalick and Lederman, 2023](#)), which highlight a naïve understanding of NOS. Such conceptions often include the perception of the scientific method as a fixed sequence of steps leading to the creation of immutable laws, while almost entirely excluding the social aspects and the view of science as a human enterprise. This perception is reinforced by the under representation of social-institutional aspects in school curricula ([Kaya and Erduran, 2016](#)), in science education research ([Cheung and Erduran, 2023](#)), and even in scientists' own accounts ([Wu and Erduran, 2024](#)). The Italian context is no exception. [Caramaschi et al. \(2022\)](#) analyzed the physics guidelines of Italian scientific high schools (*liceo scientifico*) provided by the government. Their findings indicate a focus primarily on the cognitive-epistemic aspects of NOS, with a marked under representation of the social-institutional dimensions.

As emphasized by [Abd-El-Khalick and Lederman \(2023\)](#) and [Dagher and Erduran \(2016\)](#), it is necessary to provide teachers and schools with targeted tools and resources capable of addressing these gaps in both content and pedagogy. This requires not only the development of instructional materials that explicitly integrate social-institutional aspects of NOS alongside cognitive-epistemic ones, but also sustained professional development opportunities that equip educators with the confidence, knowledge, and strategies to teach NOS in a comprehensive and contextually relevant manner.

From a conceptual point of view, over the years several approaches have been theorized to describe the main aspects of NOS, the most prominent of which are the consensus view ([Lederman, 2004](#)) and the Family Resemblance Approach ([Irzik and Nola, 2011, 2014, 2023](#)). The consensus view has traditionally emphasized a restricted set of aspects on which there is wide consensus, such as the empirical basis of science, tentativeness, subjectivity, creativity, and the role of social and cultural embeddedness. While this approach has been widely adopted due to its clarity and relative ease of curricular integration, it has also been criticized for presenting an overly narrow view of science.

In contrast, the Family Resemblance Approach (FRA), introduced by [Irzik and Nola \(2011\)](#) and later developed into the Reconceptualized Family Resemblance Approach (RFN) ([Erduran and Dagher, 2014; Erduran et al., 2019](#)), offers a broader, more comprehensive, and holistic framework. It integrates two systems: the cognitive-epistemic (including aims and values, methods and methodological rules, scientific practices, and scientific knowledge) and the social-institutional (including professional activities, scientific ethos, social

certification and dissemination, social values of science, social organizations and interactions, political power structures, and financial systems). See Table 2.2 for the description of categories.

Category	Description
Cognitive-epistemic system	
Aims and values	Cognitive and epistemic objectives of science, such as accuracy and objectivity.
Methods	Manipulative as well as non-manipulative techniques that underpin scientific investigations.
Practices	Set of epistemic and cognitive practices that lead to scientific knowledge through social certification.
Knowledge	Theories, laws, and explanations that underpin the outcomes of scientific inquiry.
Social-institutional system	
Social certification and dissemination	Social mechanisms through which scientists review, evaluate, and validate scientific knowledge, for instance through peer review systems of journals.
Scientific ethos	Norms that scientists employ in their work as well as in interactions with colleagues.
Social values	Values such as freedom, respect for the environment, and social utility.
Professional activities	How scientists engage in professional settings such as attending conferences and conducting publication reviews.
Social organisations and interactions	How science is arranged in institutional settings such as universities and research institutes.
Financial systems	Underlying financial dimensions of science, including funding mechanisms.
Political power structures	Dynamics of power that exist between scientists and within scientific cultures.

Table 2.2 – Categories of the FRA related to the Nature of Science, taken from [Kaya et al. \(2019\)](#) and originally adapted from [Erduran and Dagher \(2014\)](#).

The framework exploited for this study is the RFN, which is supported by several motivations. First, it restores complexity and diversity to the representation of scientific research, moving beyond simplified accounts that risk reducing science to a uniform set of procedures, such as the myth of a single, recipe-like scientific method, recognizing instead the plurality of methodologies and practices that characterize real scientific work. Second, it enables particular emphasis on the role of people in research, which is especially relevant in astrophysics: since the discipline often deals with distant and abstract objects, there is the risk of perceiving it as detached from everyday reality, whereas highlighting the human

dimension makes it more tangible and relatable. Third, the framework provides an holistic vision of science, systematically integrating epistemic and social-institutional aspects. Unlike the consensus view, where the latter are only marginally addressed, the RFN makes them explicit, thus acknowledging the collaborative, institutional, and societal dimensions of science. Finally, as [Dagher and Erduran \(2016\)](#) argue, weaving social-institutional aspects into cognitive-epistemic ones provides a fairer and more engaging representation of science.

In this regard, it is worth noting that the paradigm of the Public Understanding of Science referred almost exclusively to a single dimension of the Nature of Science (i.e., knowledge) assuming that promoting scientific literacy required increasing the public's factual understanding of scientific concepts. The Public Engagement model, instead, emphasizes all dimensions encompassed by the NOS framework, acknowledging that science is not only a body of knowledge but also a set of epistemic practices and social-institutional processes.

2.3.2 Attitudes and self-efficacy in astronomy

In addition to the understanding of Nature of Science, PES activities in educational contexts aim to influence how learners perceive science and their own relationship with it. For this reason, affective constructs play a central role in evaluating the impact of GBL-based interventions. In the context of astrophysics education, attitudes towards astronomy and self-efficacy emerge as particularly relevant dimensions, as they are closely linked to students' engagement with the discipline.

The construct of attitudes towards science has long been the subject of study within the field of science education. Providing a definition is by no means straightforward, given its multiple facets and the variety of approaches it entails. In general terms, it refers to students' affective, motivational, and value-laden dispositions towards science. Several studies have demonstrated that learning and career choice in a particular discipline can be influenced by attitudes towards that discipline ([Koballa and Glynn, 2013](#)), thus assuming a crucial role in fostering public engagement with science.

Astrophysics, as discussed in Section 2.2, constitutes a distinctive context that can strongly stimulate curiosity and engagement while also serving as a gateway science. However, empirical evidence indicates that students' positive attitudes towards astronomy often decline with age and progression through secondary school, particularly when the subject is absorbed into physics curricula rather than treated as an independent area of study ([Galano et al., 2024](#)). For this reason, validated quantitative instruments have been developed over the years to capture attitudes towards astronomy, such as the Astronomy and Science Student Attitudes (ASSA) survey of [Bartlett et al. \(2018\)](#).

Self-efficacy is a construct that denotes students' confidence in their ability to engage

with and succeed in learning or other task. In the context of public engagement with science, it shapes whether individuals feel capable of meaningfully interacting with scientific ideas and practices, moving beyond the prejudice that science is only for so-called “geniuses” (whatever that may mean). In the context of astronomy, self-efficacy is also closely intertwined with attitudes: according to [Galano et al. \(2024\)](#), students with more positive attitudes towards astronomy also tend to display higher levels of self-efficacy as well as a stronger disciplinary identity. Conversely, profiles characterized by negative or indifferent attitudes exhibit lower self-efficacy and achieve weaker results in knowledge tests. According to [Britner and Pajares \(2006\)](#), mastery experiences such as school achievements are the strongest predictor of self-efficacy, while emotional states such as anxiety and stress negatively affect it. In this context, the game-based experience provides a learning environment that promotes positive emotional states in a risk-free setting, where failure is perceived as more manageable.

2.4 The research questions

As introduced in Section 2.3, the two case studies presented in this thesis are guided by a common overarching research question: to what extent and through which dimensions can PES activities based on GBL be evaluated as effective in the context of astrophysics education?

Given the diversity of educational goals and target age characterizing the two board games under study, this general aim has been tailored through distinct research questions and evaluation dimensions for each case. This section introduces briefly each research question and the rationale behind them. For the detailed description of the board games see Chapter 3.

PIXEL - Picture (of) the Universe was designed to foster students’ engagement with the practices and dynamics of scientific research, with a specific focus on its social-institutional aspects as a human endeavor. For this reason, the main research question guiding this case study concerns the extent to which participation in the game supports students’ understanding of the Nature of Science. In particular, the studies presented in Chapter 4, investigate whether and how this happens within a complex GBL program or a single-session.

In addition, the PIXEL evaluation also considers affective dimensions of engagement, namely students’ attitudes toward astronomy and their perceived self-efficacy within the context of a single gameplay session. Attitudes were assessed in order to investigate whether a GBL-based approach can support the fostering of more positive attitudes toward astronomy in an age group that, according to the literature, typically experiences a marked decline in interest in science. Similarly, potential changes in self-efficacy were examined based on the assumption that the safe and risk-free environment created by the game foster

positive emotional states, which are generally associated with increased levels of perceived self-efficacy. Both dimensions were evaluated within the context of a comparative study (described in detail in Section 4.2) involving an out-of-school-time activity, with the aim of controlling for potential bias related to the novelty of an innovative methodology such as GBL.

The second board game, Cosmo Hunters, was developed with a different educational focus, centered on supporting students' understanding of the processes of astronomical observation across different wavelengths, thereby foregrounding the dimensions of cognitive-epistemic practice. Accordingly, the research question addresses the extent to which the game supports learning outcomes related to the topic.

The differentiation between the two case studies reflects a deliberate methodological choice, that narrows the wide concept of effectiveness to the specific context of the two board games.

Chapter 3

Case studies overview and rationale

This thesis evaluates the effectiveness of two cases studies: PIXEL - Picture (of) the Universe and Cosmo Hunters, which were both developed to bring PES principles into formal educational settings through GBL. Since they differ in target age, duration, and educational goals, they support different evaluation designs and enable the investigation of complementary dimensions of effectiveness. The chapter therefore provides an overview of each game, followed by a comparative discussion that highlights their key analogies and differences and anticipates how these features inform the evaluation strategies adopted.

3.1 PIXEL - Picture (of) the Universe

PIXEL - Picture (of) the Universe (Figure 3.1) is a competitive board game that simulates the world of astrophysical research, created by INAF in collaboration with the GSRC. The game is designed for 3 - 5 players aged 12 and above, with an average duration of 60 - 90 minutes.

In PIXEL, players have to manage astrophysics research centers and conduct observation of different celestial objects. Each research center is equipped with a team of researchers who are employed to carry out certain actions, a ground-based telescope, and resources that can be invested to hire new personnel, specialize staff, or upgrade the facilities. In addition, the game features a shared space telescope, which players can access by booking observation time on a first-come, first-served basis.

The goal of the game is to accumulate Scientific Prestige (SP) points. At the end of the game, the player with the highest number of SP points is declared the winner. Scientific Prestige is gained by completing projects, such as conducting observations, achieving specific upgrades of the research center, or publishing scientific results. Variability and uncertainty are introduced through random events, which can have either positive or negative effects and simulate external and often unpredictable factors that influence scientific research.



Figure 3.1 – The board game PIXEL - Picture (of) the Universe.

Observations of celestial objects represent the core of the game. Players use either their own ground-based telescope or the shared space one to observe one or more portions of celestial objects. An observation consists of placing a tile representing a section of a specific object, with a degree of resolution enabled by the level of the telescope being used (Figure 3.2). The more advanced the instrument, the higher the resolution of the tiles that can be observed, thus granting access to more detailed information.

The game was developed from the playful activity “Talkin’ about resolution” ([Varano and Ricciardi, 2018](#); [Ricciardi et al., 2019](#)), aimed at children aged 8 to 12. The goal is to use Quercetti® pegs (as an analogy for pixels) to reconstruct images at different resolutions, introducing basic astrophysical concepts such as spatial resolution of digital images ([Toniolo, 2020](#)). As seen in Figure 3.3, children directly experience how the number and size of pegs affect image quality, as well as the time required for completion: the higher the resolution, the more pegs are needed, and the longer the required time.

The choice of addressing the theme of resolution stems from its key role in astrophysics: it determines not only which details can be observed, but also the costs, times, and instruments required for research, which vary depending on the goal. For example, in the study of exoplanet transits in front of their stars, high spatial resolution is not necessary; in-

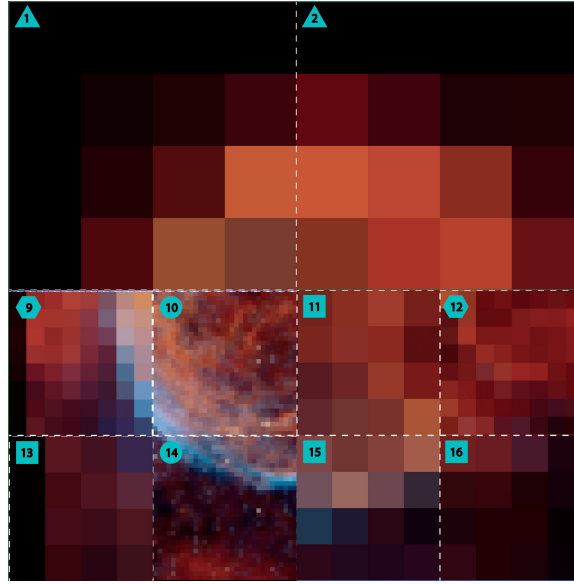


Figure 3.2 – The image of a celestial object is built using tiles of different dimensions and resolutions.

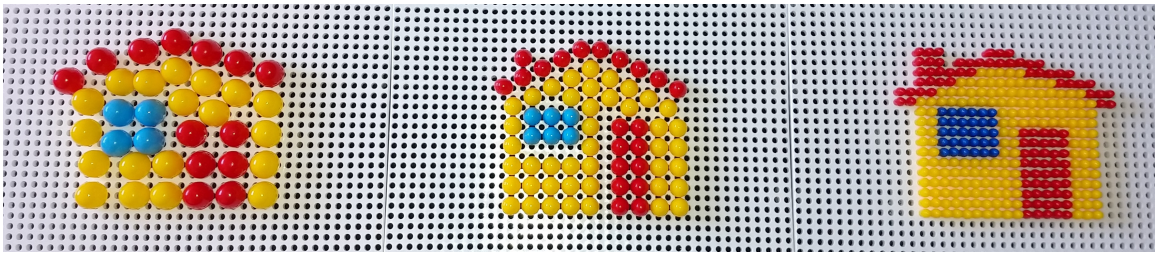
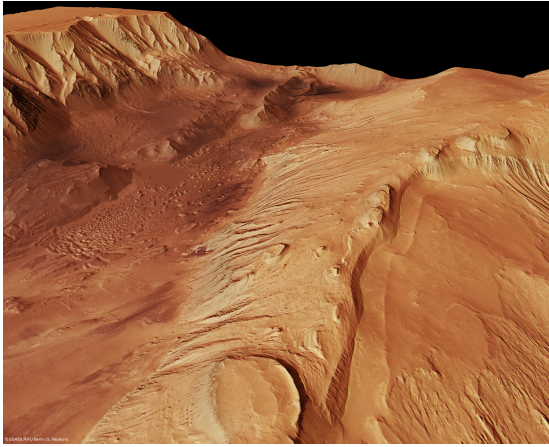


Figure 3.3 – Images of a house made with pegs of decreasing size (and therefore increasing resolution).

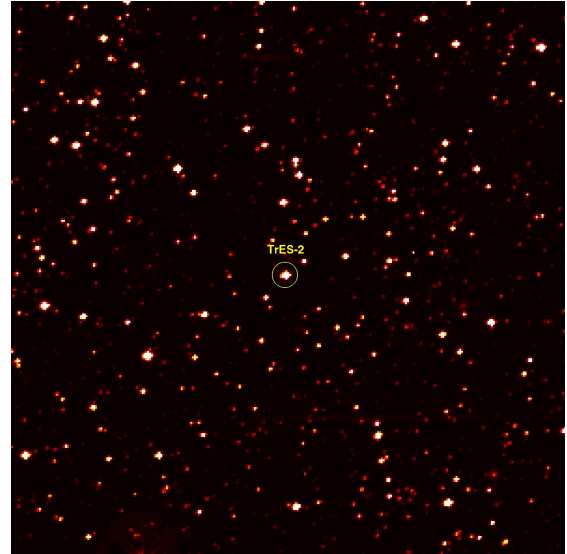
stead, many low-resolution images taken rapidly are required to detect changes in stellar brightness. Conversely, analyzing the surface of a planet in the Solar System requires high resolution to distinguish fine geological structures (Figure 3.4).

From a pedagogical perspective, resolution is a circumscribed and intuitive concept, easily relatable to children’s everyday experiences, particularly regarding the different resolutions of photos and videos on the web depending on signal intensity. These characteristics make it particularly well suited to a *micro-learning* approach, which allows the development of disciplinary and transversal skills starting from a single conceptual core.

PIXEL was developed so as to maintain the same disciplinary focus on resolution as “Talkin’ about resolution”, while adding greater emphasis on the processes of constructing scientific knowledge and on the principles of the Nature of Science, especially those related to the internal dynamics of the scientific community. Indeed, each game mechanic has been carefully designed to simulate, as faithfully as possible, the dynamics and processes that characterize contemporary astrophysics research. See Table A.1 in the Appendix for a



(a) Image credit: ESA/DLR/FU Berlin (G. Neukum)



(b) Image credit: NASA/Ames/JPL-Caltech

Figure 3.4 – (a) Image of Candor Chasma, a valley on Mars. The ground resolution is approximately 20 m/pixel. (b) Image captured by the Kepler space telescope, designed to detect exoplanets. The images are intentionally slightly defocused, reducing the number of saturated stars and improving photometric precision.

detailed mapping between the game mechanics and dynamics and the specific aspects of astrophysical research they are intended to represent.

3.1.1 Game design process

The development of PIXEL began in 2021 within a small working group that brought together astrophysicists, experts in public engagement, and professional game designers. The sharing of diverse experiences and expertise was the key ingredient in creating a game that is entertaining and engaging, while at the same time faithfully representing the dynamics of the scientific community it portrays.

After the brainstorming phase to determine which concepts should be included in the game, a process of selection and refinement followed. For instance, from the outset, there was a clear intention to retain resolution as the sole disciplinary concept, while ideas such as multi-wavelength observations and the construction of all-sky surveys were discarded to avoid overcomplicating the gameplay.

Another milestone of the entire project was the principle that the game should not require prior knowledge in order to participate or to win, in line with the considerations discussed in Chapter 2. This meant avoiding any kind of quiz or test, while at the same time making available scientific content at varying levels of depth. A clear example of this is the Publication cards (Figure 3.5a), which report details about the represented object but

whose consultation is not mandatory for the gameplay. In this way, those who wish to learn more can do so. Similarly, in the Objective cards (Figure 3.5b), the observational target is indicated both through its scientific designation and by the precise location of which tiles must be revealed, without assuming that players already know where the phenomenon in question is located.

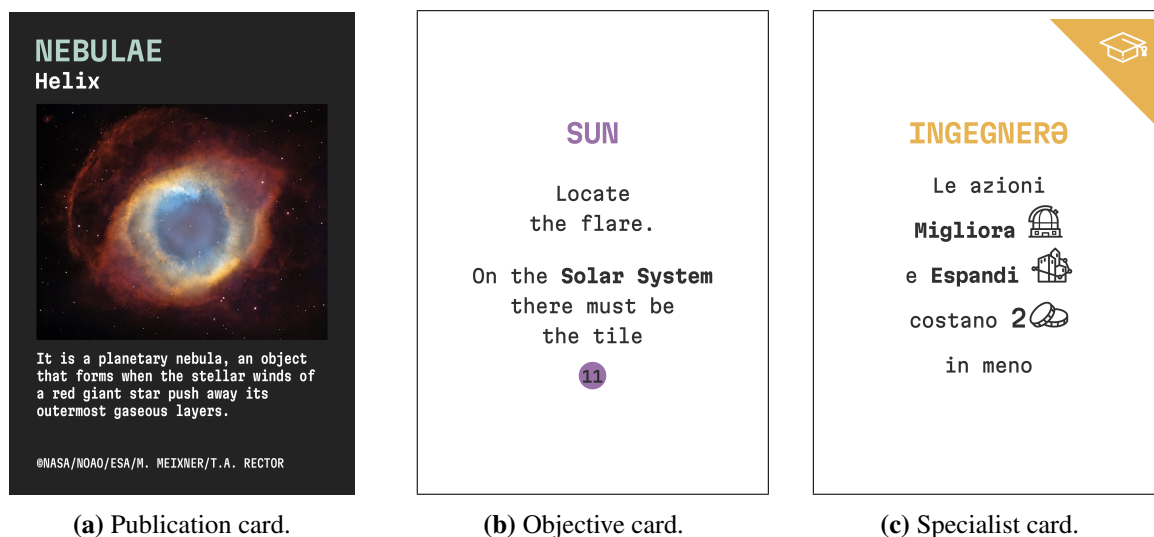


Figure 3.5 – Publication cards (left) and objective cards (center) display further information about astronomical objects. Right: example of a card where the use of the “ø” symbol makes the term *engineer* gender-neutral rather than masculine.

These considerations ultimately led to the decision to develop a board game in the *eu-rogame* genre. This type of game is highly strategic and emphasizes mechanics over theme or conflict simulation. It is also a complex system with multiple winning strategies and unlimited combinations of play, which made it possible to introduce a variety of game elements and mechanics that represent the processes and dynamics of astrophysical research, while ensuring replayability.

During the process, particular attention was devoted to adopting the most gender-neutral language possible, with the idea of representing science as a gender neutral discipline, not contributing to gender-biased narration. This proved challenging in Italian, where most nouns have masculine or feminine forms and the masculine plural is conventionally used for collective and generic references. For instance, the English term *researcher* translates as *ricercatore* (masculine) or *ricercatrice* (feminine), while *researchers* is typically translated as *ricercatori* (masculine plural). To address this issue, collaboration was established with Dr. Vera Gheno, a sociolinguist specialized in inclusive language. Whenever possible, gendered expressions were minimized by using alternative yet natural paraphrases to easily overcome the issue. Elsewhere (e.g., cards, where short texts were mandatory, such as the one on Figure 3.5c), the symbol “ø” (schwa) was adopted as a proposed solution for

neutralizing gendered terms and avoiding binary references (Ghenò, 2024) – for example, using “ricercatorə” to indicate both singular and plural forms of *researcher*.

The finalization of the first PIXEL prototype was completed in 2021, marking the beginning of the playtest phase aimed at evaluating the functionality of the mechanics and the game itself. The first gameplay sessions took place in 2021 at the PLAY Festival in Modena and at Lucca Comics & Games, two of the main Italian events dedicated to games and play, where the prototype was presented to the general public and tested in sessions run by the authors. Given the context, the playtesters were primarily experienced adult players, whose familiarity with board games made them focus on the actual enjoyability of the game, despite a certain complexity, that could have intimidated non-expert players but was deliberately retained, as it allowed for a richer representation. Feedback was collected through direct observation and a debriefing discussion at the end of the session, then implemented following an iterative process that led to the achievement of the final prototype in 2023.

3.1.2 Evaluation of engagement and scientific relevance

After the playtest phase, a preliminary evaluation stage followed, aimed at measuring the level of engagement and the perception of the scientific content conveyed by the game after a single gameplay. The analysis involved 210 players – 80% of whom were between 16 and 19 years old – who played PIXEL between 2021 and 2022 at PLAY Festival (Figure 3.6), Lucca Comics & Games, and Rome Science Festival. The data were collected through a two-section questionnaire.



Figure 3.6 – A game session held during the Play Festival in 2022.

The first section asked participants to rate, on a Likert scale from 1 (not at all) to 7 (very much), a set of statements related to PIXEL concerning five positive characteristics (funny, absorbing, pleasant, interesting, and entertaining) and two negative ones (not attracting, boring), as described in Table 3.1.

Characteristics	Statements
Boring	I thought it was a boring game
Pleasant	I really enjoyed playing this game
Not attracting	This game did not capture my attention at all
Entertaining	While I was playing this game, I was thinking about how entertaining it was
Funny	The game was fun
Absorbing	I thought the game was exciting
Interesting	I would describe this game as very interesting

Table 3.1 – Characteristics evaluated through specific statements.

Figure 3.7 reports the average scores (M) for each item. Overall, the game was perceived mostly as funny and pleasant (both $M = 5.6$), interesting ($M = 5.4$), absorbing ($M = 5.2$) and entertaining ($M = 4.7$). Boring ($M = 1.9$) and not attracting ($M = 1.8$) received the lowest score. These values were used to obtain the overall engagement index, which was calculated averaging the mean scores, after inverting the two negative attributes (boring and not attracting) allowing direct comparison with positive characteristics. For instance, in the case of the attribute “boring”, a reported mean value of 1.8 was converted into 5.2. The final index resulted in an average score of 5.2 out of 7, thus indicating generally positive feedback on the entertaining nature of PIXEL.

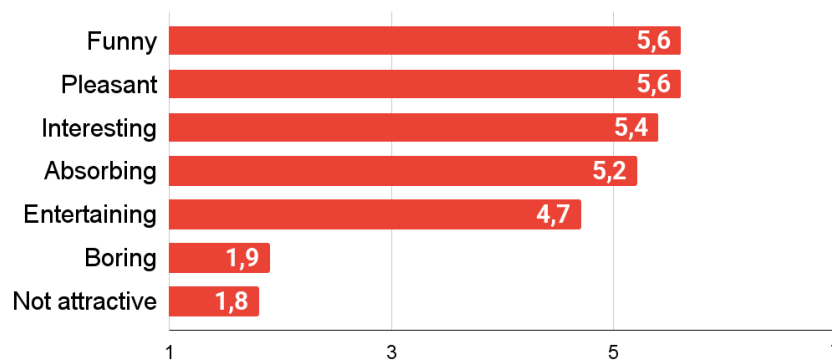


Figure 3.7 – Average scores of the PIXEL characteristics.

In the second section of the questionnaire, players were asked to indicate, from a list of topics presented in Table 3.2, the three that in their opinion emerged from the game. The

pool of topics from which to choose was constructed to include seven items directly related to the game and three unrelated ones (nos. 1, 7, 9), which served as control items.

Topics
1. Publications are the final goal of astrophysics research.
2. High resolution images require more observation time.
3. Astrophysics research is shared with the scientific community.
4. Scientific research is a collaboration of different professionalities.
5. Improvement of technologies brings to an improvement of observation resolution.
6. Publications are extremely useful for research centers.
7. Astrophysics research requires to travel a lot.
8. An important research center is one with a more scientific distinction rather than more resources.
9. Italian astrophysics research is paid less than abroad.
10. Investments in space telescopes are convenient because they are immune to Earth's high variation of observation conditions.

Table 3.2 – Statements used to assess the scientific relevance of PIXEL.

Figure 3.8 presents the percentage frequency distribution of the 571 responses. The most frequently selected items are “Improvement of technologies brings to an improvement of observation resolution” (19.8%), “Scientific research is a collaboration of different professionalities” (16.3%), and “High resolution images require more observation time” (14.9%). “Astrophysics research is shared with the scientific community” (12.6%) and “Publications are extremely useful for research centers” (12.3%) obtained similar results, followed by “Investments in space telescopes are convenient because they are immune to Earth’s high variation of observation conditions” (11.2%), and “An important research center is one with a more scientific prestige rather than more resources” (6.1%). The three control item (“Publications are the final goal of astrophysics research”, “Italian astrophysics research is paid less than abroad”, and “Astrophysics research requires to travel a lot”) scored less than 7% all together.

This initial analysis, discussed in detail by [Inchingolo et al. \(2023\)](#), confirmed that PIXEL successfully achieved its design intent: to create an engaging board game capable of stimulating reflection on the processes underlying scientific research even after a single gameplay session. In particular, players recognized the relationship between image resolution, the required observation time, and the technological development necessary to achieve such resolution. They also identified the collaboration of different professional roles as essential for the functioning of a research center, another key feature of the game.



Figure 3.8 – Percentage frequency distribution of scientific topic addressed by PIXEL.

3.2 Cosmo Hunters

While PIXEL is designed to support the experience of scientific research as a process, Cosmo Hunters addresses astrophysics from a complementary perspective, focusing on observational methods and conceptual understanding.

Cosmo Hunters (Figure 3.9) is a card game that aims to introduce the topic of astronomical observations of celestial objects at different frequencies of the electromagnetic spectrum to upper primary and middle school students (aged 10-14).

The design, built on the GBL principles discussed in Chapter 2, was guided by three main challenges commonly faced by teachers when implementing GBL activities in classroom settings. First, the selection or development of an appropriate educational game requires an attentive consideration of the game media literacy (Swertz, 2019) and familiarity with board games (Ligabue, 2020) to identify a suitable match between game features and learning goals. At first glance, many game rules may appear confusing or overly complex, potentially discouraging students who are unfamiliar with board gaming from fully engaging with the activity.

A second challenge concerns the considerable effort required from teachers to prepare and facilitate GBL sessions. Andreoletti et al. (2023) summarized the various roles teachers may assume during GBL activities – player, instructor, guide, observer, director, assistant, initiator, referee, game character, subject expert, and evaluator – many of which unfold before, during, and after the activity itself, thus increasing the workload and potentially



Figure 3.9 – Students playing Cosmo Hunters.

discouraging implementation.

Lastly, there is a strictly time constraint that makes it difficult to adopt GBL in classroom settings. Indeed, [Inchingolo et al. \(2024\)](#) noted that most astrophysics-themed board games available on the market have a duration ranging from 60 to 120 minutes. Considering the additional time needed to explain the rules and conduct a debriefing, it becomes difficult to fit such activities within a standard Italian school lesson, which typically lasts a maximum of 100 minutes.

These challenges are specifically addressed in Cosmo Hunters through deliberate design choices. The gameplay was kept short to allow the activity to fit easily within standard lesson schedules while still leaving room for discussion and reflection. The game mechanics were intentionally designed to be simple and intuitive, enabling students and teachers with little or no prior experience with board games to immediately engage with the activity without long explanations or complex rules. In addition, a set of freely accessible print&play materials was developed¹, including both the game components and a structured guide for the debriefing session. This supplementary material supports teachers in organizing and facilitating the activity, significantly reducing preparation time and the overall workload associated with GBL implementations.

The game features two main types of cards – Observation and Telescope – divided into six categories corresponding to the main frequency bands of the electromagnetic spectrum: radio, infrared, visible, ultraviolet, x-rays, and gamma rays. Each Observation card represents a celestial object observed at a specific frequency (Figure 3.10a) and is associated

¹Available here: play.inaf.it/en/cosmo-hunters-en/

with a certain number of points, while Telescope cards (Figure 3.10b) represent the instruments needed to perform those observations within the corresponding frequency band. A third category, Event cards (Figure 3.10c), introduces variability and unpredictability during play through the application of specific effects.

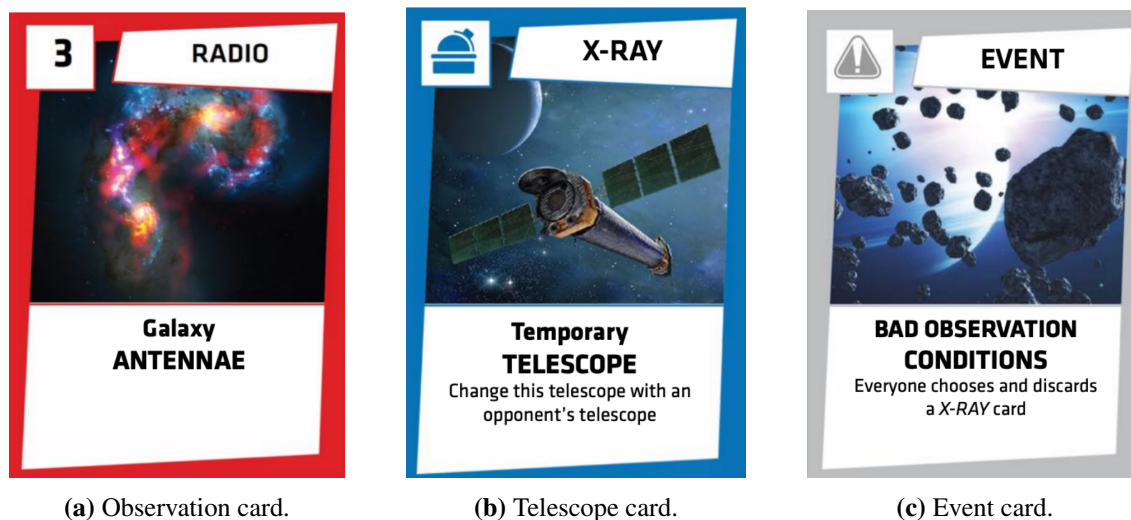


Figure 3.10 – Example of the types of card featured in Cosmo Hunters.

In each round, a fixed number of cards are revealed randomly from the deck, which contains all types of cards shuffled together. Players take turns selecting one card at a time. Observation and Telescope cards are accumulated in the player deck, while Event cards are immediately activated as indicated and then discarded. The round continues until all available cards have been selected. At that point, new cards are revealed, and a new round begins.

At the end of the game, each player's score is calculated by summing points of the Observation cards collected, only if the player also owns a Telescope card of the same frequency category. For instance, in the example shown in Figure 3.11, there is one ultraviolet Observation card, one infrared Telescope card, and one infrared Observation card. In this configuration, only the 3 points from the infrared Observation card are valid, since the player owns the corresponding Telescope card. The points of the ultraviolet Observation card, on the other hand, do not count because the matching Telescope card is missing.

Additionally, players can earn bonus points by collecting specific sets of Observation cards identified by shared symbols. These sets represent the same celestial object observed at multiple frequencies. As shown in Figure 3.12, the Centaurus-A galaxy appears in three different frequency bands. If, at the end of the game, a player owns all three corresponding Observation cards together with their matching Telescope cards, each Observation card will be worth 5 points instead of 3, for a total of 15 points.

The aesthetic design of the cards employs authentic astrophysical images of celestial

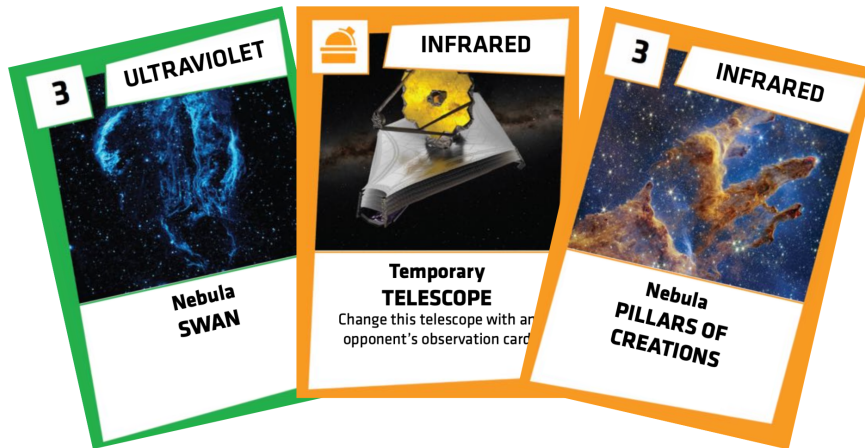


Figure 3.11 – Example configuration.

objects and telescopes, creating an evocative setting that enhances the scientific content conveyed through gameplay. In particular, the Telescope cards illustrate the diversity of instruments used in contemporary astrophysics, emphasizing that telescopes come in many forms, often very different from the familiar optical model.

The game mechanics are purposefully designed to simulate core processes of astrophysical research. The first one – requiring players to pair Observation and Telescope cards – reflects the idea that observing celestial objects requires specific instruments (telescopes) sensitive to certain regions of the electromagnetic spectrum. Consequently, astrophysicists must employ different types of telescopes to study different frequency bands.

The second mechanic – awarding bonus points for multi-frequency observations of the same object – embodies the astrophysical concept of multi-wavelength analysis, which is fundamental for understanding the physical processes occurring within celestial bodies. Since different types of radiation originate from distinct physical mechanisms, combining data from multiple frequency bands enables a more comprehensive interpretation of an object's nature.



Figure 3.12 – Centaurus-A Galaxy Observation card in three different frequencies.

3.3 Design choices and implications for evaluation

The choice of PIXEL and Cosmo Hunters as the case studies of this thesis is motivated by the fact that both games were developed with the explicit intention of exploiting the potential of GBL to introduce PES principles within a formal educational context such as school. In both cases, the focus is placed on the experience of science as a process rather than on the transmission of predefined factual knowledge. This shift – which reflects the broader transition from the PUS paradigm to that of PES – represents a relatively novel approach within the game-based learning literature, where the majority of studies have traditionally focused on content acquisition.

At the same time, the selection of two activities that differ substantially in terms of target audience, duration, and educational goals might appear problematic, as it prevents a direct comparison of outcomes between the two case studies. However, this heterogeneity also constitutes a strength, as it allows different dimensions of effectiveness to be investigated simultaneously, without restricting the analysis to a single construct or age group. Specifically, PIXEL was designed for upper secondary school students, whereas Cosmo Hunters targets lower secondary school students. Moreover, the difference on the duration of the activities leads to distinct modes of educational implementation. PIXEL has a duration of approximately two hours, making it a self-contained activity that can be integrated either within school curricula or in informal educational contexts such as science centers or game festivals. Cosmo Hunters, by contrast, has a much shorter duration (approximately 20 - 30 minutes), which makes it suitable to be included within a broader and more articulated

educational activity.

As a consequence, the educational objectives of the two games also differ. PIXEL aims to simulate scientific research as a complex process, with particular attention to the social and human dynamics that characterize the construction of scientific knowledge. Cosmo Hunters, instead, focuses primarily on the methodology underlying astronomical observations and on key conceptual aspects of astrophysics related to observations at different wavelengths. From the perspective of the RFN, PIXEL addresses both the cognitive-epistemic system (e.g., scientific practices, methods, and knowledge construction) as well as the social-institutional system (e.g., collaboration, competition, publication, and professional activities), whereas Cosmo Hunters is mainly oriented toward the cognitive-epistemic dimensions, with a particular emphasis on methods and knowledge.

These similarities and differences are necessarily reflected in the evaluation strategies adopted for the two case studies, which were designed to be coherent with their respective goals and contexts. As introduced in Section 2.3, in the case of PIXEL effectiveness is first and foremost investigated in relation to the understanding of NOS, interpreted as an exploration of which characteristics emerge most prominently through interaction with the game. This was addressed through the development of an articulated educational pathway. In addition, the effects of a single gameplay session were examined in terms of NOS, self-efficacy, and attitudes, by comparing the outcomes with those of an out-of-school activity.

In the case of Cosmo Hunters (Chapter 5) effectiveness is instead operationalized in terms of learning outcomes related to basic concepts of the electromagnetic spectrum and to the mechanisms underlying astronomical observations. In this context, a comparative study was conducted to evaluate the learning outcomes of an activity based on Cosmo Hunters and a frontal lecture addressing the same topics.

Chapter 4

PIXEL - Picture (of) the Universe: experiencing NOS

As stated in the previous chapter, PIXEL was developed with the explicit intention of simulating scientific research as a complex process, with particular attention to the social and human dynamics of astrophysics research. Within the context of PES, this reflects in the design of game mechanics that make players experience and engage with different dimensions of NOS.

To investigate the effectiveness of this approach, PIXEL was implemented and evaluated in two complementary contexts. First, the game was embedded within an extended GBL program, designed to provide students with repeated gameplay, facilitation experiences, and structured moments of reflection. Second, the impact of a single gameplay session was tested in a school setting, comparing PIXEL with an out-of-school activity and focusing on a set of constructs aligned with the game's framework.

These two studies presented in the following sections are intended as complementary lenses on the same research question: what can the game elicit, in terms of students' understanding of Nature of Science and their affective engagement with it?

4.1 PIXEL embedded in a GBL program

The first study of PIXEL involved a GBL program implemented to assess the effectiveness of this approach in terms of understanding of NOS. The project was built around the game and provided ample space for debriefing, that is the reflective discussion that occurs after a learning activity, in which participants analyze their experiences, connect them to underlying concepts, and consolidate knowledge or skills, turning play into learning (Crookall, 2010).

The research question that guided the design and evaluation of the program is the following:

Which aspects of NOS emerge from students' interaction with PIXEL when the game is embedded within an extended GBL program?

4.1.1 Sample and research design

The study involved 32 students of upper secondary school aged 15 to 17 in five Italian cities (Cagliari, Bologna, Rome, Palermo, Florence). The distribution by city, gender, and school type is reported in Table 4.1.

City	Type of school	Male	Female
Bologna	Scientific lyceum	1	4
Cagliari	Scientific lyceum	0	5
Florence	Scientific lyceum	4	0
Rome	Scientific lyceum	3	3
Palermo	Vocational-technical	7	5

Table 4.1 – Distribution of participants by city, type of school, and gender.

Participation in the activity was voluntary and integrated within the framework of a PCTO program (Pathways for Transversal Skills and Orientation, *Percorsi per le Competenze Trasversali e l'Orientamento* in Italian), a curricular initiative of Italian high schools that aims to provide students with a broader educational experience extending beyond formal learning. Students were not chosen randomly, but were encouraged to participate by their teachers. This resulted in the participation of students that were already interested in scientific-related topics.

The project lasted 40 hours distributed across 20 sessions held between February and June 2023. The main purpose of the program was to train participants to guide gameplay and debriefing sessions with their classmates, and to collect their impressions through reflective diaries. For this reason, the program was structured around the alternation of four phases reported hereafter (the detailed plan of interventions is reported in Table A.2 in the Appendix).

Training sessions. Students learned PIXEL rules, the disciplinary content embedded in it, the basics of GBL, and how to act as expert facilitators of the game. The role of the facilitator is to guide the gameplay without influencing the players' choices, i.e., ensuring that the rules are respected, guiding the group through the different phases of the game, and resolving any doubts. This must be done with care not to provide strategic advice. At the end of the gameplay, the facilitator is also responsible for leading the debriefing session.

Facilitation experiences. After the training, participants put the facilitation techniques into practice by running PIXEL gameplay sessions and debriefing sessions with their classmates. The debriefing phase was organized around four guiding questions: (1) How did

you feel at the beginning and during the game? (2) What surprised or impressed you about the game? (3) What strategies/ideas did you use to achieve your goals? (4) What challenges are represented in the game?

Players were asked to write down the answers. The facilitator then read the responses and initiated a discussion on the four themes, with the aim of encouraging participants to collectively build meaning from their experiences.

Compilation of reflective diary. Between each session mentioned above, participants independently studied the provided materials, analyzed the debriefing data collected during the facilitation experiences, and completed a reflective diary, which constitutes one of the main sources of data for this study.

Feedback session. At the midpoint and at the conclusion of the educational program, astrophysicists and INAF researchers conducted two feedback and support meetings. The purpose was to analyze the progress of the facilitation experiences and to discuss about them.

Reflective diaries were chosen as a evaluation tool because they represent a well established practice for stimulating metacognition, engagement, and critical reflection. A reflective diary is a form of reflective writing in which the writer describes situations, events, interactions, or memories and adds a personal consideration of their meaning. This practice is typically guided by open-ended questions such as “What did I learn?”, “How has this changed me?”, or “What might I have done differently?” (Helyer, 2015). The literature highlights that reflective diaries can probe and promote students’ engagement (Tang, 2002), support the development of self-regulated learning (Wallin and Adawi, 2018), and enhance critical thinking, professional growth, and the integration of theory with practice (Tang, 2002). Moreover, they can be considered formative assessment tools, providing a valuable instrument for the qualitative evaluation of students’ learning.

At the same time, the use of reflective diaries is not only a method of data collection but also an integral part of the learning process itself. Reflective practice requires students to observe their own actions, evaluate them, and generate and monitor improvement, thereby making the act of documenting their own change. In this sense, the research design aligns with the tradition of action research (Cunningham, 1976; Lewin, 1948), in which the act of observation can directly influence the phenomenon under study because of the reflective nature of the instrument.

For this study, the evaluation was conducted through nine reflective diaries containing open-ended questions on different themes. The complete set of questions is presented in Table A.3 in the Appendix.

4.1.2 Data analysis

The responses from the reflective diaries (32 entries for each question) were analyzed using a technique inspired by the reflexive thematic analysis of [Braun and Clarke \(2019\)](#).

For the purposes of this study, only the first and last sessions of the reflective diary (Diary #1 and #9 in Table A.3) are presented in the following sections, since the others were used to keep track of the students' progress in facilitating and understanding the game. Sessions #1 and #9 include the same three questions and can therefore be examined using a pre-post approach: "What is science for you?" (pre: 1.1, post: 9.1), "How does scientific research work?" (pre: 1.2, post: 9.2), and "How do researchers work?" (pre: 1.3, post: 9.3).

The procedure of data analysis involved the following steps. A group of ten INAF and GSRC researchers, working in pairs, carefully read each response and independently identified tags representing the themes expressed. The resulting tags were then analyzed and discussed one by one in a plenary session in order to converge and assign the definitive ones. This process followed an iterative path, with several cycles of tag creation, evaluation, and adjustment conducted to establish the final pool of tags. For example, to the answer "Observation of a phenomenon, formulation of a law, and experiment to verify the law" was assigned the tag *scientific method*, as it directly reproduces the classical linear description of the scientific process. On the other hand, to the entry "After this experience I understood that scientific research is composed of the joint work of several researchers engaged in a common goal" was assigned the tags *collaboration*, *commitment*, *objectives*, and *researchers*. These tags capture the emphasis placed on the cooperative nature of scientific work, the recognition of the dedication required to pursue research, the focus on the presence of common aims as driving forces, and the explicit mention of the role of researchers as individuals.

During the analysis process, an additional dimension emerged that could not be fully addressed through thematic analysis alone. This dimension concerned the tone of the responses and the degree of personal style adopted. To account for this, an additional level of content evaluation and interpretation – called modal analysis – was implemented, categorizing the responses according to a set of specific tags described in Table 4.2. This categorization was employed to quantify a shift in the students' expressive register, from bookish and stereotypical to more personal accounts.

Then, for each question, the number of occurrences of each tag was counted across the entire sample, in order to obtain a precise indication of how many times specific themes were mentioned and to identify their relative weight within students' responses. Subsequently, these frequencies were used to calculate the percentage frequency for each tag.

Finally, the tags that emerged from the responses were grouped within the two systems of the RFN: the cognitive-epistemic and the social-institutional. The classification is

Tag	Description	Example
Formal response	A response that follows a formal structure or style, often derived from what has been memorised in school, but may lack a deep understanding of the concept.	<i>1.2 How does scientific research work?</i> “Scientific research is based on the scientific method which involves: observation of a phenomenon, formulation of a hypothesis through questions, verification of the hypothesis, formulation of the conclusion supported by an experiment that verifies its validity and possible application.”
Personal elaboration	A response that reflects a deeper understanding of the concept, often enriched by personal experiences and knowledge, indicating a more profound reflection.	<i>9.2 How does scientific research work?</i> “Scientific research is a collaborative work but also strategy and a thoughtful commitment of resources.”
Superficial	A response given quickly and superficially, indicating low engagement and minimal attention to the task. Usually brief and lacking detail or in-depth explanation.	<i>1.2 How does scientific research work?</i> “Obtain the appropriate information that is needed to put future research into practice.”
Imprecise terminology	A response that uses vague or ambiguous words or phrases, potentially causing confusion or a lack of clarity in meaning but not necessarily incorrect.	<i>1.3 How do researchers work?</i> “They work following several written and unwritten rules of science.”
Off-topic	A response not pertinent to the question or topic under consideration.	<i>1.2 How does scientific research work?</i> “Go to the root of events.”
Incorrect answer	An objectively incorrect response.	<i>1.3 How do researchers work?</i> “Group of researchers work to find new astrological phenomena.”

Table 4.2 – Tags and descriptions used to categorize the responses, with illustrative examples.

presented in Table A.5 in the Appendix. For example, the term *collaboration* was placed within the social-institutional system because it pertains to the category of social organizations and interactions, whereas *discovery* was included in the cognitive-epistemic system, as it represents one of the aims and values of science.

The adoption of multiple levels of analysis reflects the multidimensional nature of the research question. The thematic analysis was employed to identify the specific concepts and ideas students associated with science, scientific research, and researchers. However, focusing solely on what students mention does not fully capture how their understanding evolves. For this reason, a modal analysis was introduced to account for changes in the tone, structure, and personal involvement expressed in the responses. Finally, the mapping of themes onto the RFN framework made it possible to interpret these changes at a higher conceptual level.

4.1.3 Results

This section reports the data analysis results divided into the three main questions, the modal analysis and the mapping within the RFN systems. For a detailed report of data see Table A.4 in the Appendix.

What is science for you?

Figure 4.1 reports the percentage of themes that emerged in the responses to the question “What is science for you?” asked both in Diary #1 (pre) and in Diary #9 (post).

In the pre-test, ten distinct themes emerged, whereas in the post-test, sixteen were identified – indicating not only an expansion in the range of concepts mentioned but also a diversification in how students conceived the nature and role of science.

At the beginning of the project, the most frequently cited theme was *study of nature* (26.7%), followed by *knowledge* (15.6%) and *scientific method* (13.3%). Together, these categories reveal that students initially viewed science primarily as a descriptive and procedural activity in line with the way science is taught in Italian schools. However, these same themes also exhibited the most pronounced decline by the end of the project. *Study of nature* decreased by 12.4% and *scientific method* by 11.5%, while *knowledge* showed only a slight reduction (-3.1%). This pattern suggests a gradual departure from decontextualized and definition-like conceptions of science.

Among the new themes emerged in the post-test, *experimentation* (+7.1%) and *positive feelings* (+7.1%) experienced the highest increase. While *experimentation* points to the recognition of science as a process of trial and errors, *positive feelings* introduces a novel affective dimension, such as in the two illustrative answers reported below:

For me, science represents curiosity, the beauty of discovering our planet,

which we gradually learn more about.

Science is unity, people brought together by a passion for knowledge.

This emotional engagement contrasts sharply with the earlier dominance of cognitive or procedural categories such as *scientific method* and *study of nature*, suggesting a shift from definitional or formal views of science toward experiential and value-laden perspectives.

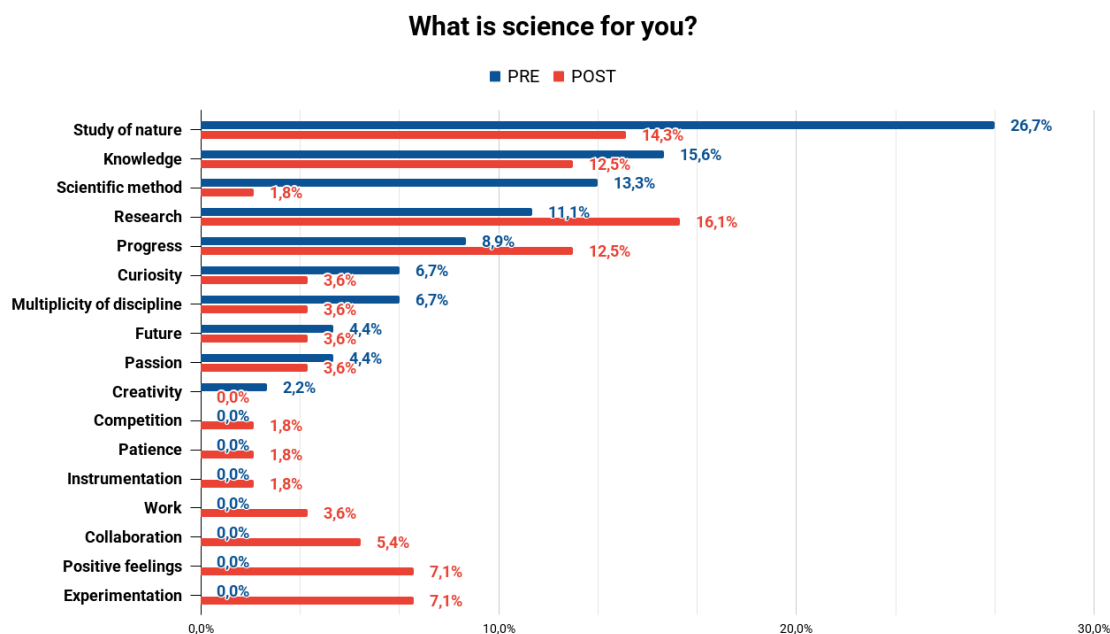


Figure 4.1 – Percentage of the themes emerged in the answers to the question “What is science for you?”.

How does scientific research work?

Figure 4.2 shows the distribution of themes regarding the question “How does scientific research work?”.

Students’ responses revealed an even greater expansion in thematic diversity compared to the previous question, going from 7 themes in the pre-test to 18 in the post-test. At the beginning, the vast majority of students described research almost exclusively in terms of the *scientific method* (71.4%), while by the end of the project, the theme dropped by 61.5%.

Moreover, a wide range of new themes emerged, many of which directly reference the specific language and context of PIXEL. Words such as *resources* (4.9%), *strategy* (4.9%), *improvements* (4.9%), *contingencies* (1.2%), *time* (3.7%), *astronomical observation* (7.4%), *instrumentation* (6.2%) and *objectives* (9.9%) reflected the influence of the game’s mechanics in shaping students’ reflections. At the same time, more traditional but

conceptually rich themes also increased in frequency, such as *publication* (+4.5%), *collaboration* (+5.7%), and *researchers* (+5.4%). Their rise suggests that students recognized the social and institutional dimensions of science, aspects that were explicitly represented within the game dynamics.

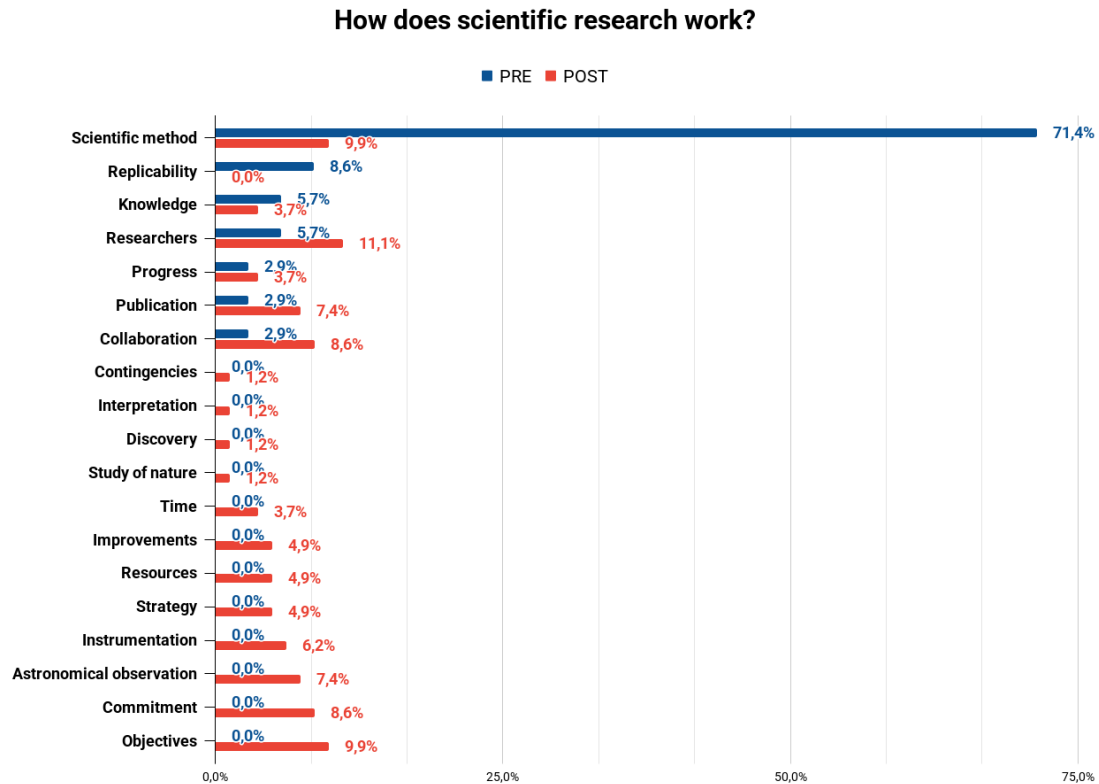


Figure 4.2 – Percentage of the themes emerged in the answers to the question “How does scientific research work?”.

How do researchers work?

Figure 4.3 displays the percentage frequencies of the themes emerged from responses to “How do researchers work?”.

As in earlier results, *scientific method* experienced a substantial decrease, dropping from 26.7% to 9.3% (-17.4%). Another emerged pattern is the emergence of new themes that directly reflect the terminology and mechanics of PIXEL as happened in the previous question. Terms such as *resources* (7.4%), *planning* (1.9%), *instrumentation* (3.7%), and *outreach* (1.9%) appeared only after the activity. Among these *objectives* experienced the highest increase (18.5%), encompassing also ideas of goal-setting within the broader scientific community, such as in these illustrative answers:

Researchers work collaboratively with each other to achieve a goal. Very often, even if the final goals are not quite the same, one researcher's work can benefit all.

You work with your own goals in mind, in parallel with colleagues who are proceeding in the same way, trying to have strategies to pursue. It is important to invest in what you think is most useful.

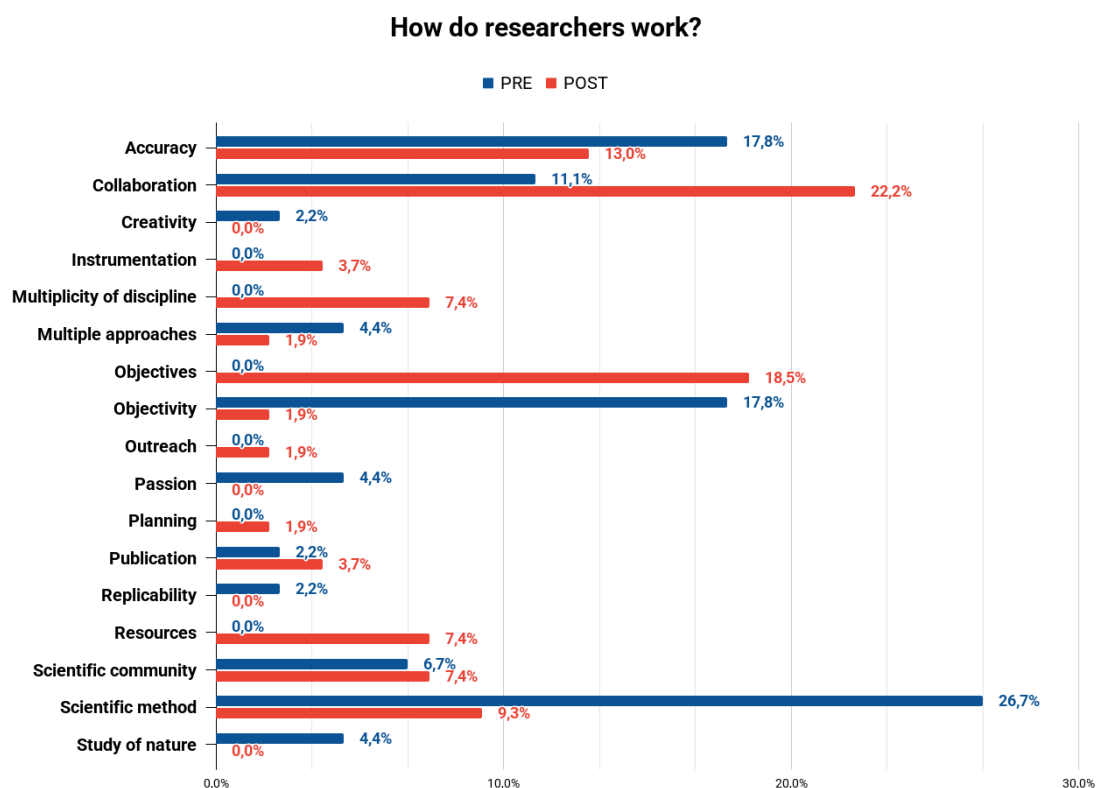


Figure 4.3 – Percentage frequencies of the themes emerged in the answers to the question “How do researchers work?”.

On the other hand, the theme of *objectivity* declined sharply (from 17.8% to 1.9%). The game portrayed research as a human endeavor shaped by cooperation and competition, undermining the idea of science as a purely objective, impersonal process. This was balanced by an increase in *collaboration* (+11.1%) and the emergence of *resources* (+7.4%), *multiplicity of disciplines* (+7.4%), and *instrumentation* (+3.7%).

A brief reflection is warranted on the variation of *replicability* (-2.2%), which disappeared also in the previous question. This aspect, typically emphasized in school education – where students learn that science must be replicable in order to be refutable – is not represented in the game. This absence reflects the idea that any simulation of the scientific

process, whether through games or active learning activities, can never be a perfectly faithful reproduction of reality. By definition, a simulation inevitably omits certain elements. What is essential, therefore, is to be aware of which aspects are being emphasized and which are being overlooked, maintaining a conscious balance between simplification and representation.

Modal analysis

Figure 4.4 shows the results of the modal analysis, based on a comparison of the distribution of response categories (Table 4.2) before and after the activity across the three questions discussed above.

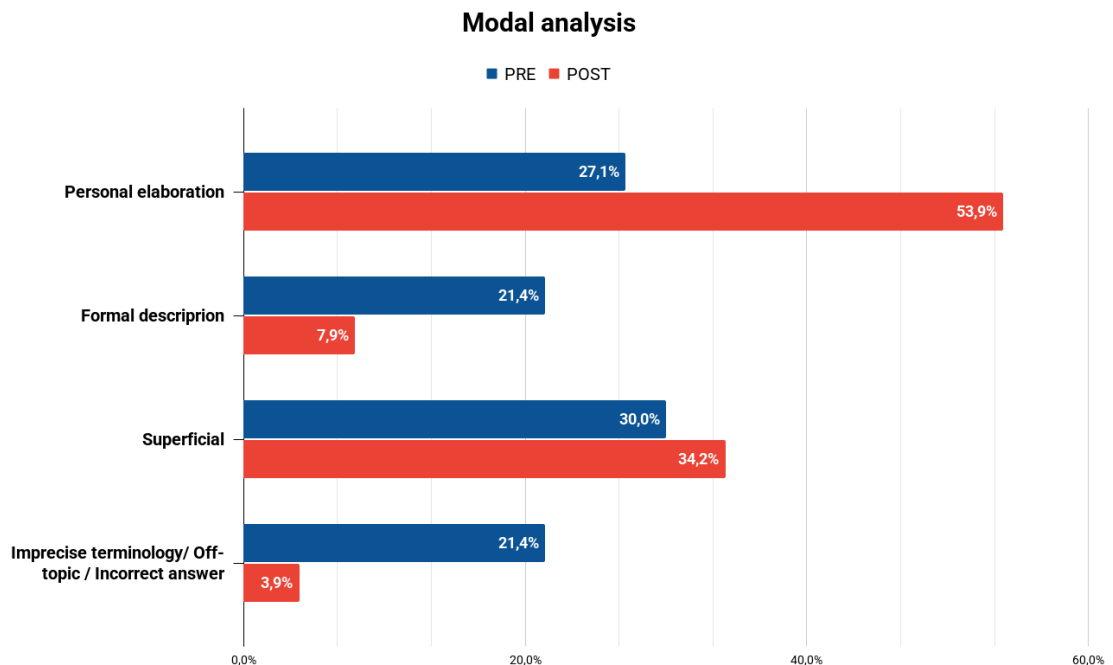


Figure 4.4 – Frequency distribution of the tags of the modal analysis.

Before the activity, answers were almost equally distributed among *personal elaboration* (27.1%), *formal description* (21.4%), *superficial* (30.0%), and *imprecise / incorrect / off-topic* categories (21.4%). At the end of the project, the latter category nearly disappeared, while a slight increase in *superficial* (+4.2%) responses may reflect fatigue from the frequency of reflective journals (nine in total), which imposed a notable workload.

The most significant transformation lies in the shift from *formal description* (-13.5%) to *personal elaboration* (+26.8%). Students increasingly abandoned textbook-like formulations in favor of answers shaped by personal reasoning and experience. This transition

signals a deeper change in how students perceive science: from something to be described formally, to something they can speak about in their own terms.

RFN categories mapping

Figure 4.5 shows the mapping of the themes onto the RFN framework, displayed as percentage distribution. See Table A.5 in the Appendix for the details of themes classification.

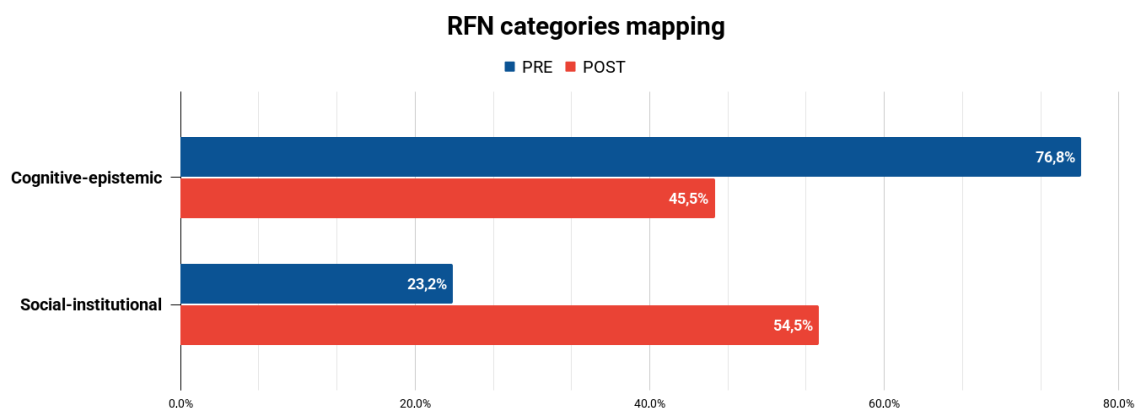


Figure 4.5 – Percentage frequencies of the themes that address the two RFN system.

The themes related to the RFN cognitive-epistemic system accounted for 76.8% of the total in the pre-phase and 45.5% in the post-phase, showing a decrease of 31.3%. Conversely, the social-institutional system increased from 23.2% to 54.5%, with a corresponding growth of 31.3%.

This shift indicates a progressive balancing in students' representations of science: while initially focused on knowledge and other epistemic aspects, their post responses placed greater emphasis on the human and social dimensions of scientific work. Such a transformation confirmed the ability of PIXEL, integrated in a GBL program, to give a representation of science focused mainly on social-institutional aspects of it.

4.1.4 Discussion

With respect to the research question guiding this study – namely, which aspects of the Nature of Science emerge from students' interaction with PIXEL when the game is embedded within an extended GBL program – the analysis highlights a shift from representations focused predominantly on cognitive-epistemic dimensions toward more articulated views that also include social-institutional aspects, mirroring the paradigm shift from PUS to PES.

At the beginning of the program, students' accounts largely reflected a school-driven and procedural image of science, consistent with previous research showing the persistence

of naïve or reductionist NOS views in formal education contexts (see, for example, [Abd-El-Khalick and Lederman, 2023](#)). By the end of the program, however, students increasingly referred to collaboration, objectives, resources, contingencies, and professional roles – elements that mirror the mechanics of PIXEL. This alignment suggests that students did not merely recall abstract ideas discussed during the activities, but appropriated concepts that had been experienced through gameplay.

A key factor in this transformation lies in the structure of PIXEL and the GBL program itself. The combination of repeated gameplay, facilitation experiences, and reflective diaries provided students with multiple opportunities to oscillate between action and reflection. In particular, the reflective diaries played a central role in making implicit experiences explicit, allowing students to connect in-game dynamics with real scientific practices. This finding supports the view that meaningful NOS learning requires explicit reflection approach. Indeed, [Buaraphan \(2012\)](#) states that students need the opportunity to map connections between the activities performed during the experience and those of scientists to make meaning of it, a process that was possible thanks to the reflective practice of diaries.

Importantly, what emerges from this study is the impact of PIXEL integrated within a carefully designed educational framework. The game functioned as a mediating tool that structured experiences, while learning outcomes were shaped by the surrounding pedagogical design. This distinction is crucial to avoid attributing to the game effects that are instead produced by the broader GBL ecology.

4.2 Short-term impact: a comparison with an out-of-school activity

The previous section proved that PIXEL, integrated into an extended GBL program for high school students, can foster a more nuanced and human perception of science and its nature. This approach, although effective, is complex to implement within the standard school routine due to its scope and duration.

For this reason, this second study adopts a comparative design, benchmarking PIXEL against an out-of-school-time (OST) activity: a guided visit to a radioastronomy facility. This comparison is motivated by the fact that OST activities are commonly regarded as highly engaging and educationally valuable, thus providing a meaningful baseline for an in-class activity that is itself non-traditional, such as PIXEL. By doing so, the study aims to mitigate the novelty bias that often positively affects students' responses to activities that differ substantially from standard lecture-based instruction in formal educational settings.

Therefore, the study evaluate the impact of a single play session in school contexts on three dimensions informed by PIXEL. The first dimension concerns students' perception of science as a social-institutional system, with the aim of exploring what can emerge after

a single gameplay session (instead of a longer program). The second and third dimensions address attitudes and self-efficacy in astronomy. These constructs were included to assess the impact of the intervention in terms of affective engagement with the discipline within a broader and less self-selected sample than the one involved in the previous study.

The research questions (RQ) are the following:

RQ1: Can a single gameplay session of PIXEL, compared to an out-of-school activity, foster students' understanding of the social-institutional aspects of science?

RQ2: Can a single gameplay session of PIXEL, compared to an out-of-school activity, positively influence their attitudes and their self-efficacy in astronomy?

In addition, qualitative feedback from teachers of the experimental group classes was gathered through interviews, with the aim of complementing the quantitative data by providing deeper insights into students' behaviors and into the educational impact of the activity as close observers of their students' daily interactions. Teachers were involved because they have a privileged perspective on classroom dynamics. From this privileged point of view, they are often able to notice aspects of GBL activities that may go unnoticed by both the students and external observers.

4.2.1 Sample and research design

The study involved 188 students from fourth year of high-school (Grade 11, aged 17 and 18 years old) according to a convenience sample, i.e., based on their easy accessibility and availability. The participating classes were divided into two groups: an experimental group (involved in the GBL activity) and a control group (involved in the OST activity).

Students in the experimental group participated in a PIXEL session facilitated by INAF researchers within their classroom. The activity lasted two hours and was carried out in three high schools (lyceum) located in Bologna and Modena, involving a total of five classes and 113 students.

The control group took part in OST guided tour to the Medicina radiotelescopes¹ of the same duration, structured as follows: a 40-minute introductory lecture on radioastronomy, the screening of a 12-minute 3D video illustrating both cognitive-epistemic and social-institutional processes of the discipline, a guided tour of the antennas, and a visit to the interactive museum section of the visitor center. The visits were conducted by INAF researchers as well. The four classes involved in the control group were of a vocational-technical high school in the province of Urbino, comprising a total of 55 students.

The gender distribution of the two groups is presented in Table 4.3.

¹See the website of the visitor center: centrovisite.ira.inaf.it

Group	Male	Female	Other
Experimental	38	33	0
Control	28	2	1

Table 4.3 – Gender distribution of the two groups.

For both groups, data collection was carried out through the administration of a pre-test (online, three days before the activity) and a identical post-test immediately after the activities. The time interval between the two administrations was designed to minimize memory effects. The questionnaire, presented in detail in Table A.6 in the Appendix, consisted of statements for which students were asked to indicate their level of agreement on a 5-point Likert scale, and one open-ended question. Students were also asked to indicate, in addition to their gender, their class number and student ID, in order to create a unique anonymous code for each participant.

The students’ understanding of NOS was evaluated through the Social-Institutional Questionnaire (SIQ) developed by [Akbayrak and Kaya \(2020\)](#), consisting of 15 Likert-scale items. It was selected among all the evaluation instruments developed within this framework because it offers a concise yet comprehensive measure of students’ understanding of the social and institutional dimensions of science. Even if it was specifically designed and validated for primary-school students (fifth grade), it is more appropriate for the target population of this study than other instruments developed for university students or pre-service teacher samples (such as the one of [Kaya et al., 2019](#)). Moreover, its brevity and clear item formulation further ensured suitability for administration within a single session, without compromising the reliability of the results.

In addition to the SIQ, an open-ended question was introduced, designed to promote deeper reflection on the topic, in line with the considerations drawn from the previous study. The question asked to state which are the professional activities carried out by astrophysics researchers in their job.

Attitudes towards astronomy was evaluated with the scale developed by [Galano et al. \(2024\)](#). The tool is subdivided into three categories of attitude: intrinsic (interest and enjoyment in the study of astronomy), extrinsic (intention to pursue a career in astronomy), and utilitarian (perception of the relevance of astronomy for society). The decision to use this scale was lead by two main aspects. First, it has been validated with a sample of upper secondary students (9-13 Grade) comparable in age to the participants involved in this study. Then, its tripartite structure aligns closely with the potential impact of the game-based experience. Indeed, the intrinsic dimension may be enhanced through the engaging challenges of PIXEL, which foster curiosity and enjoyment in exploring astrophysical concepts. The extrinsic component may be informed by the game’s realistic representation of multi-

disciplinary and complex astrophysics research processes. On the other hand, although the utilitarian dimension is not explicitly addressed in the game's mechanics, participating in an activity that frames science as a human endeavor may nonetheless contribute to a more positive overall perception of its social value.

Self-efficacy was evaluated through items adapted from [Galano et al. \(2024\)](#). Since the original scale had been developed and validated for a formal educational context, only the questions concerning astronomy in general, rather than as a school subject, were selected.

Unlike the previous study, the assessment tools described above were not specifically developed for this experience but were instead adapted from existing studies. This may be considered a limitation, given that these dimensions are notoriously difficult to influence over the short duration of a single PIXEL gameplay. However, these tools remain the most reliable means of providing quantitative evidence of potential changes. For this reason, the present study should be regarded as an initial step that paves the way for a broader investigation, a testing ground aimed at identifying and validating suitable instruments for assessing the educational impact of a single PIXEL gameplay session.

4.2.2 Data analysis

Once data collection was completed, only questionnaires for which both pre and post-test responses had the same unique code were included in the analysis. Questionnaires of students that were completed in only one of the two phases were excluded. This procedure resulted in a final dataset of 71 valid cases for the experimental group and 31 for the control group, corresponding to approximately 62.8% and 56.4% of the original participants, respectively. This significant difference between the number of participants and the questionnaires collected can be attributed to several factors. Some students were present on the day of the pre-test but not on that of the post-test, or vice versa. In addition, since participation was voluntary, some students chose not to complete the questionnaire. Furthermore, a number of questionnaires were excluded due to the impossibility of matching them to a unique identification code.

Then, the analysis followed two different approaches depending on the dimension evaluated. Data from the SIQ, self-efficacy, and attitude tests were analyzed using t-tests and ANCOVA, statistical methods commonly employed in education research. For a detailed description of these procedures see Section A.2 in the Appendix. The open-ended question was analyzed following a procedure analogous to that described in Section 4.1.2, namely through the definition of tags representing the main concepts expressed in the students' responses and mapping them in the RFN systems.

For the teachers' interview, questions focused on their observations of student engagement, interaction dynamics, role changes, skills and competences activated during the activity, connections with curricular learning, and their overall evaluation of the experience

(see Table A.7 in the Appendix for details). The questions were delivered via e-mail some weeks after the activity, and three out of five teachers responded.

4.2.3 Results

This section presents the results of the study divided into the dimensions evaluated: students' understanding of NOS, self-efficacy, attitude and teachers' perception of the activity. The details of the descriptive statistics are reported in Appendix A.

Students' understanding of the social-institutional aspects of science

The SIQ datas were compared using a paired-samples t-test between pre and post-test scores within both the control and experimental groups. For the experimental group ($N = 68$), the pre-test mean (58.37 ± 0.57) and post-test mean (58.87 ± 0.53) showed no statistically significant difference in the test ($t(67) = -0.90$, $p = 0.374$, $d = -0.11$). For the control group ($N = 30$), the mean pre-test score (56.47 ± 1.10) was slightly lower than the post-test score (58.67 ± 0.89), but the difference was no statistically significant in the test ($t(29) = -1.87$, $p = .071$, $d = -0.34$).

Within the experimental group, male students ($N = 35$) obtained a mean pre-test scores of 58.40 ± 0.81 in the pre-test and 59.69 ± 0.68 in the post-test, whereas female students ($N = 33$) scored 58.33 ± 0.83 in the pre-test and 58.00 ± 0.80 in the post-test. The two-tailored t-test showed no significant differences between pre and post scores for either males ($t(34) = -1.45$, $p = 0.157$, $d = -0.244$) or females ($t(32) = 0.519$, $p = 0.607$, $d = 0.09$).

These results are visually confirmed by Figure 4.6, which displays the mean pre and post scores for both groups and gender division for the experimental group, with error bars representing ± 2 SE (95% confidence interval). Although the two samples shows a slightly pre-post variation, the mean values fall well within the error bars, indicating that the observed differences are not statistically meaningful.

The ANCOVA results showed that the overall model was significant ($F(4, 92) = 6.16$, $p < 0.001$), explaining approximately 21% of the variance in post-test scores ($R^2 = 0.21$, adjusted $R^2 = 0.18$). Pre-test scores were a significant predictor of post-test performance ($F(1, 92) = 20.76$, $p < 0.001$), indicating that students with higher initial SIQ scores tended to maintain higher scores after the intervention. However, the main effects of group ($F(1, 92) = 1.03$, $p = 0.313$), and gender ($F(1, 92) = 0.05$, $p = 0.825$) were not statistically significant. These results suggest that, after controlling for pre-test differences, the intervention did not produce measurable effects, and no gender-based differences were observed.

For the open-ended question "In your opinion, what kind of activities do astronomers carry on during their work?", were compared 62 paired pre-post responses from the experimental group and 27 from the control group. The list of themes that emerged, organized

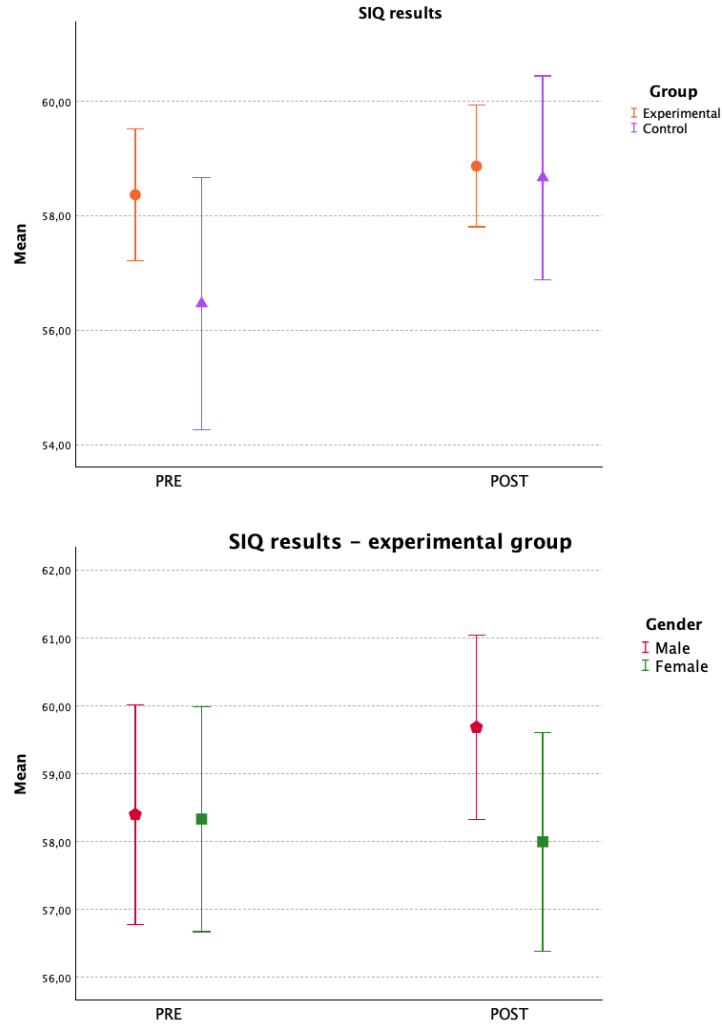


Figure 4.6 – Mean scores of SIQ divided in groups (top) and gender of the experimental group (down). Error bars are equal to 2SE, corresponding approximately to 95% confidence intervals.

according to the RFN system categories, is reported in Table A.8 in the Appendix. Responses that were brief or lacking in detail were categorized as *superficial* and included in the final percentage frequency representation of Figure 4.7.

In the pre-test, 82.6% of the experimental group's responses referred to the cognitive-epistemic system, while only 3.0% mentioned social-institutional aspects, and 14.4% were classified as *superficial*. After the intervention, references to the cognitive-epistemic system decreased to 72.3% (−10.3%), whereas those to the social-institutional system increased to 16.3% (+13.3%), suggesting a shift toward a broader and more socially contextualized understanding of science. The proportion of *superficial* answers slightly decreased to 11.3% (−3.1%), indicating a modest improvement in the depth and articulation of responses.

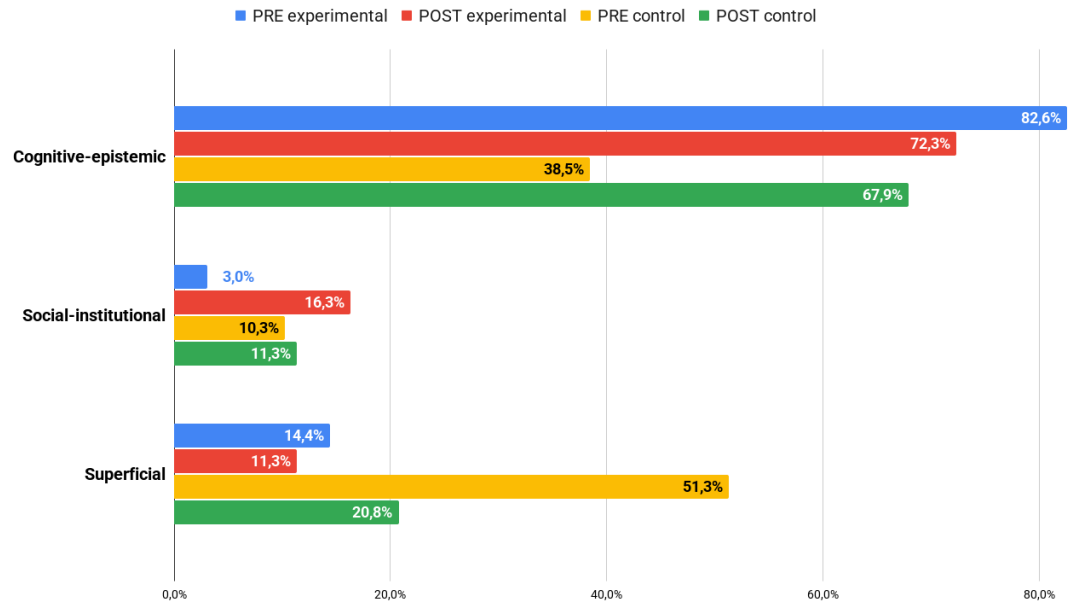


Figure 4.7 – Frequency distribution of themes addressing RFN systems.

By contrast, the control group exhibited different changes. Cognitive-epistemic themes increased from 38.5% to 67.9% (+29.4%), while social-institutional themes remained almost stable (10.3% to 11.3%), and *superficial* responses dropped substantially from 51.3% to 20.8% (-29.5%).

These variations can be interpreted in light of the different composition of the two groups. The experimental group consisted of students from lyceum (academic-oriented high schools), which in the Italian educational system are characterized by a strongly theoretical curriculum. This explains the initially high proportion of responses related to the cognitive-epistemic system, as well as the relatively stable proportion of *superficial* answers across pre and post phases. Conversely, the control group was composed of students from a technical institute, whose curriculum places greater emphasis on practical rather than theoretical learning. This accounts for the lower initial proportion of cognitive-epistemic responses and the higher incidence of *superficial* ones. For these students, the visit to the radiotelescopes likely represented one of their first opportunities to interact with an astronomical research environment, rather than with the technological aspects that are more common in their studies.

Attitudes towards astronomy

Attitudes is evaluated considering its three sub-dimensions: extrinsic (intentions related to pursuing astronomy-related studies or careers), intrinsic (interest in and enjoyment), and

utilitarian (perceived relevance and usefulness for society).

Based on the t-test results of the intrinsic attitude, for the experimental group ($N = 69$), the mean pre-test score (14.99 ± 0.37) and post-test score (14.75 ± 0.30) did not differ significantly ($t(68) = 0.90$, $p = 0.373$, $d = 0.11$), indicating a negligible effect size. Similarly, for the control group ($N = 30$), the pre-test (13.87 ± 0.56) and post-test (13.70 ± 0.48) means were not significantly different ($t(29) = 0.31$, $p = 0.758$, $d = 0.06$). Within the experimental group, neither males ($t(35) = 0.65$, $p = 0.523$, $d = 0.11$) nor females ($t(32) = 0.66$, $p = 0.516$, $d = 0.11$) showed statistically significant differences between pre and post-test scores.

These results are also visually represented in Figure 4.8, which displays the mean pre and post-test scores for the intrinsic sub-dimension of attitude. In both representations – by group and by gender – the mean values fall within the 95% confidence interval. Still, the graph shows a slight discrepancy between the two groups, with the experimental group displaying higher values than the control group both in the pre and post-test.

The overall ANCOVA model was significant ($F(4, 94) = 17.92$, $p < 0.001$), explaining approximately 43% of the variance in post-test scores ($R^2 = 0.43$, adjusted $R^2 = 0.41$). Pre-test scores were a significant predictor of post-test performance ($F(1, 94) = 64.85$, $p < 0.001$), indicating that students with higher initial scores on the intrinsic attitude tended to maintain higher post-test scores. In contrast, the main effects of group ($F(1, 94) = 1.10$, $p = 0.296$), and gender ($F(1, 94) = 0.27$, $p = 0.603$) were not statistically significant.

For the extrinsic sub-dimension, in the experimental group ($N = 69$) the difference between pre-test (8.72 ± 0.45) and post-test (8.78 ± 0.38) scores was not statistically significant ($t(68) = -0.21$, $p = 0.836$, $d = -0.03$). Similarly, for the control group ($N = 29$), no significant difference emerged between pre-test (8.59 ± 0.60) and post-test (8.10 ± 0.52) scores ($t(28) = 0.90$, $p = 0.375$, $d = 0.17$). Within the experimental group, neither males ($t(35) = 0.06$, $p = 0.951$, $d = 0.01$) nor females ($t(32) = -0.48$, $p = 0.633$, $d = -0.08$) showed significant differences between pre and post-test scores.

These findings are further illustrated in Figure 4.9, which depicts the mean pre and post-test scores for the extrinsic dimension of attitude. Both the group-level and gender-level comparisons show that the mean values remain within the 95% confidence intervals.

The overall model resulted by the ANCOVA was significant ($F(4, 93) = 28.71$, $p < 0.001$), explaining approximately 55% of the variance in post-test scores ($R^2 = 0.55$, adjusted $R^2 = 0.53$). Pre-test scores were a significant predictor of post-test performance ($F(1, 93) = 107.19$, $p < 0.001$), indicating that students' initial levels of extrinsic attitude strongly influenced their post-test outcomes. A significant main effect of group was found ($F(1, 93) = 5.13$, $p = 0.026$), suggesting that, after controlling for pre-test differences, students in the experimental group reported slightly higher post-test scores compared to the control group. In contrast, the main effect of gender was not statistically significant ($F(1, 93) = 3.71$, $p = 0.057$), indicating that male and female students of the experimental group did not differ

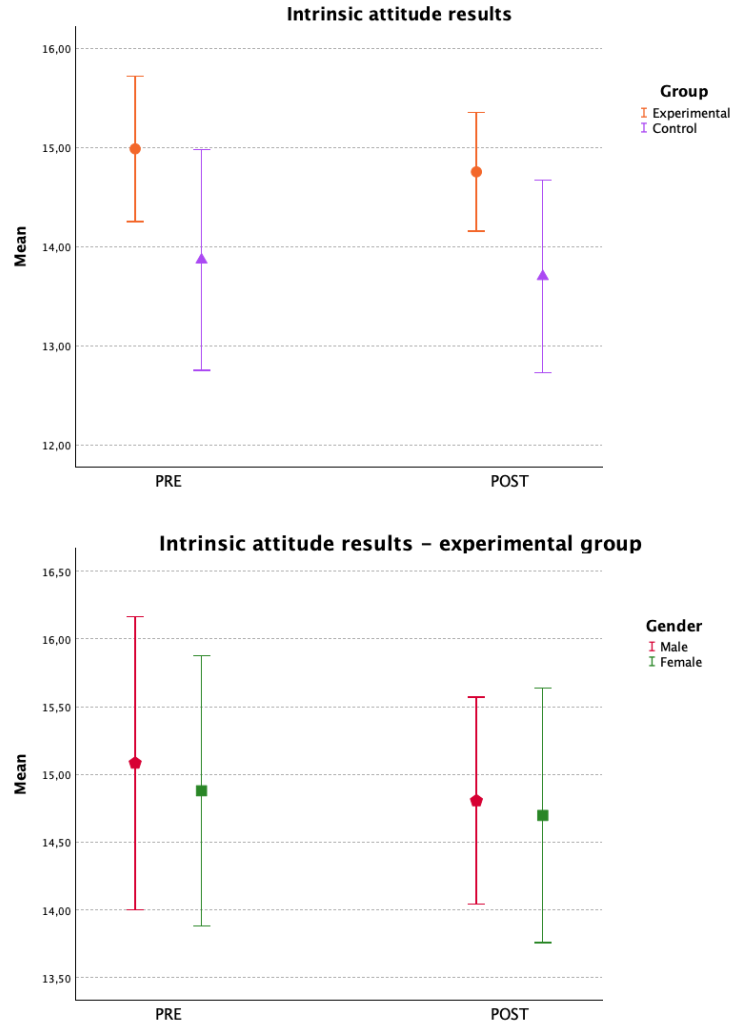


Figure 4.8 – Mean scores of intrinsic attitude divided in groups (top) and gender of the experimental group (down). Error bars are equal to 2SE, corresponding approximately to 95% confidence intervals.

meaningfully in their post-test scores on the extrinsic dimension of attitude.

For the utilitarian sub-dimension, based on the t-test results of the experimental group ($N = 67$), the mean pre-test score (16.93 ± 0.40) and post-test score (17.87 ± 0.38) differed significantly ($t(66) = -2.98$, $p = 0.004$, $d = -0.36$), indicating a small but statistically significant improvement following the intervention. For the control group ($N = 30$), the difference between pre-test (16.80 ± 0.42) and post-test (18.17 ± 0.40) scores was also significant ($t(29) = -3.18$, $p = 0.004$, $d = -0.58$), corresponding to a medium effect size. When the experimental group was examined by gender, male students ($t(34) = -2.90$, $p = 0.006$, $d = -0.49$) showed a significant increase, whereas female students ($t(31) = -1.49$, $p = 0.145$, $d = -0.26$) did not exhibit a statistically significant difference.

These findings are illustrated in Figure 4.10, which shows the mean pre and post-test

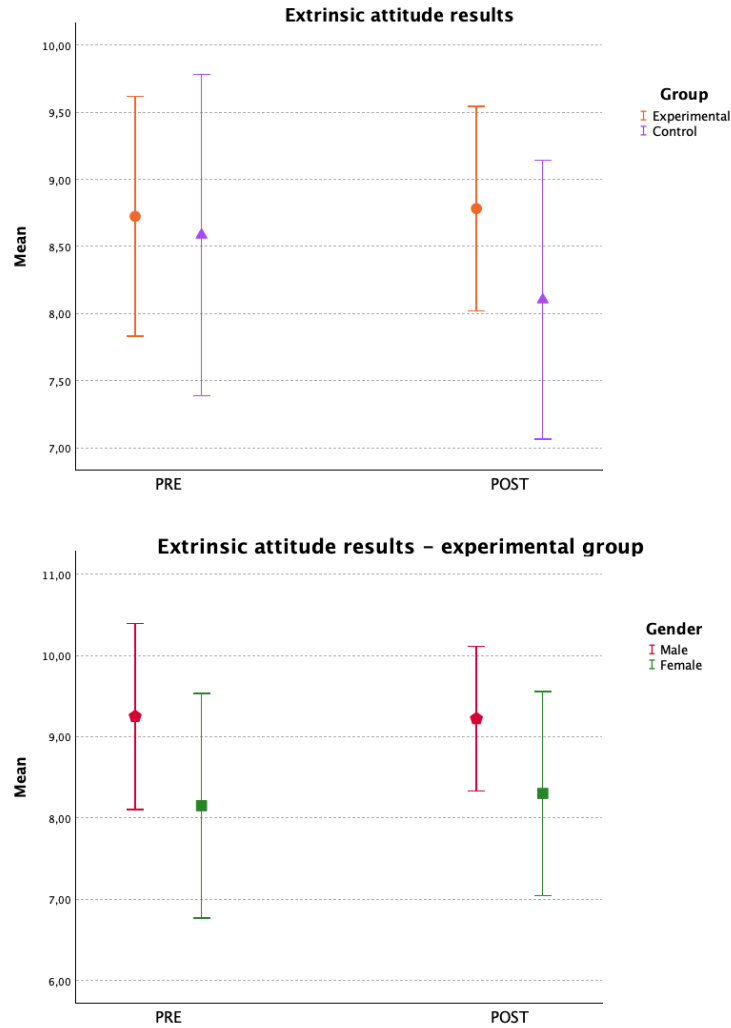


Figure 4.9 – Mean scores of extrinsic attitude divided in groups (top) and gender of the experimental group (down). Error bars are equal to 2SE, corresponding approximately to 95% confidence intervals.

scores.

The overall model of the ANCOVA was significant ($F(4, 92) = 16.58, p < 0.001$), explaining approximately 42% of the variance in post-test scores ($R^2 = 0.42$, adjusted $R^2 = 0.39$). Pre-test scores were a significant predictor of post-test performance ($F(1, 92) = 62.00, p < 0.001$), while neither the main effect of group ($F(1, 92) = 0.04, p = 0.850$) nor that of gender ($F(1, 92) = 0.43, p = 0.515$) reached statistical significance.

Taken together, these results suggest that both the game-based and the OST activity didn't substantially modify students' intrinsic or career-oriented attitudes toward astronomy. However, both experiences contributed to a more positive perception of the discipline's broader societal value.

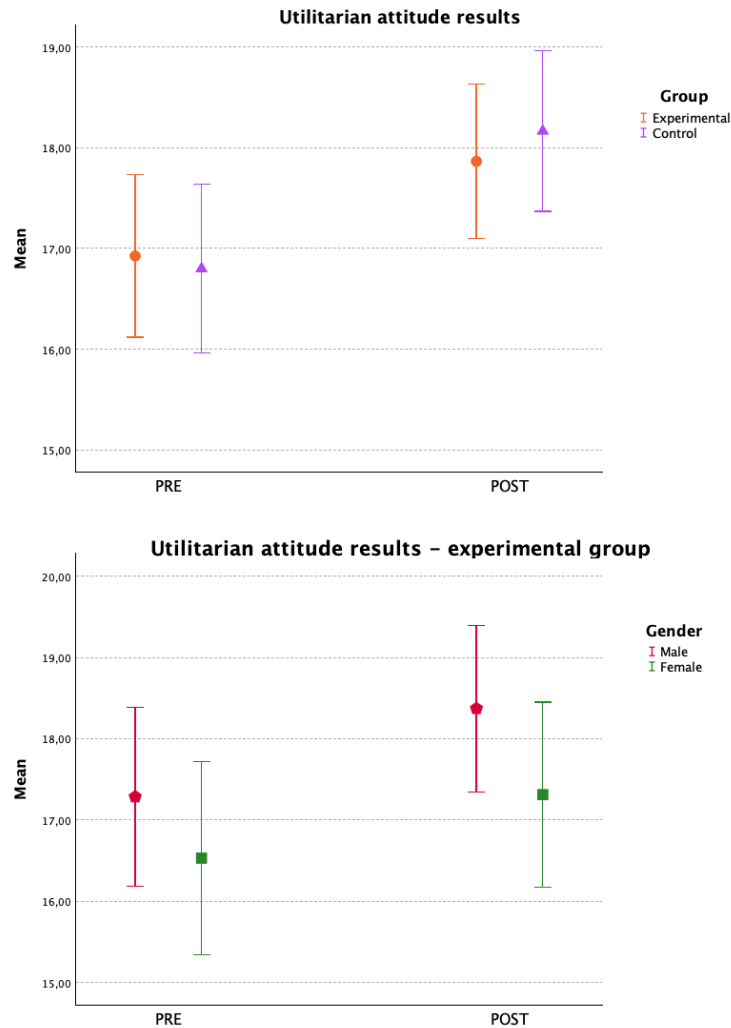


Figure 4.10 – Mean scores of utilitarian dimension of attitude divided in groups (top) and gender of the experimental group (down). Error bars are equal to 2SE, corresponding approximately to 95% confidence intervals.

Self-efficacy in astronomy

Considering self-efficacy, for the experimental group ($N = 70$), the difference between pre-test (10.70 ± 0.30) and post-test (10.34 ± 0.31) is not statistically significant ($t(69) = 1.41$, $p = .163$, $d = 0.17$). Similarly, the control group ($N = 31$) showed no significant difference between pre-test (8.90 ± 0.59) and post-test (9.45 ± 0.47 , $t(30) = -1.03$, $p = 0.309$, $d = -0.19$). Within the experimental group, neither males ($t(36) = 1.42$, $p = .165$, $d = 0.23$), nor females ($t(32) = 0.51$, $p = 0.612$, $d = 0.09$), exhibited significant differences between pre and post-test scores.

These results are illustrated in Figure 4.11, which shows the mean values for the self-efficacy dimension. The graph shows a slight discrepancy between the two groups, with

the experimental group displaying higher values than the control group both in the pre and post-test.

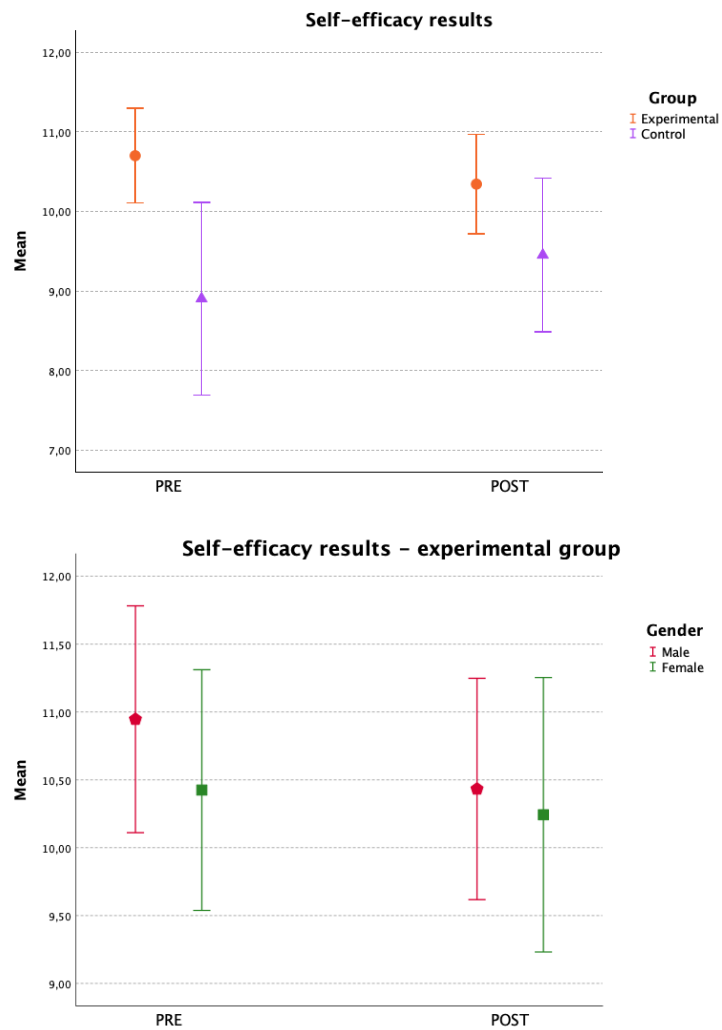


Figure 4.11 – Mean scores of self-efficacy dimension divided in groups (top) and gender of the experimental group (down). Error bars are equal to 2SE, corresponding approximately to 95% confidence intervals.

The overall model resulted by the ANCOVA was significant ($F(4, 95) = 15.26, p < 0.001$), explaining approximately 39% of the variance in post-test scores ($R^2 = 0.39, R^2_{\text{adjusted}} = 0.37$). Pre-test scores were a significant predictor of post-test performance ($F(1, 95) = 55.00, p < 0.001$). However, the main effects of group ($F(1, 95) = 0.47, p = 0.494$), and gender ($F(1, 95) = 0.49, p = 0.486$), were not statistically significant.

Teachers' perception of the activity

Concerning the interview conducted with three teachers of the experimental group, all of them highlighted a high level of student engagement, particularly in classes that are usually less motivated. Terms such as *interest*, *participation*, *enthusiasm*, and *focus* were mentioned, describing how the game successfully captured students' attention and maintained their involvement throughout the session. It is worth noting that students played in pairs, a setup that encouraged communication and mutual support.

A second recurring element concerns collaboration and social interaction. Teachers observed strong teamwork and cooperative behaviors, noting that even typically shy or reserved students took on a more proactive role. As reported by one teacher:

I noticed great interest and participation in a class that, normally, isn't always so interested in the proposed learning activities.

Only one teacher reported not observing anything particularly new in the students' interactions over the usual dynamics, yet still emphasizing the strong sense of cooperation and group cohesion that the game fostered "as if they were used to playing a team game together" (cit).

Regarding the skills and competences activated during the activity, teachers mentioned team building, cooperative learning, collaboration, risk-taking, trial and error, and the ability to assume specific roles within their teams. One teacher, in particular, described spontaneous role allocation and the emergence of leadership, often driven by curiosity rather than external pressure:

The students attempted to implement strategies to achieve their desired goal through trial and error. They were able to do so without pressure, entirely spontaneously, driven by a healthy curiosity.

As for the post-activity interactions, the three teachers expressed different views. One reported no follow-up discussions, another noted that students talked about the activity showing satisfaction, while a third perceived some frustration due to the limited time available for play. These differences also reflect the teachers' perceptions of possible connections between the game and curricular content: only one teacher believed that the more attentive students might have made conscious links to what they study in class, whereas the other two expressed negative reflections in this regard.

Finally, two teachers expressed their intention to repeat the experience with other classes, emphasizing its potential for Civic Education – a cross-curricular subject of the Italian high school curriculum designed to promote students' awareness of citizenship – and for team building. They also acknowledged, however, that time constraints in regular school routines could limit the feasibility of implementing such activities more broadly.

4.2.4 Discussion

Regarding RQ1 (Can a single gameplay session of PIXEL, compared to an out-of-school activity, foster students' understanding of the social-institutional aspects of science?) quantitative analyses revealed that, overall, no statistically significant differences emerged between pre and post-test scores. Students' understanding of the social-institutional aspects of science (measured by the SIQ) remained stable in both the experimental and control groups. On the other hand, the qualitative analysis of the open-ended question provided complementary insights that help interpret the quantitative results. Although changes were less pronounced than those observed in the extended GBL program described in Section 4.1, the data suggest a similar trend. Specifically, responses from the experimental group showed a shift from a purely cognitive-epistemic representation of science toward a broader understanding that includes social and institutional dimensions. This indicates that even a single session of PIXEL can prompt students to perceive science not merely as a set of methods and knowledge, but as a human enterprise. The smaller magnitude of change compared to previous findings is consistent with the more limited scope of the intervention, reaffirming that the duration and contextual integration of the game are crucial to achieving deeper conceptual transformations.

Concerning RQ2 (Can a single gameplay session of PIXEL, compared to an out-of-school activity, positively influence their attitudes toward astronomy and their self-efficacy in this domain?), no significant changes were observed in intrinsic and extrinsic attitudes toward astronomy, or in self-efficacy levels, both in the experimental and in the control group. A slight increase was detected in the utilitarian attitude dimension across both groups. These findings align with expectations, considering the short duration of the experience and the small scale of the sample: modifying constructs such as attitudes and self-efficacy typically requires prolonged exposure and repeated reinforcement, whereas a single gameplay session or OST activity primarily serves as an introductory stimulus.

The results of the teacher interviews further reinforce the pedagogical potential of GBL in formal education contexts. Teachers highlighted the high levels of engagement and motivation observed among students, including those who are typically less active or interested in traditional lessons. The cooperative and interactive nature of PIXEL fostered teamwork, communication, and problem-solving skills, while also promoting positive emotional states and a sense of agency. Teachers also emphasized the game's value as a tool for Civic Education lecture and team building, although they noted practical challenges such as time constraints and the difficulty of linking the activity directly to curricular content.

In conclusion, the results indicate that the evaluating tools employed were not sensitive enough to detect potential changes elicited by a single PIXEL gameplay session. In contrast, the qualitative instruments specifically developed for this study revealed some variations. Although the project involved a small sample and a considerable compositional

bias between the two groups, it represents a valuable starting point for a broader and more comprehensive investigation.

4.3 Conclusion

This chapter investigated what PIXEL can elicit, and under which conditions, in terms of students' understanding of science and their affective engagement with it. By adopting two complementary research designs, the chapter tailored effectiveness and clarified how different GBL approach can promote the application of PES principles within formal educational context.

When PIXEL is embedded within an extended and structured GBL program, the results showed clear transformations in students' representations of science. Through repeated gameplay, facilitation experiences, and explicit reflective practices, students progressively shifted from predominantly procedural and method-centered views of science toward more articulated conceptions that explicitly included its social and human dimensions. This change was evident not only in the content of students' responses, but also in their expressive register, which evolved from formal and textbook-like formulations to more personal and experience-based. These findings support the idea that learning about the Nature of Science requires time, iteration, and explicit reflection, and that games such as PIXEL are most effective when integrated into a broader pedagogical ecology.

The single-session study offered a deliberately constrained perspective, aimed at exploring what can emerge from a single gameplay session. As expected, no statistically significant effects were detected on quantitative measures of social-institutional understanding, attitudes toward astronomy (except for the utilitarian dimension), or self-efficacy when compared to an OST activity. These results confirm that constructs such as attitudes and self-efficacy are relatively stable and unlikely to change following brief experiences. At the same time, qualitative evidence from open-ended responses and teachers' observations revealed signs of increased engagement, collaboration, and attention to the human and social dimensions of scientific work. Although modest, these signals suggest that even a single gameplay session can activate alternative ways of thinking about science, which may represent necessary precursors to deeper conceptual change.

Some limitations must be acknowledged. The extended GBL program involved a self-selected group of students already interested in science, which may have amplified its effects. Conversely, the single-session study relied on heterogeneous classroom samples and on assessment tools not specifically designed to capture subtle shifts generated by GBL. Moreover, the exploratory and qualitative nature of part of the analysis limits the generalizability of the findings. These limitations, however, reflect broader methodological challenges in evaluating complex educational interventions.

Overall, this chapter provides evidence that the educational value of PIXEL does not reside solely in the game itself, but in the conditions under which it is implemented. PIXEL functions as a mediating artifact that makes visible and experientially accessible the social and human dimensions of scientific research, following the principles of PES. When supported by time and reflection, it can contribute to the development of richer and more realistic understandings of science.

Chapter 5

Cosmo Hunters: from game mechanics to learning outcomes

The second case study addressed in this thesis is Cosmo Hunters, a card game that aims to introduce the topic of astrophysical observations of celestial objects at different frequencies of the electromagnetic spectrum to upper primary and middle school students (Toniolo et al., 2025). The description of the game mechanics is reported in Section 3.2, while Table 5.1 summarizes the correspondence between the game mechanics and the scientific aspects (SA) they aim to represent.

Scientific Aspects (SA)	Game Mechanic
SA1. There are different kinds of light (frequencies of the electromagnetic spectrum) emitted by objects.	Different types of Observation cards represent emissions at specific electromagnetic frequencies.
SA2. Different kinds of telescopes are used to observe celestial objects.	For each frequency category, there is a corresponding Telescope card required to score points.
SA3. The same celestial object can emit different kinds of radiation simultaneously.	Extra points are awarded when Observation cards of the same object at different frequencies are collected together.

Table 5.1 – Correspondence between the scientific aspects represented and the game mechanics.

Within the broader framework of PES outlined in Chapter 1, Cosmo Hunters was designed to support learning outcomes while fostering engagement and active participation among students. Rather than aiming at experiencing the social-institutional aspect of NOS as PIXEL did, this case study focuses on disciplinary learning as a foundational component of meaningful engagement with astrophysics.

In this context, the GBL activity is compared to a frontal lecture addressing the same

topics and the effectiveness is operationalized in terms of learning outcomes. Therefore, the overarching research question is tailored as follows:

To what extent does Cosmo Hunters, compared to a traditional frontal lecture, support students' learning outcomes related to astrophysical observations across the electromagnetic spectrum?

5.1 Sample and research design

The sample is composed of 175 students in the last year of an Italian lower secondary school (aged 13–14). The school is located in the city center of Florence, in an area that is neither particularly wealthy nor disadvantaged, with approximately 12% of students coming from immigrant families. The decision to limit the study to this age target was made considering the Italian school curriculum, in which the topic of light is briefly introduced at that age but still not deeply explored. Moreover, this target age is considered at higher risk for school drop-out, as reported by the Italian Ministry of Education ([Ministero dell'Istruzione e del Merito, 2024](#)). The intervention aimed to involve students who are in a critical stage of their educational paths, during which they have to choose which high school is best suited for them.

The students were randomly divided into two groups, as reported in Table 5.2: the lesson group (control) and the game group (experimental).

Group	Female	Male	Total
Lesson (control)	40	36	76
Game (experimental)	53	46	99

Table 5.2 – Composition of the two experimental groups.

A two-hour intervention was delivered to each group, as represented in Figure 5.1. For the game group, a GBL activity using Cosmo Hunters was implemented, including a debriefing session. For the lesson group, a frontal lecture covering the same educational content as the game was conducted. For both groups, a pre-test was administered before starting the activity, a post-test at the end of it, and a follow-up test four weeks later.

For the game group, the GBL activity was organized into two phases. In the first phase, participants engaged in two or more gameplay sessions of Cosmo Hunters. Students played in pairs, an effective approach in preventing them from feeling intimidated by the activity and fostering the exchange of ideas and strategies, leading to a deeper processing and understanding of the topics ([Ligabue, 2020](#)). The aim of the second phase – the 45-min debriefing session – was to reinforce the knowledge conveyed by the game and to foster discussion among students. The role of the facilitators was to pose guiding questions

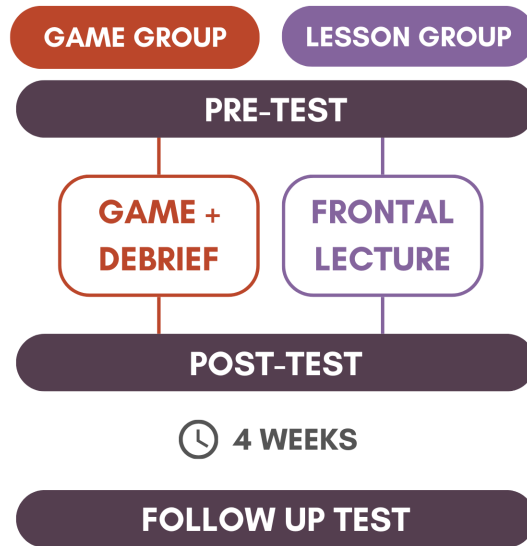


Figure 5.1 – Timeline of the activity sessions conducted during spring 2024.

to stimulate dialogue and guide students toward finding the answers by themselves. The questions used in this phase are reported in Section B.1 in the Appendix.

On the other hand, the lesson group attended a frontal lecture delivered by astrophysics outreach experts. The lecture, whose structure is reported in Section B.2 in the Appendix, covered the same topics as Cosmo Hunters, with the use images identical to those displayed on the game cards. It is worth noting that the format allowed for a deeper exploration of specific aspects of the content compared to the GBL activity.

The pre-test was administered to both groups right before the activity and aimed to assess students' prior knowledge of the electromagnetic spectrum and astrophysical observations. By comparing its results with those of the post-test (compiled at the end), the evaluation sought to assess learning outcomes. The long-term impact was assessed through the follow-up test four weeks after the intervention. Table 5.3 summarizes the questions of the three questionnaires, while Figure 5.2 reports the images displayed in it. Each questionnaire required approximately 15 minutes to complete.

The images used in Q1 were selected based on existing literature on students' misconceptions about light and vision, as these often lead learners to believe that the eye or a mirror emit light (Besson, 2015). The correct images in question Q3 and those in question Q4 are the same as those displayed in the game and shown during the frontal lecture.

Question	Type	SA	Test
E: On a scale from 1 (not at all) to 7 (very much), how much did you enjoy the activity?	Likert scale	–	Post
Q1: Which of these objects emit light? (Fig. 5.2a)	Multiple choice	SA1	Pre; follow-up
Q2: List the types of light you know. In the pre-test, this question was linked to Q1 (Specify what type of light the selected objects emit).	Open-ended	SA1	Pre; post; follow-up
Q3: Which of these instruments are telescopes? (Fig. 5.2c)	Multiple choice	SA2	Pre; post; follow-up
Q4: What do the three images shown here represent? (Fig. 5.2c) (A) All the same nebula, but observed at different frequencies (B) A galaxy, a nebula, and a black hole (C) All galaxies, but of different types (D) All the same nebula but observed at three different stages of its evolution.	Single choice	SA3	Pre; post; follow-up

Table 5.3 – List of questions presented in different tests and addressing the scientific aspects (SA) covered by the game mechanics.



(a) Q1: Which of these objects emit light? (6 correct)



(b) Q3: Which of these instruments are telescopes? (5 correct)



(c) Q4: What do the three images shown here represent?

Figure 5.2 – Images used in the questionnaires.

5.2 Data analysis

Data collected from the three tests were analyzed using a combination of quantitative and qualitative methods. For the details of collected data see Section B.3 in the Appendix. The engagement level of participants was analyzed using descriptive statistics, and the average scores were compared to determine differences in perceived engagement between the game and lesson groups.

For the multiple-choice questions with numerical scoring (Q1 and Q3), responses were coded by assigning a value of 1 to each correct option that was selected and to each incorrect option that was not selected, while a value of 0 was assigned to each incorrect option selected or correct option not selected. Table 5.4 provides an illustrative scheme of coding mechanism.

	Selected	Non selected
Correct option	1	0
Incorrect option	0	1

Table 5.4 – Coding scheme used for multiple-choice questions (Q1 and Q3).

The individual scores for each option were then summed to obtain a total performance score per question for each student, ranging from 0 to 10 for both Q1 and Q3.

To compare performance between groups (lesson and game) and testing phases (pre, post and follow-up), a multiple regression analysis was applied. This statistical method, widely used in educational research, is based on the same principles as analysis of covariance (ANCOVA, see Section A.2 in the Appendix for further details). It builds a predictive linear model of the results as a function of the testing phase and the group. Through this model, the estimated marginal means were calculated, representing the average predicted scores for each group and phase, as if all groups had started from identical initial conditions. In practical terms, they show what the mean score for each group would be if all groups started from the same initial conditions.

For the single-choice question (Q4), responses were analyzed using a logistic multiple regression, a variant of multiple regression employed when the dependent variable is categorical rather than continuous. In this case, the outcome variable was binary, since each response could only be either correct (1) or incorrect (0). Logistic regression models the relationship between categorical outcomes and a set of independent variable (testing phase and group). Unlike standard linear regression, which predicts a continuous value, logistic regression estimates the probability of the occurrence of a specific categorical event – in this case, the probability that a student gives a correct answer. The output represents the likelihood, expressed as a value between 0 and 1, that a student belonging to a given group would respond correctly at a specific testing phase.

Both statistical models were applied to account for the influence of the testing phase and group on the final outcomes, thus providing a more robust and reliable analysis than a calculation of mean values.

The open-ended question (Q2) was evaluated qualitatively through keywords analysis, identifying the presence of scientifically relevant terms such as *radio*, *infrared*, *visible*, *ultraviolet*, *x-rays*, and *gamma rays*. The proportion of responses containing one or more of these keywords was then quantified to determine the percentage of correct or partially correct answers. Responses listing unrelated types of light (e.g., *natural/artificial light*, *thermal light*, *LED*) were classified as incorrect.

5.3 Results

This section presents the results of the pre-, post-, and follow-up questionnaires, divided in engagement assessment and learning outcomes.

5.3.1 Engagement

In the post-test, students were asked about their perceived enjoyment of the activity. Figure 5.3 shows the distribution of responses on a 7-point Likert scale, where the lowest value (1) corresponds to “not at all enjoyed” and the highest value (7) to “very much enjoyed”.

The distribution for the game group shows an average enjoyment score of 6.0, with more than 75% of students selecting high enjoyment values between 6 and 7. Conversely, the lesson group had a lower average enjoyment score (5.3), with more than 80% of responses between 5 and 6, and fewer than 4% selecting “very much enjoyed” (7). These results proved that the game activity was perceived as more engaging than the frontal lecture.

5.3.2 Learning outcomes

The second aim of the evaluation was to assess the effectiveness of Cosmo Hunters in terms of learning outcomes, through the use of questions developed ad-hoc (Q1, Q2, Q3, Q4 in table 5.3).

Questions Q1 and Q2 addressed the existence of different frequencies of the electromagnetic spectrum emitted by objects. The game group experienced this concept through the various frequency categories depicted in the cards (Q2). The same content was also covered in the lecture delivered to the lesson group. Furthermore, during both debriefing sessions, reflections on how some daily-life objects emit different frequencies (Q1) were encouraged, using the same images displayed in the questionnaire.



Figure 5.3 – Answers to the question: “On a scale from 1 (not at all) to 7 (very much), how much did you enjoy the activity?”

The goal of Q3 was to determine whether students paid attention to the images of telescopes presented through the Telescope cards in the game or during the lecture. Similarly, Q4 tested whether students understood the game mechanics related to multifrequency observation. The images shown in these questions (Figure 5.2) were the same as those discussed during the lecture and represented in the game cards.

Figure 5.4 shows the evolution of the estimated marginal mean of scores between the pre-test and follow-up for Q1. In the pre-test, the mean score was 7.08 ± 0.16 for the game group and 6.98 ± 0.18 for the lesson group. In the follow-up, they both achieved a score of 8.38 ± 0.16 and 8.38 ± 0.18 , respectively. No statistically significant differences were observed between the two groups in either phase, as both mean values fell within the other’s standard error range.

Figure 5.5 presents the results for Q2. The lesson group achieved 5.7% correct or partially correct answers in the pre-test, 54.4% in the post-test, and 50.0% in the follow-up. The game group achieved 9.3%, 81.8%, and 78.9%, respectively. The initial difference between the two groups performance was minimal (3.6%), indicating comparable baseline knowledge. However, the game group outperformed the lesson group both in the post-test (+27.4%) and in the follow-up (+28.9%), demonstrating a stronger and more persistent learning effect. A slight decline was observed between the post-test and follow-up for both groups (-4.4% for the lesson group and -2.9% for the game group), which was expected given the four-week interval between the two assessments.

For Q3 (Figure 5.6), the estimated marginal mean of scores in the pre-test was 6.75 ± 0.17 for the game group and 6.40 ± 0.19 for the lesson group. In the post-test, the scores were 8.05 ± 0.17 and 8.77 ± 0.19 , and in the follow-up 8.63 ± 0.17 and 8.48 ± 0.19 , respec-

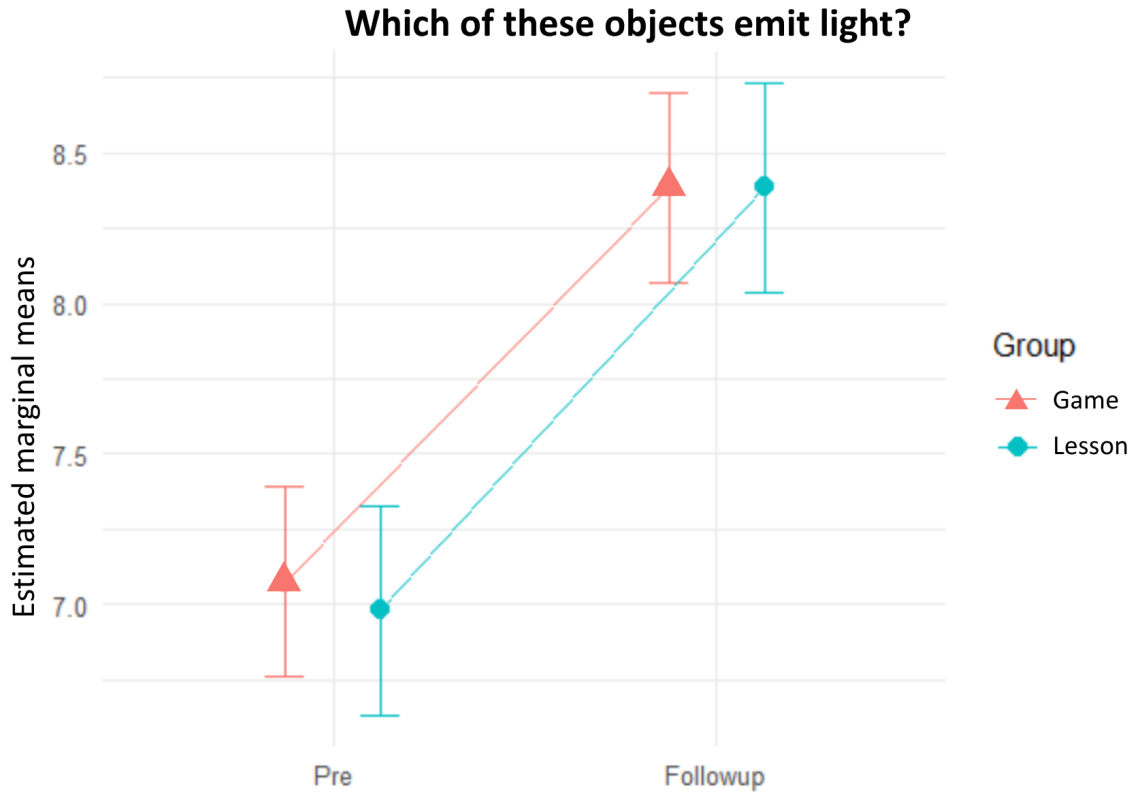


Figure 5.4 – Estimated marginal means over time for Q1. The error bar represents twice the standard error.

tively. Both groups showed similar prior knowledge regarding different kinds of telescopes, within the error. After the activity, the lesson group performed better at recognizing telescopes, likely due to the more detailed analysis of telescope types during the lecture. However, after four weeks (follow-up), the game group's performance improved and became comparable to that of the lesson group.

The predicted probability for Q4 (Figure 5.7) in the pre-test was 0.51 ± 0.07 for the game group and 0.24 ± 0.07 for the lesson group. In the post-test, the probabilities were 0.84 ± 0.05 and 0.94 ± 0.03 , and in the follow-up 0.73 ± 0.06 and 0.84 ± 0.06 , respectively. Therefore, the game group demonstrated greater prior knowledge of the images shown in Q4, as reflected by higher pre-test scores. After the activities, both groups increased and performed comparably within the standard error.

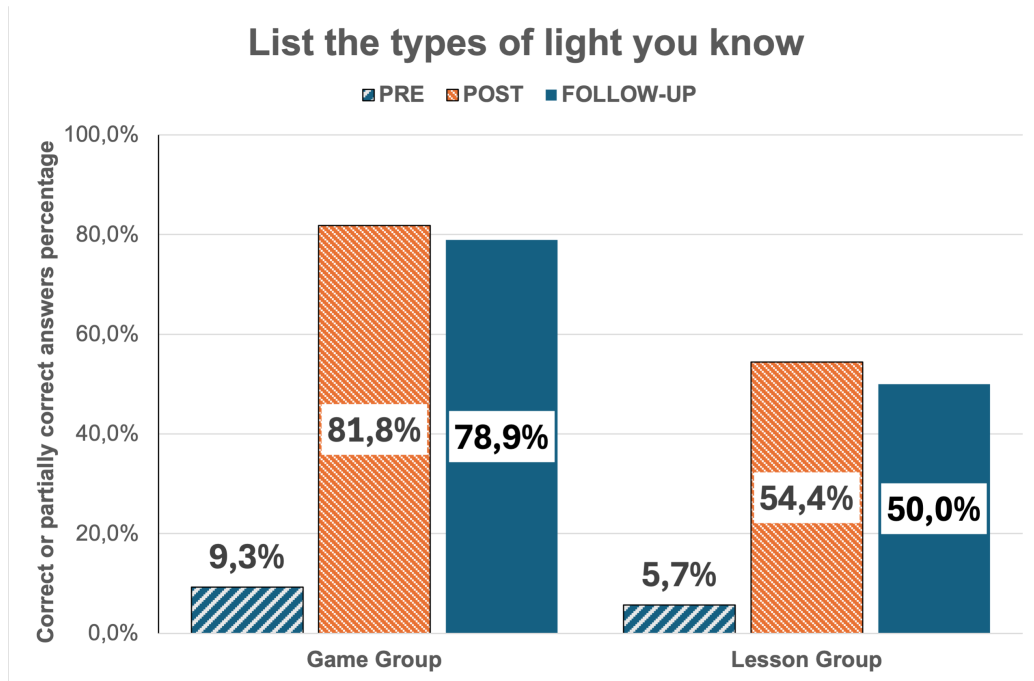


Figure 5.5 – Percentage of correct or partially correct answers to Q2.

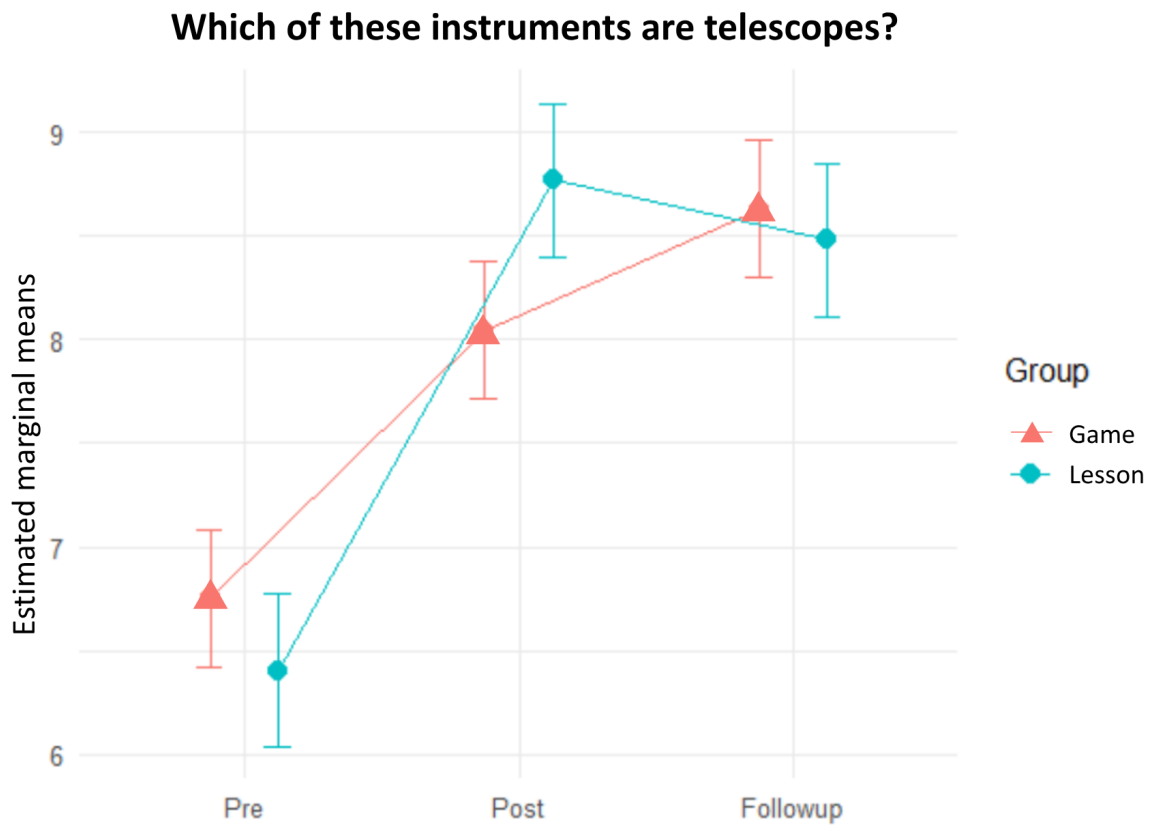


Figure 5.6 – Estimated marginal mean over time for Q3. The error bar represents twice the standard error.

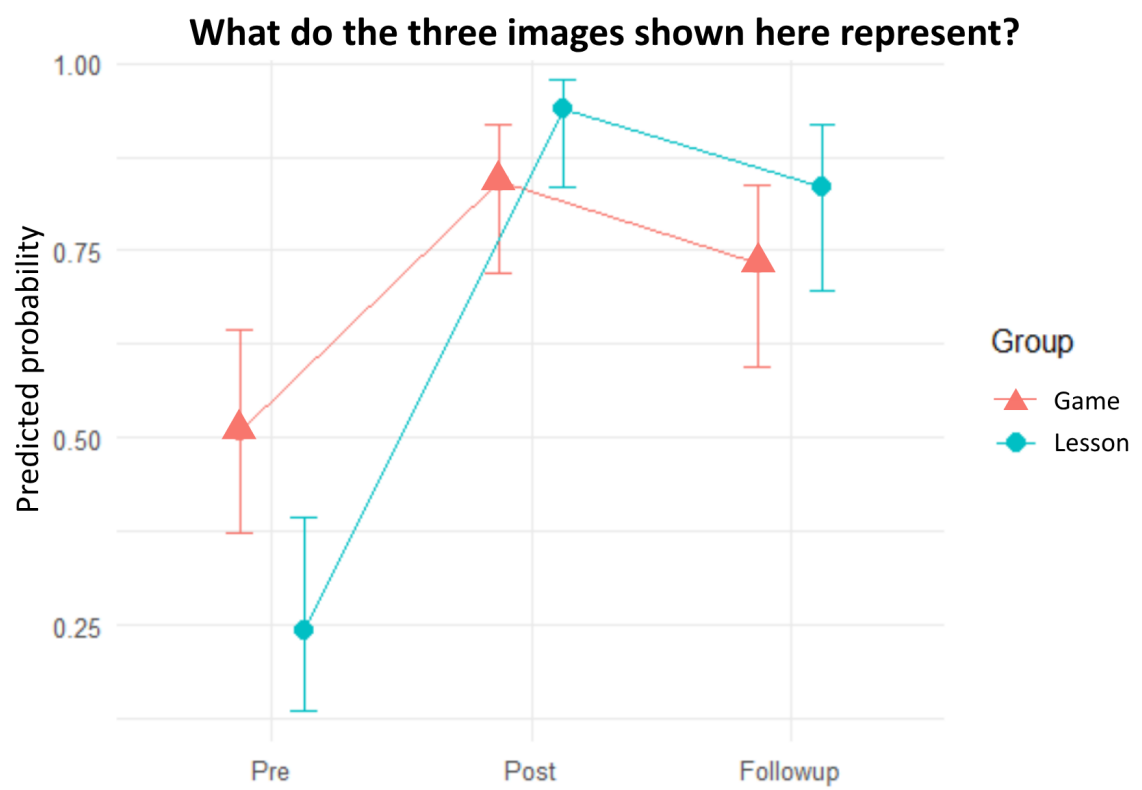


Figure 5.7 – Predicted probability over time for Q4.

5.4 Discussion

This study aimed to address the following research question: to what extent does Cosmo Hunters, compared to a traditional frontal lecture, support students' learning outcomes related to astrophysical observations across the electromagnetic spectrum?

As expected, the GBL activity proved to be more engaging than a traditional frontal lecture, both in terms of average value and distribution. This outcome is consistent with the implementation of GBL, which use instrument (games) that are intrinsically perceived as fun and enjoyable. On the other hand, both the activity (game and frontal lecture) was perceived as something new and different from the usual classroom routine, and therefore appreciated also for its novelty effect.

Regarding the scientific aspects embedded within Cosmo Hunters, the results reported in Section 5.3 indicate that, overall, the GBL activity was able to promote learning with outcomes comparable to those of a frontal lecture. However, some differences emerged between the three scientific aspects. For the nomenclature of electromagnetic spectrum frequencies (SA1), the game group produced a higher percentage of correct or partially correct answers than the lesson group. This result can be linked to the intrinsic mechanics of Cosmo Hunters that required players to familiarize themselves with the terminology of the electromagnetic spectrum in order to play effectively. This helped students to retain the terminology more firmly and to associate it more deeply with the concept of the electromagnetic spectrum, which was later reinforced during the debriefing, compared to the lecture format.

A different pattern was observed for the images of telescopes (SA2). The lesson group performed better than the game group immediately after the activity, while the predicted scores were similar (no statistical difference) in the follow-up. This difference is likely due to the structure of the lecture, which allowed more time to visualize various telescopes used in astrophysical observation. In contrast, in the GBL activity, both the gameplay and the debriefing focused more on the necessity of using different types of telescopes to observe specific types frequencies rather than on the detailed characteristics or names of these telescopes, mainly due to time constraints.

For the understanding of multifrequency observation (SA3), the two groups performed comparably both in the post-test and in the follow-up. It is worth noting that SA3 in the GBL activity was conveyed more through a strategic choice than a core game mechanic. Players could choose whether to collect Observation cards of the same object at different frequencies to gain more points or to adopt alternative strategies. Thus, even without an explicit rule pushing towards multifrequency observation, the strategic need to maximize points through such combinations facilitated a deeper understanding of this astrophysical concept, leading to results comparable to those obtained through the frontal lecture.

Overall, Cosmo Hunters proved to be an effective tool to engage and promote learning

among 13-year-old students in educational contexts on topics related to the electromagnetic spectrum and astrophysical observations. The game represents a valid alternative to a traditional lecture, offering higher engagement perceived by participants and requiring limited effort for implementation.

In particular, the results suggest that the scientific aspects stimulated different levels of learning according to how they were implemented within the game design. SA1 was implemented as a core game mechanic. This represents the primary level of engagement and access to the game, and for this reason, the terminology of the electromagnetic spectrum used in this mechanic had a stronger impact on students' learning. SA2 was implemented both as a mechanic (different telescopes required for observation) and as an aesthetic feature (different images for the telescopes). Consequently, while the mechanic component of SA2 attracted students' attention, the aesthetic component explored in Q3 appeared to be less impactful. SA3 was implemented as a strategic choice, requiring a deeper understanding of the game's dynamics to be fully perceived. Nevertheless, the simplicity of Cosmo Hunters enabled a learning level in this topic comparable to that achieved through the frontal lecture.

This hierarchy of results aligns with the literature suggesting that learning outcomes can be achieved more effectively through game mechanics rather than through explicit content alone (Bayeck, 2020; Yoon and Khambari, 2022).

5.5 Conclusion

This study assessed the potential of Cosmo Hunters to engage 13-year-old students in educational contexts and to promote learning in three scientific aspects addressed by the game: (SA1) the existence of different frequencies of the electromagnetic spectrum, (SA2) the necessity of using different types of telescopes for astrophysical observations, and (SA3) the possibility of observing the same astrophysical object at different frequencies to obtain more information. The engagement and learning outcomes of Cosmo Hunters were compared to those of a traditional frontal lecture, and their long-term endurance was assessed after four weeks.

Two groups of students participated in the study: the game group and the lesson group. The game group (99 students) attended a GBL activity composed of gameplay sessions followed by a debriefing, whereas the lesson group (76 students) attended a frontal lecture covering the same content as the game. Data were collected through three questionnaires administered before the activity (pre-test), immediately after (post-test), and four weeks later (follow-up).

Results showed that the GBL activity was perceived as more engaging than the frontal lecture, with an average enjoyment of 6.0/7 for the game group and 5.3/7 for the lesson

group. The analysis of students' learning in the scientific topics produced different outcomes for the three aspects addressed by the game. SA1, implemented as a core game mechanic, stimulated more learning outcomes than the traditional lecture. SA2, introduced through an aesthetic features of the game, was more effectively learned through the lecture. SA3, addressed as a strategic component rather than a game mechanic, required a deeper understanding of the game's dynamics to emerge, yet it still resulted in learning outcomes comparable to those of the lecture.

These findings indicate that Cosmo Hunters represents an effective and accessible tool for introducing the basic concepts of the electromagnetic spectrum and demonstrating how astrophysical observations exploit it to study celestial objects, while exploiting the potential of a PES informed approach. Owing to its short gameplay duration (15-20 minutes), the game can also be used as a follow-up or flexible-use activity in the classroom, without requiring extensive preparation or teacher intervention, reinforcing learning in an enjoyable and interactive environment.

The game is freely available, along with a print&play version¹ that allows educators to reproduce the activity and a guide for facilitating the GBL activity.

¹Available here: play.inaf.it/en/cosmo-hunters-en/

Chapter 6

Conclusion

This thesis stems from the necessity to assess to what extent, and through which dimensions, PES activities based on GBL can be evaluated as effective in the context of astrophysics education.

Starting from the evolution of science communication models – from the deficit-oriented logic of the Public Understanding of Science to the participatory paradigm of Public Engagement – this work proposes game-based and playful learning methodologies as a concrete way to operationalize PES principles within science education. To this end, two board-game case studies were analyzed: PIXEL, which aims to introduce the human and social dimensions of astrophysics research, and Cosmo Hunters, which focuses on the fundamental concepts underlying astrophysical observations.

The development of PIXEL involved a multidisciplinary team composed of astrophysicists, public engagement experts, and professional game designers. This collaboration ensured that the game was enjoyable and a faithful representation of the complexity of scientific research. A first evaluation focused on engagement and on the scientific content that could be conveyed through a single gameplay session. Conducted with over 200 participants from different backgrounds, the study revealed high levels of engagement. Participants were able to recognize the relationship between image resolution, observation time, and technological development, and identified collaboration among different professional roles as essential for the functioning of a research center.

Cosmo Hunters, by contrast, is a card game designed to introduce players to the basic concepts of the electromagnetic spectrum and to the ways astrophysicists exploit different frequencies to study celestial objects. Unlike PIXEL, which primarily addresses the social and institutional dimensions of scientific research, Cosmo Hunters emphasizes the procedural and conceptual practices of astrophysics.

Since the two case studies addressed different audiences and learning outcomes, the research question was tailored to the specific context of each intervention. In the case of PIXEL, the impact of the game was investigated both within a long-term, 40-hour GBL

program and as a single gameplay session.

In the extended program, 32 high school students participated in a pathway combining repeated gameplay sessions with explicit reflective activities. Results showed a significant shift in students' conceptions of science, from formal and predominantly cognitive-epistemic views to more articulated representations that included social and institutional dimensions of scientific practice. This demonstrates that, when embedded in a broader educational framework, PIXEL can effectively promote a deeper and more nuanced understanding of science as a human and collective endeavor.

In the second PIXEL study, the game's impact was examined in a formal educational setting through single-play sessions, compared with an out-of-school-time activity addressing the same topic. Although the short intervention naturally produced limited effects than the extended program, students who played PIXEL still demonstrated a better grasp of the social aspects of scientific work than those who attended the OST activity. The same study also included an evaluation of dimensions such as self-efficacy and attitudes toward astronomy using validated instruments. No significant changes were observed in either group, a result that can be attributed to the limited sensitivity of tools not specifically calibrated for capturing the nuances of a single gameplay experience. This finding highlights the need for future studies to develop assessment instruments specifically designed for short-term GBL interventions. At the same time, interviews with teachers provided complementary insights, highlighting high levels of engagement and motivation among students, including those typically less involved in traditional lessons. Teachers also reported the activation of transversal competences such as teamwork and problem solving.

The evaluation of Cosmo Hunters focused on engagement and learning outcomes through a comparative study between a game-based activity and a traditional frontal lecture on the same topics, involving 175 students aged 13. Results showed that, overall, both approaches were similarly effective in terms of knowledge acquisition, demonstrating that the game conveyed scientific content as successfully as the lecture. However, Cosmo Hunters was rated as more enjoyable by participants, confirming the potential of game-based methodologies to enhance motivation and interest without compromising learning outcomes.

Despite the relevance of the findings, some limitations must be acknowledged. First, the sample sizes involved in the studies are relatively small, particularly in the extended GBL program based on PIXEL, in which participation was also voluntary and partly mediated by teachers, resulting in a self-selection bias. While this context was appropriate for exploring the potential of the game within a complex educational framework, it limits the generalizability of the results.

A second limitation concerns the assessment instruments used in the short-term PIXEL study. Although validated tools were employed to measure constructs such as self-efficacy and attitudes, these instruments were not specifically tailored to the characteristics of the PIXEL experience. As a result, they may not have been sensitive enough to capture sub-

tle changes elicited by a single gameplay session. This highlights a broader challenge in the evaluation of game-based public engagement activities, where traditional quantitative instruments may fail to fully reflect context-specific impact.

Future research should therefore focus on the development and validation of quantitative assessment tools explicitly designed for game-based public engagement with science. Such instruments should be capable of being administered to larger samples, while remaining meaningful to the educational context in which the activities take place. Combining context-sensitive quantitative measures with qualitative approaches would allow for more robust and scalable evaluations, supporting a deeper understanding of how GBL-based PES activities can be considered effective.

Chapter 7

AMACA: the challenge of multisensoriality

Grounded on the results obtained so far, this thesis paves the way for another future development: the challenge of multisensoriality. In fact, as already anticipated in Section 2.2, astrophysics is commonly perceived as a visual science, largely because its public communication relies on spectacular images. Despite the fascination that such representations evoke, this visually oriented approach naturally excludes audiences who have sensory disabilities. On the other hand, methodologically speaking, astrophysics is inherently multisensory because its knowledge production relies on the transformation of non-visible signals across the electromagnetic spectrum – as well as other messengers such as gravitational waves and neutrinos – into visual representations. This concept open to the possibility of a paradigm shift mirroring the one from PUS to PES, aimed at making visible the processes and methodologies behind stunning results, while accounting for accessibility to everyone.

Following this research strand, in recent years innovative approaches have been developed to make astrophysics communication multisensory ([Pérez-Montero, 2019](#); [Varano and Zanella, 2023](#); [Noel-Storr and Willebrands, 2022](#); [Foran et al., 2022](#); [Harrison et al., 2022](#); [Varano and Di Giacomo, 2024](#)). An increasing number of projects focus on developing tactile resources ([Bonne et al., 2018](#); [Paredes-Sabando and Fuentes-Muñoz, 2021](#); [Arcand, 2022](#)) and/or audible resources ([Quinton et al., 2021](#); [Tomlinson et al., 2017](#); [Bieryla et al., 2020](#); [Elmquist et al., 2021](#); [García-Benito and Pérez-Montero, 2022](#); [Bardelli et al., 2021](#); [Harrison et al., 2023](#); [Guiotto Nai Fovino et al., 2024](#)). In particular, the process of turning data into sound, called “sonification”, is gaining momentum in the astrophysical community with applications in both astrophysics research and communication ([Zanella et al., 2022](#); [Lenzi and Ciuccarelli, 2020](#)). These pioneering efforts have shown that a multisensory approach makes astrophysics communication not only more inclusive but also more engaging for everyone.

While encouraging, these attempts still face limitations preventing them from becoming

widely shareable. First, many resources are being developed for specific one-time events (Bieryla et al., 2020; Elmquist et al., 2021); they often lack substantial documentation and are not easily accessible. Second, several of these resources, particularly those adopting tactile approaches, require materials that are difficult to manufacture, expensive, and/or time-consuming to produce for large audiences (Bonne et al., 2018; Usuda-Sato et al., 2019; Paredes-Sabando and Fuentes-Muñoz, 2021; Pérez-Montero et al., 2023; Arcand, 2022). Third, many resources have been created by teams of astrophysicists alone, without collaboration from multi-disciplinary experts (e.g., sound designers, psychologists) and have not been properly tested with users before being adopted by the public. Lastly, such resources are generally disseminated by the same professionals who created them, making them difficult to share widely and delivered by non-specialists (Harrison et al., 2024).

Within this context rely the AMACA project¹ (Astronomy education with a Multisensory, Accessible, and Circular Approach), that aims to develop multisensory playful learning activities that are easily replicable and adaptable to different contexts. As for the characteristics of playful learning presented in Chapter 2, the activities focus on methodologies and practices (e.g., how exoplanet transits are discovered with the light curve method; how the chemical composition of comets and planets' atmosphere are investigated), rather than presenting only spectacular results. These activities are developed within an educational path (described in detail in Section C.1 in the Appendix) that involves multiple audiences, such as PhD candidates in astrophysics, psychologists, sound designers, high school students, the general public, and primary and secondary school teachers. The interaction among these different groups follows a bottom-up and peer-to-peer approach, reflecting the two-way participation principles of PES.

The whole approach is framed within the methodology of Universal Design for Learning (UDL), a pedagogical framework aimed at fostering inclusive and accessible learning environments by placing learners' diversity – rather than an average ideal type – at the center of the design process (Savia, 2016). UDL has its roots in cognitive psychology, drawing on the theories of Piaget, Bruner, and Vygotsky, as well as in neuroscience. Its implementation is supported by a growing body of evidence-based research demonstrating its effectiveness in producing significant improvements in learning outcomes for all students, regardless of background, prior knowledge, or ability (Schelly et al., 2011).

The UDL approach is synthesized into three core instructional principles. First, it emphasizes the provision of multiple means of representation and presentation of content, allowing students to access information and knowledge through different modalities. Second, it promotes multiple means of action and expression, enabling learners to demonstrate their understanding in diverse and flexible ways. Finally, it foregrounds the use of multiple means of engagement, offering students varied opportunities to sustain motivation and actively participate in the learning process. AMACA applies these principles through the

¹See the AMACA website: universointuttiisensi.it/en/progetto-amaca-2/

development of multi-sensory activities that allows each participant to choose the mode of engagement that best suits their needs, personalizing the experience and making it more meaningful.

The preliminary evaluation conducted within the AMACA project ([Toniolo et al., 2026](#)) suggested positive impacts across different audiences. Among these, the experience involving PhD candidates is particularly relevant to this thesis, as it sheds light on the concept of learning by design within a PES-informed framework.

PhD candidates were involved through a training course in which they were guided to design playful, multisensory activities related to astrophysics, adapted for both school settings and science festivals. Although highly specialized in their scientific domains, they had limited prior experience with PES initiatives and were accustomed to top-down communication approaches such as PUS. Engaging in the design and testing of playful activities therefore required them to rethink how complex astrophysical concepts could be translated into small, accessible units of meaning and how the dynamics of scientific research could be conveyed through participatory and multisensory experiences.

Through the project recurring patterns emerged, including a progressive shift toward more focused content, increased attention to the human and collaborative dimensions of scientific work, and a more explicit consideration of the target audience. These changes reflect a broader conceptual shift from an approach focused exclusively on scientific content, much like the logic of PUS, to one centered on the human processes that lead to the creation of that content, echoing the principles of PES. In this sense, the activity's design process itself became an exercise in engagement, encouraging the PhD candidates to rethink their role as communicators and to adopt a more participatory perspective on how science can be shared with the society.

This project represents the final component of this thesis, which has explored how PES-informed activities grounded in active learning methodologies such as GBL and PL can serve as effective tools for restoring the complexity of scientific research as a human endeavor. This applies not only to school contexts – examined through the case studies of PIXEL and Cosmo Hunters – but also to higher education settings, as illustrated by the AMACA project.

Thus, the trajectory of this work culminates in a central claim: bringing PES principles into science education through GBL and PL is not only effective, but necessary. Science, like every human activity, is deeply intertwined with society and thus must be able to return to where it started – to everyone.

List of References

- Abd-El-Khalick, F. and Lederman, N. G. (2023). Research on teaching, learning, and assessment of nature of science. *Handbook of research on science education*, pages 850–898.
- Akbayrak, M. and Kaya, E. (2020). Fifth-grade students’ understanding of social-institutional aspects of science. *International Journal of Science Education*, 42(11):1834–1861.
- Allchin, D. (2011). Evaluating knowledge of the nature of (whole) science. *Science Education*, 95(3):518–542.
- Andreoletti, M., Tinterri, A., et al. (2023). *Apprendere con i giochi. Esperienze di progettazione ludica*. Carocci Editore.
- Arcand, K. (2022). Chandra’s Accessible Universe: From Sight to Sound & Touch. In *Revista Mexicana de Astronomia y Astrofisica Conference Series*, volume 54 of *Revista Mexicana de Astronomia y Astrofisica Conference Series*, pages 53–56.
- Bailey, J. (2024). *Engagement by design: Engagement through design*. PhD thesis, Te Herenga Waka-Victoria University of Wellington.
- Bardelli, S., Ferretti, C., Ludovico, L. A., Presti, G., and Rinaldi, M. (2021). A sonification of the zCOSMOS galaxy dataset. In *International Conference on Human-Computer Interaction*, pages 171–188. Springer.
- Bartlett, S. (2024). Evaluation Framework & Toolkit: For OAE NAECs, Centers & Nodes. Version 4.0.
- Bartlett, S., Fitzgerald, M. T., McKinnon, D. H., Danaia, L., and Lazendic-Galloway, J. (2018). Astronomy and Science Student Attitudes (ASSA): A Short Review and Validation of a New Instrument. *Journal of Astronomy & Earth Sciences Education (JAESE)*, 5(1):1–22.
- Bayeck, R. Y. (2020). Examining board gameplay and learning: A multidisciplinary review of recent research. *Simulation & Gaming*, 51(4):411–431.

-
- Besson, U. (2015). *Didattica della fisica*. Carocci editore.
- Bieryla, A., Diaz-Merced, W., Davis, D., Hyman, S., Troncoso Iribarren, P., García, B., and Labbe, E. (2020). LightSound: The Sound of An Eclipse. *Communicating Astronomy with the Public Journal*, 28:38.
- Binkley, M., Erstad, O., Herman, J., Raizen, S., Ripley, M., Miller-Ricci, M., and Rumble, M. (2011). Defining Twenty-First Century Skills. In *Assessment And Teaching Of 21st Century Skills*, pages 17–66. Springer.
- Bonne, N. J., Gupta, J. A., Krawczyk, C. M., and Masters, K. L. (2018). Tactile Universe makes outreach feel good. *Astronomy and Geophysics*, 59(1):1.30–1.33.
- Borchelt, R. and Hudson, K. (2008). Engaging the scientific community with the public: Communication as a dialog, not a lecture. *Science Progress*. Accessed: 2025-07-07.
- Bowater, L. and Yeoman, K. (2012). *Science communication: A practical guide for scientists*. John Wiley & Sons.
- Braun, V. and Clarke, V. (2019). Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health*, 11(4):589–597.
- Britner, S. L. and Pajares, F. (2006). Sources of science self-efficacy beliefs of middle school students. *Journal of Research in Science Teaching: The Official Journal of the National Association for Research in Science Teaching*, 43(5):485–499.
- Bruner, J. S. (1966). *Toward a Theory of Instruction*. Harvard University Press, Cambridge, MA.
- Bruner, J. S. (1977). *The Process of Education*. Harvard University Press, Cambridge, MA. Original work published 1961.
- Bryant, C. (2003). Does Australia need a more effective policy of science communication? *International Journal for Parasitology*, 33(4):357–361.
- Buaraphan, K. (2012). Embedding nature of science in teaching about astronomy and space. *Journal of Science Education and Technology*, 21(3):353–369.
- Bucchi, M. and Trench, B. (2014). Science communication research: themes and challenges. In *Routledge handbook of public communication of science and technology*, pages 17–30. Routledge.
- Burns, T. W., O'Connor, D. J., and Stocklmayer, S. M. (2003). Science communication: a contemporary definition. *Public understanding of science*, 12(2):183–202.

-
- Caillois, R. (2001). *Man, play, and games*. University of Illinois press.
- Calleja, G. (2022). *Unboxed: Board Game Experience And Design*. MIT Press.
- Caramaschi, M., Cullinane, A., Levrini, O., and Erduran, S. (2022). Mapping the nature of science in the Italian physics curriculum: from missing links to opportunities for reform. *International Journal of Science Education*, 44(1):115–135.
- Cardinot, A. and Fairfield, J. A. (2022). Game-based learning to engage students with physics and astronomy using a board game. In *Research Anthology on Developments in Gamification and Game-Based Learning*, pages 785–801. IGI Global.
- Chen, J. (2007). Flow In Games (And Everything Else). *Communications Of The ACM*, 50(4):31–34.
- Cheung, K. K. C. and Erduran, S. (2023). A systematic review of research on family resemblance approach to nature of science in science education. *Science & Education*, 32(5):1637–1673.
- Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. routledge.
- Crookall, D. (2010). Serious games, debriefing, and simulation/gaming as a discipline. *Simulation & gaming*, 41(6):898–920.
- Csikszentmihalyi, M., Larson, R., et al. (2014). *Flow And The Foundations Of Positive Psychology*, volume 10. Springer.
- Cunningham, B. (1976). Action research: toward a procedural model. *Human relations*, 29(3):215–238.
- Dagher, Z. R. and Erduran, S. (2016). Reconceptualizing the nature of science for science education: Why does it matter? *Science & Education*, 25(1):147–164.
- Davies, S. R. (2021). An empirical and conceptual note on science communication’s role in society. *Science Communication*, 43(1):116–133.
- de Freitas, S. (2006). *Learning In Immersive Worlds: A Review Of Game-Based Learning*. Report, JISC, Bristol, UK. ISBN: 991005543282807891.
- Delaney, N. and Tornasi, Z., editors (2020). *Science with and for Society in Horizon 2020 – Achievements and Recommendations for Horizon Europe*. Publications Office of the European Union, Luxembourg. European Commission, Directorate-General for Research and Innovation; Dagher, R.; Monachello, R.; Warin, C.

-
- Deterding, S., Dixon, D., Khaled, R., and Nacke, L. (2011). From game design elements to gamefulness: defining "gamification". In *Proceedings of the 15th International Academic MindTrek Conference: Envisioning Future Media Environments*, pages 9–15.
- Dewey, J. (1916). *Democracy and Education: An Introduction to the Philosophy of Education*. The Macmillan Company, New York.
- Dewey, J. (1938). *Experience And Education*. Macmillan, New York.
- Driver, R., Leach, J., Millar, R., and Scott, P. (1996). *Young people's images of science*. McGraw-Hill Education (UK).
- Durant, J. R., Evans, G. A., and Thomas, G. P. (1989). The public understanding of science. *Nature*, 340(6228):11–14.
- EHT Collaboration (2019). Press Release (April 10, 2019): Astronomers Capture First Image Of A Black Hole. <https://eventhorizontelescope.org/press-release-april-10-2019-astronomers-capture-first-image-black-hole>. Accessed: 2025-09-09.
- Einsiedel, E. F. (2005). Understanding “publics” in the public understanding of science. In *Between understanding and trust*, pages 156–162. Routledge.
- Elmqvist, E., Ejdbo, M., Bock, A., and Rönnerberg, N. (2021). OpenSpace Sonification: Complementing Visualization of the Solar System with Sound. In *Proceedings of the 26th International Conference on Auditory Display (ICAD 2021)*, pages 135–142, Virtual / Silent Sky Studios, Trento, Italy. Georgia Institute of Technology.
- Erduran, S. and Dagher, Z. R. (2014). Reconceptualizing nature of science for science education. In *Reconceptualizing the nature of science for science education: Scientific knowledge, practices and other family categories*, pages 1–18. Springer.
- Erduran, S., Dagher, Z. R., and McDonald, C. V. (2019). Contributions of the family resemblance approach to nature of science in science education: A review of emergent research and development. *Science & Education*, 28(3):311–328.
- Foran, G., Cooke, J., and Hannam, J. (2022). The Power of Listening to your Data: Opening Doors and Enhancing Discovery Using Sonification. In *Revista Mexicana de Astronomia y Astrofisica Conference Series*, volume 54 of *Revista Mexicana de Astronomia y Astrofisica Conference Series*, pages 1–8.
- Fraenkel, J. R. and Wallen, N. E. (1990). *How to design and evaluate research in education*. ERIC.

-
- Galano, S., Palazzo, L., and Testa, I. (2024). A latent profile analysis of students' attitudes towards astronomy across grades 9–13. *International Journal of Science Education*, 46(14):1480–1516.
- García-Benito, R. and Pérez-Montero, E. (2022). Painting graphs with sounds: CosMonic sonification project. In *Revista Mexicana de Astronomía y Astrofísica Conference Series*, volume 54 of *Revista Mexicana de Astronomía y Astrofísica Conference Series*, pages 28–33.
- Gheno, V. (2024). Gender Inclusiveness In A Binary Language: The Rise Of The Schwa In Italian And The Discussion Surrounding It. In *Inclusiveness Beyond The (Non) Binary In Romance Languages*, pages 50–65. Routledge.
- Gross, A. G. (1994). The roles of rhetoric in the public understanding of science. *Public understanding of science*, 3(1):3.
- Guenther, L. and Joubert, M. (2017). Science communication as a field of research: identifying trends, challenges and gaps by analysing research papers. *Journal of Science Communication*, 16(2):A02.
- Guiotto Nai Fovino, L., Zanella, A., Di Mascolo, L., Ginolfi, M., Carpita, N., Trovato Manuncola, F., and Grassi, M. (2024). Evaluating the effectiveness of sonification in science education using Edukoi. *Personal and Ubiquitous Computing*.
- Harrison, C., Trayford, J., Harrison, L., and Bonne, N. (2022). Audio universe: tour of the solar system. *Astronomy and Geophysics*.
- Harrison, C., Zanella, A., and Girdhar, A. (2024). Evaluation and insights from a sonification-based planetarium show intended for improving inclusivity. *arXiv preprint arXiv:2401.04188*.
- Harrison, C. M., Zanella, A., and Girdhar, A. (2023). Evaluation and insights from a sonification-based planetarium show intended for improving inclusivity. *Communicating Astronomy with the Public Journal*, 33:6.
- Helyer, R. (2015). Learning through reflection: the critical role of reflection in work-based learning (WBL). *Journal of Work-Applied Management*, 7(1):15–27.
- Her Majesty's Stationery Office (1993). *Realising Our Potential: A Strategy for Science, Engineering and Technology*.
- Hilgartner, S. (1990). The dominant view of popularization: Conceptual problems, political uses. *Social studies of science*, 20(3):519–539.

-
- Holden, C. (2002). From PUS to PEST. *Science*, 298(5591):49.
- House of Lords Select Committee on Science and Technology (2000). Science and Technology Third Report. Also known as the "Science and Society" report.
- Huizinga, J. (1949). *Homo Ludens: A Study of the Play-Element in Culture*. Routledge & Kegan Paul, London. Original work published 1938.
- Illingworth, S. et al. (2019). Developing Science Tabletop Games: Catan And Global Warming. *JCOM: Journal Of Science Communication*, 18(4).
- Inchingolo, G., Toniolo, R., Varano, S., Ligabue, A., Ricciardi, S., et al. (2023). PIXEL: Challenges of designing a professional board game for astronomy education. In *European Conference on Games Based Learning*, volume 17, pages 278–288. Academic Conferences International Limited.
- Inchingolo, G., Toniolo, R., Varano, S., Zanazzi, A., Ligabue, A., and Ricciardi, S. (2024). A board game ludography analysis for Game-Based Learning in Astronomy, Astrophysics and Space Science. *Astronomy Education Journal*, 4(1).
- Irzik, G. and Nola, R. (2011). A family resemblance approach to the nature of science for science education. *Science & education*, 20(7):591–607.
- Irzik, G. and Nola, R. (2014). New directions for nature of science research. In *International handbook of research in history, philosophy and science teaching*, pages 999–1021. Springer.
- Irzik, G. and Nola, R. (2023). Revisiting the foundations of the family resemblance approach to nature of science: Some new ideas. *Science & Education*, 32(5):1227–1245.
- Jones, R. A. L. (2011). Public engagement in an evolving science policy landscape. In Bennett, D. J. and Jennings, R. C., editors, *Successful Science Communication: Telling It Like It Is*, pages 1–14. Cambridge University Press, Cambridge.
- Kaya, E. and Erduran, S. (2016). From FRA to RFN, or how the family resemblance approach can be transformed for science curriculum analysis on nature of science. *Science & Education*, 25(9):1115–1133.
- Kaya, E., Erduran, S., Aksoz, B., and Akgun, S. (2019). Reconceptualised family resemblance approach to nature of science in pre-service science teacher education. *International Journal of Science Education*, 41(1):21–47.
- Koballa, T. R. and Glynn, S. M. (2013). Attitudinal and motivational constructs in science learning. In *Handbook of research on science education*, pages 75–102. Routledge.

-
- Lederman, N. G. (1992). Students' and teachers' conceptions of the nature of science: A review of the research. *Journal of research in science teaching*, 29(4):331–359.
- Lederman, N. G. (2004). Syntax of nature of science within inquiry and science instruction. *Scientific inquiry and nature of science*, pages 301–317.
- Lederman, N. G. and Zeidler, D. L. (1987). Science teachers' conceptions of the nature of science: Do they really influence teaching behavior? *Science Education*, 71(5):721–734.
- Lenzi, S. and Ciuccarelli, P. (2020). Intentionality and design in the data sonification of social issues. *Big Data & Society*, 7(2):205395172094460.
- Lewin, K. (1948). *Resolving Social Conflicts: Selected Papers On Group Dynamics*. Harper, New York.
- Ligabue, A. (2020). *Didattica ludica: Competenze in gioco*. Edizioni Centro Studi Erickson.
- Mäyrä, F. (2008). *An introduction to game studies*. Sage.
- Miller, S. (2001). Public understanding of science at the crossroads. *Public understanding of science*, 10(1):115–120.
- Ministero dell'Istruzione e del Merito (2024). Focus: La Dispersione scolastica negli anni scolastici 2019/2020, 2020/2021 e 2021/2022. https://www.foe.it/files/2024/01/Focus-La-Dispersione-scolastica-aa.ss_.1920_2021-2021_2122.pdf. Accessed: 2025-10-13.
- Murray, C., Dunstan, M., Heron, C., Holland, L., Palmer, S., Price, D., and Basham, M. (2022). Diamond: The Game – a board game for secondary school students promoting scientific careers and experiences. *Research for All*, 6(1):1–18.
- NASA, ESA, CSA, and STScI (2022). Cosmic Cliffs In The Carina Nebula (NGC 3324). <https://webbtelescope.org/contents/media/images/2022/031/01G77PKB8NKR7S8Z6HBXMYATGJ>. Accessed: 2025-09-09.
- NCCPE (2023). The Beacons for Public Engagement. Technical report, NCCPE. Accessed on July 14, 2025.
- Nesti, R. (2017). Game-based learning. *Gioco e progettazione ludica in educazione*. Pisa: Edizioni ETS.
- Nisbet, M. C. and Markowitz, E. (2015). Public engagement research and major approaches. *AAAS Center for Public Engagement with Science and Technology Washington: Washington, DC, USA*.

-
- Noel-Storr, J. and Willebrands, M. (2022). Accessibility in astronomy for the visually impaired. *Nature Astronomy*, 6:1216–1218.
- Osborne, J. and Collins, S. (2001). Pupils' views of the role and value of the science curriculum: a focus-group study. *International journal of science education*, 23(5):441–467.
- Paredes-Sabando, P. and Fuentes-Muñoz, C. (2021). Dedoscopio Project: Making Astronomy Accessible to Blind and Visually Impaired (BVI) Communities Across Chile. *Communicating Astronomy with the Public Journal*, 29:27.
- Pekrun, R., Goetz, T., Frenzel, A. C., Barchfeld, P., and Perry, R. P. (2011). Measuring emotions in students' learning and performance: The Achievement Emotions Questionnaire (AEQ). *Contemporary educational psychology*, 36(1):36–48.
- Pérez-Montero, E. (2019). Towards a more inclusive outreach. *Nature Astronomy*, 3:114–115.
- Pérez-Montero, E., Lanzara, M., Ortiz-Gil, A., Villaverde, M., García-Benito, R., Gallego-Calvente, T., and García Gómez-Caro, E. (2023). A tactile model of the night summer northern sky for the teaching of astronomy to the BVI. In Manteiga, M., Bellot, L., Benavidez, P., de Lorenzo-Cáceres, A., Fuente, M. A., Martínez, M. J., Vázquez Acosta, M., and Dafonte, C., editors, *Highlights on Spanish Astrophysics XI*, page 468.
- Piaget, J. (1962). *Play, Dreams And Imitation In Childhood*. Norton, New York. English translation of *La Formation Du Symbole Chez L'Enfant* (1945).
- Pink, D. H. (2009). *Drive: The Surprising Truth About What Motivates Us*. Riverhead Books, Penguin Group (USA), New York.
- Pitrelli, N. (2003). The crisis of the "Public Understanding of Science" in Great Britain. *JCOM - Journal of Science Communication*, 2(01).
- Pivec, M., Dziabenko, O., and Schinnerl, I. (2003). Aspects of game-based learning. In *3rd International Conference on Knowledge Management, Graz, Austria*, volume 304, pages 14–16.
- Qian, M. and Clark, K. R. (2016). Game-based Learning and 21st Century Skills: A review of recent research. *Computers in Human Behavior*, 63:50–58.
- Quin, M. (1991). The Interactive Science and Technology Project: the Nuffield Foundation's Launchpad for a European Collaborative. *International journal of science education*, 13(5):569–573.

-
- Quinton, M., McGregor, I., and Benyon, D. (2021). Sonification of Planetary Orbits in Asteroid Belts. In *Proceedings of the 16th International Audio Mostly Conference*, AM '21, page 72–80, New York, NY, USA. Association for Computing Machinery.
- Ricciardi, S., Varano, S., Zanazzi, A., et al. (2019). Talkin' About The Resolution. In *Conference Proceedings. New Perspectives In Science Education 2019*.
- Rogers-Hayden, T. and Pidgeon, N. (2007). Moving engagement “upstream”? Nanotechnologies and the Royal Society and Royal Academy of Engineering’s inquiry. *Public Understanding of Science*, 16(3):345–364.
- Ryan, R. M. and Deci, E. L. (2000). Self-Determination Theory And The Facilitation Of Intrinsic Motivation, Social Development, And Well-Being. *American Psychologist*, 55(1):68.
- Ryan, R. M., Rigby, C. S., and Przybylski, A. (2006). The Motivational Pull Of Video Games: A Self-Determination Theory Approach. *Motivation And Emotion*, 30(4):344–360.
- Salimpour, S., Bartlett, S., Fitzgerald, M. T., McKinnon, D. H., Cutts, K. R., James, C. R., Miller, S., Danaia, L., Hollow, R. P., Cabezon, S., et al. (2021). The gateway science: A review of astronomy in the OECD school curricula, including China and South Africa. *Research in Science Education*, 51(4):975–996.
- Salimpour, S. and Fitzgerald, M. T. (2024). Astronomy and Culture: A Social Semiotic Perspective on the Role of Culture in Astronomy Education. *Science & Education*, 33(2):405–426.
- Sandrelli, S. (2023). Comunicare la scienza: il public understanding of science. <https://edu.inaf.it/rubriche/oltre-orizzonte/pus>.
- Savia, G. (2016). *Universal Design for Learning: La Progettazione Universale per l'Apprendimento per una didattica inclusiva*. Edizioni Centro Studi Erickson.
- Schelly, C. L., Davies, P. L., and Spooner, C. L. (2011). Student Perceptions Of Faculty Implementation Of Universal Design For Learning. *Journal Of Postsecondary Education And Disability*, 24(1):17–30.
- Schreiner, C. and Sjøberg, S. (2004). Sowing the seeds of ROSE. Background, rationale, questionnaire development and data collection for ROSE (the Relevance of Science Education) – a comparative study of students’ views of science and science education. Technical Report 4, Acta Didactica, Oslo, Norway.

-
- Setiawan, H. and Phillipson, S. (2019). The effectiveness of Game-Based Science Learning (GBSL) to improve students' academic achievement: A meta-analysis of current research from 2010 to 2017. *REiD (Research and Evaluation in Education)*, 5(2):7.
- Short, D. B. (2013). The public understanding of science: 30 years of the Bodmer report. *School Science Review*, 95(350):39–44.
- Squire, K. (2003). Video Games In Education. *Int. J. Intell. Games & Simulation*, 2(1):49–62.
- Squire, K. (2005). Changing The Game: What Happens When Video Games Enter The Classroom? *Innovate: Journal Of Online Education*, 1(6).
- Stilgoe, J. and Sykes, K. (2009). *The Road Ahead: Public Dialogue on Science and Technology*. Sciencewise, Didcot. Accessed: 2025-07-02.
- Sturloni, G. (2003). La paura vien mangiando - Comunicare il rischio alimentare. *JCOM*, 2:1.
- Suits, B. (1978). *The Grasshopper: Games, Life and Utopia*. University of Toronto Press, Toronto.
- Swertz, C. (2019). Game media literacy. *The International Encyclopedia of Media Literacy*, pages 1–8.
- Tang, C. (2002). Reflective diaries as a means of facilitating and assessing reflection. In *Quality conversations: Proceedings of the 29th HERDSA Annual Conference Perth*, pages 7–10.
- The Royal Society (1985). The Public Understanding of Science (Report of a Royal Society ad hoc Group). *Royal Society*.
- Tomlinson, B., Winters, R., Latina, C., Bhat, S., Rane, M., and Walker, B. (2017). Solar System Sonification: Exploring Earth and Its Neighbors Through Sound. In *The 23rd International Conference on Auditory Display (ICAD 2017)*, pages 128–134.
- Toniolo, R. (2020). Didattica della Astrofisica con un approccio innovativo: un caso studio di game-based learning. Tesi di laurea magistrale, Università degli Studi di Padova. Anno accademico 2020–2021 (17 settembre 2020), 95 p.
- Toniolo, R., Inchingolo, G., Casu, S., Contino, G., La Parola, V., Leoni, R., Varano, S., Zanazzi, A., Ligabue, A., and Ricciardi, S. (2023). Pixel-Picture (Of) The Universe: an educational board game about astrophysics research world. In *EDULEARN23 Proceedings*, pages 4202–4209. IATED.

-
- Toniolo, R., Inchingolo, G., Zanazzi, A., and Casu, S. (2025). Cosmo Hunters-a card game for middle-school (K12) Astronomy Education. *Italian Journal of Educational Technology*, 33(1):13–30.
- Toniolo, R., Zanella, A., Cottinelli, A., Liuzzi, G., Ricciardi, S., Grassi, M., and Delle Monache, S. (2026). Amaca: astronomy education with a multi-sensory, accessible, and circular approach. *International Journal of Science Education, Part B*, pages 1–25.
- Trench, B. and Bucci, M. (2010). Science communication, an emerging discipline. *Journal of science communication*, 9(3):C03.
- Usuda-Sato, K., Nakayama, H., Fujiwara, H., and Usuda, T. (2019). Touch the Universe: Developing and Disseminating Tactile Telescope Models Created with a 3D Printer. *Communicating Astronomy with the Public Journal*, 26:24.
- Varano, S. and Di Giacomo, F. (2024). Design of a multisensory planetarium. *arXiv e-prints*, page arXiv:2407.05801.
- Varano, S. and Ricciardi, S. (2018). Representing The Universe: A Hands-On Challenge. In *Communicating Astronomy With The Public Conference*.
- Varano, S. and Zanella, A. (2023). Design and evaluation of a multi-sensory representation of scientific data. In *Frontiers in Education*, volume 8, page 1082249. Frontiers Media SA.
- Vygotsky, L. S. (1978). *Mind in Society: The Development of Higher Psychological Processes*. Harvard University Press, Cambridge, MA. Edited by Michael Cole, Vera John-Steiner, Sylvia Scribner, and Ellen Souberman.
- Wallin, P. and Adawi, T. (2018). The reflective diary as a method for the formative assessment of self-regulated learning. *European Journal of Engineering Education*, 43(4):507–521.
- Weingart, P., Joubert, M., and Connaway, K. (2021). Public engagement with science - Origins, motives and impact in academic literature and science policy. *PloS one*, 16(7):e0254201.
- Whitton, N. J. (2018). Playful Learning: Tools, Techniques, And Tactics. *Research In Learning Technology*, 26:1–12.
- Wu, J.-Y. and Erduran, S. (2024). Investigating scientists’ views of the family resemblance approach to nature of science in science education. *Science & Education*, 33(1):73–102.

-
- Wynne, B. (1991). Knowledges in context. *Science, Technology, & Human Values*, 16(1):111–121.
- Wynne, B. (1992). Misunderstood misunderstanding: social identities and public uptake of science. *Public understanding of science*.
- Wynne, B. (1995). Public understanding of science. In *Handbook of Science and Technology Studies, Revised Edition*, pages 361–388. SAGE Publications, Inc.
- Yoon, C. S. and Khambari, M. N. M. (2022). Design, Development, and Evaluation of the Robobug Board Game: An Unplugged Approach to Computational Thinking. *International Journal of Interactive Mobile Technologies*, 16(6).
- Zanella, A., Harrison, C., Lenzi, S., Cooke, J., Damsma, P., and Fleming, S. (2022). Sonification and sound design for astronomy research, education and public engagement. *Nature Astronomy*, 6(11):1241–1248.
- Zarzycki, A. (2016). Epic Video Games: Narrative Spaces And Engaged Lives. *International Journal Of Architectural Computing*, 14(3):201–211.
- Ziman, J. (1992). Not knowing, needing to know, and wanting to know. In Lewenstein, B. V., editor, *When Science Meets the Public*, pages 13–20. American Association for the Advancement of Science, Washington, DC.

Appendix A

PIXEL supporting materials

A.1 Tables

Table A.1 – Correspondence between PIXEL game mechanics and counterparts in astrophysics research

Game mechanics and dynamics	Counterparts in astrophysics research
One can win the game using different strategies, for example focusing on publications, observations, or completing projects and objectives	Scientific research comprehends a multitude of approaches, methodologies, and strategies
The number and variety of Publication cards give Scientific Prestige (SP) points at the end of the game	Astrophysicists communicate with their peers by writing papers for peer-reviewed journals. This process may help a researcher or a research group build a reputation
At the end of the game, the player with the higher SP points is the winner	Different kinds of rewards drive astrophysics researchers. One of the key ones is the recognition and prestige in the scientific community and society
Most of the actions needed to proceed in the game require the investment of resources, but they are not a direct way to collect SP points	Financial resources are necessary to conduct research but, by themselves, cannot guarantee a good scientific outcome

Continued on next page

Game mechanics and dynamics	Counterparts in astrophysics research
At the beginning of each turn, players receive resources based on the growth of their research center, calculated by adding the size of the research center, the capabilities of the telescope, and the number of researchers	Public funds are distributed based on the size of the research center and its productivity
Players have to manage a research team that can carry out different actions	Scientific investigation is a collaborative and team effort of researchers, not a stand-alone activity
Resources can be invested to hire new researchers from the University slot	Research centers can be improved and benefit through the acquisition of new personnel
Researchers can be specialized through the acquisition of Specialist Cards that benefit the player	Managing a research center requires a multidisciplinary team of researchers and professionals who help with fundraising, public communication, and technology implementation
Research centers can be improved through the acquisition of Improvement Cards that represent infrastructures or processes such as supercomputers, workshops, scientific collaborations, and fundamental research	Research involves a plurality of facilities (e.g., supercomputers, workshops) and processes (e.g., scientific collaboration, fundamental research) that positively impact research effectiveness
Observations can be obtained with a ground telescope (private) and a space one (shared)	Astrophysics research can be both competitive (e.g., racing to observe a phenomenon first) and collaborative
Doing an observation benefits the player who made it, but the tile revealed on the board is available to all players and counts for everyone's projects and objectives	Being the first to study a specific object gives an advantage to the research team; however, data and results are shared with the scientific community
Leading the observation of a specific object makes players gain SP points	Focusing research on a specific topic builds expertise in that area
To place a higher-resolution tile, the corresponding lower-resolution one needs to be already placed on the board	Astrophysics research often requires low-resolution surveys to identify regions of interest for subsequent high-resolution studies

Continued on next page

Game mechanics and dynamics	Counterparts in astrophysics research
The highest-resolution tiles are smaller than the lowest-resolution ones	Obtaining a high-resolution image requires more observation time (modeled as placing a tile on the board)
Investing in the technological development of telescopes and research centers allows for placing higher-resolution tiles	Continual investment in technology development is required to enable more detailed observations of the Universe
Improving the space telescope requires fewer resources than improving the ground ones	Space telescopes are typically the result of international collaborations
Ground observation can be affected randomly during the game by observational conditions, while the space telescope is immune to this effect	Atmospheric conditions affect ground observations, whereas space telescopes, orbiting above Earth's atmosphere, are immune

Table A.2 – Overview of activities and session types in the GBL program.

Session description	Type
1. Compilation of reflective diary #1, aimed at eliciting students' initial conceptions of science, research, and researchers.	Diary (1h)
2. Introduction to the group and to the overall project. First gameplay session facilitated and explained by researchers, followed by a debriefing.	Training (3h)
3. Compilation of reflective diary #2, focused on first impressions of the game.	Diary (1h)
4. Second gameplay session facilitated by researchers, with debriefing on the role of the facilitator. Distribution of the facilitator's guide.	Training (3h)
5. Compilation of reflective diary #3, reflecting on interaction and analogies between game dynamics and research processes.	Diary (1h)
6. Introductory training on resolution in astrophysics and on game-based learning, followed by gameplay and debriefing.	Training (4h)
7. Compilation of reflective diary #4, focused on strategies, representation of resolution, and the publication system.	Diary (1h)
8. Gameplay session facilitated by students (first attempt).	Training (4h)
9. Gameplay session facilitated by students (second attempt).	Training (4h)
10. Compilation of reflective diary #5, reflecting on facilitation experiences (Sessions 8 and 9).	Diary (1h)

Continued on next page

Session description	Type
11. Facilitation of gameplay with classmates by students, followed by debriefing supported by INAF researchers.	Experience (2h)
12. Compilation of reflective diary #6, reflecting on the first facilitation experience.	Diary (1h)
13. Facilitation of the second gameplay session by students, followed by debriefing with classmates.	Experience (2h)
14. Compilation of reflective diary #7, reflecting on the meaning of Specialist Cards.	Diary (1h)
15. Support and feedback meeting on students' facilitation experiences.	Feedback (2h)
16. Facilitation of the third gameplay session by students, followed by debriefing with classmates.	Experience (2h)
17. Compilation of reflective diary #8, reflecting on peers' engagement and perception of science.	Diary (1h)
18. Facilitation of the fourth gameplay session by students, followed by debriefing with classmates.	Experience (2h)
19. Compilation of reflective diary #9, final reflections on science, research, and researchers.	Diary (1h)
20. Final focus group on experiences gathered throughout the facilitation activities.	Feedback (3h)

Table A.3 – Questions of the reflective diaries.

Diary #1
1. What is science to you?
2. How does scientific research work?
3. How do researchers work?
4. What are your favorite subjects?
5. What are your personal interests?
6. How often do you play board games?
7. What is your favorite game?
8. Do you play other types of games?

Continued on next page

Questions of the reflective diaries

Diary #2

1. How did you feel at the beginning, during, and at the end of the game?
 2. What did you understand about scientific research through the game?
 3. How was scientific research represented in the game?
-

Diary #3

1. What kind of interaction occurred between you and the rest of the group?
 2. Did the actions of others help you in any way, and vice versa? When?
 3. What difficulties did you experience?
 4. In your opinion, is there a counterpart in the real research world?
-

Diary #4

1. What does the game reward the most?
 2. Which strategies are the most effective?
 3. How is resolution represented in the game?
 4. Why does only the first player reaching the publication marker obtain the Publication card?
 5. In the game, publications can be obtained in three ways: by spending one resource, by being the first to reach the publication marker of an object, or by leading on that track. Why do you think this is so?
-

Diary #5

1. Did being in the role of facilitator make you notice anything new about the game?
 2. Did you develop new reflections on science and scientific research?
-

Diary #6

1. Which Improvement card is the most useful and why?
 2. In the game, how are ground-based telescopes different from the space telescope?
 3. How did you feel facilitating in front of your classmates?
 4. Did you notice differences between playing as a participant and facilitating?
-

Diary #7

1. Which Specialist card is the most useful and why?
 2. What do Specialist cards represent in the research world?
-

Continued on next page

Questions of the reflective diaries

3. Were there things you had not noticed or understood about the game that became clearer during facilitation?

Diary #8

1. Based on your facilitation experience, what effects did you observe on your classmates?
 2. Did the game stimulate interest in some of them?
 3. Do you think it may have changed their idea of science?
-

Diary #9

1. After this experience with PIXEL, how would you answer the question: “What is science to you?”
 2. After this experience with PIXEL, how would you answer the question: “How does scientific research work?”
 3. After this experience with PIXEL, how would you answer the question: “How do researchers work?”
-

Table A.4 – Frequencies of the themes emerged in the responses to the questions (pre/post):
 (1) What is science for you? (2) How does scientific research work? (3) How do researchers work?

Themes	(1)		(2)		(3)	
	PRE	POST	PRE	POST	PRE	POST
Accuracy	0	0	0	0	8	7
Collaboration	0	3	1	7	5	12
Competition	0	1	0	0	0	0
Scientific community	0	0	0	0	3	4
Knowledge	7	7	2	3	0	0
Creativity	1	0	0	0	1	0
Curiosity	3	2	0	0	0	0
Outreach	0	0	0	0	0	1
Future	2	2	0	0	0	0
Commitment	0	0	0	7	0	0
Contingencies	0	0	0	1	0	0

Continued on next page

Themes	(1)		(2)		(3)	
	PRE	POST	PRE	POST	PRE	POST
Interpretation	0	0	0	1	0	0
Work	0	2	0	0	0	0
Scientific method	6	1	25	8	12	5
Improvements	0	0	0	4	0	0
Multiple approaches	0	0	0	0	2	1
Multiplicity of disciplines	3	2	0	0	0	4
Objectives	0	0	0	8	0	10
Objectivity	0	0	0	0	8	1
Astronomical observation	0	0	0	6	0	0
Passion	2	2	0	0	2	0
Patience	0	1	0	0	0	0
Planning	0	0	0	0	0	1
Progress	4	7	1	3	0	0
Publication	0	0	1	6	1	2
Replicability	0	0	3	0	1	0
Research	5	9	0	0	0	0
Researchers	0	0	2	9	0	0
Resources	0	0	0	4	0	4
Discovery	0	0	0	1	0	0
Positive feelings	0	4	0	0	0	0
Experimentation	0	4	0	0	0	0
Strategy	0	0	0	4	0	0
Instrumentation	0	1	0	5	0	2
Study of nature	12	8	0	1	2	0
Time	0	0	0	3	0	0
Personal elaboration	0	12	8	18	11	11
Incorrect answer	0	0	1	0	1	0
Formal description	0	2	14	4	1	0
Not knowing	0	0	0	0	2	0
Off-topic	0	0	1	0	1	0

Continued on next page

Themes	(1)		(2)		(3)	
	PRE	POST	PRE	POST	PRE	POST
Superficial	8	8	3	6	10	12
Imprecise terminology	0	0	8	0	3	3

Table A.5 – List of themes emerged from students’ responses addressing the RFN cognitive-epistemic system (left) and the social-institutional system (right).

Cognitive-epistemic system	Social-institutional system
Accuracy	Collaboration
Astronomical observation	Commitment
Discovery	Competition
Experimentation	Contingencies
Interpretation	Creativity
Knowledge	Curiosity
Multiplicity of discipline	Future
Multiple approaches	Improvements
Objectives	Instrumentation
Objectivity	Outreach
Replicability	Passion
Research	Patience
Scientific method	Planning
Study of nature	Positive feelings
	Progress
	Publication
	Researchers
	Resources
	Scientific community
	Strategy
	Time
	Work

Table A.6 – Complete list of questionnaire items for the evaluation of the single play session.

Section 0 - Introduction

1. Write your class and student ID number.
 2. (*Open-ended*) Indicate the gender that best represents you.
-

Section A – Understanding of social-institutional aspects of NOS

On a scale from 1 (= not at all) to 5 (= completely), indicate how much you agree with the following statements.

1. There is a relationship between science and money
2. Scientists examine and evaluate the work of others
3. Science takes place in institutions such as universities and research centers
4. Science originates within society
5. Politics influences science
6. Scientists should respect the environment
7. Scientists share their research with the community
8. Scientists need money to carry out research
9. States influence the development of scientific knowledge
10. Scientists earn money through their work
11. Ethnicity of scientists influences their work
12. Scientists write articles for academic journals
13. Scientists attend conferences to share their research
14. Scientists must be honest
15. Scientists interact with other scientists during their research

(Open-ended) In your opinion, what kind of activities do astronomers carry on during their work?

Section B – Attitudes towards astronomy

On a scale from 1 (= not at all) to 5 (= completely), indicate how much you agree with the following statements.

1. I am interested in the results of research in astronomy
 2. I find astronomy interesting
 3. I find the study of astronomy fulfilling
 4. I would like to use a telescope
-

Continued on next page

(continued)

5. I would like to become an astronomer
 6. I would like to get a job involving astronomy
 7. I would invest my future in the profession of astronomer
 8. Knowing astronomy will be useful in my profession
 9. I believe that the astronomer carry an important work
 10. I think that in today's society, astronomers are highly esteemed
 11. I think that there are many jobs that astronomy or astrophysics graduates can do
 12. I think that a career as a researcher in astronomy is rewarding and of high status
 13. I think that astronomy is important for technological progress
-

Section C – Self-efficacy in astronomy

On a scale from 1 (= not at all) to 5 (= completely), indicate how much you agree with the following statements.

1. Even if the work in astronomy is hard, I can learn it.
 2. I can do astronomy
 3. I can do even the hardest work in astronomy if I try
-

Table A.7 – Teacher interview questions about the activity.

Interview questions for teachers.

1. What did you notice while observing students during the activity?
 2. Did you observe anything particular in the interactions among students?
 3. Did you notice changes in roles or modes of participation compared to other school activities?
 4. How would you describe the level of student engagement?
 5. In your opinion, what types of skills or competencies were activated?
 6. Did you have the impression that students made connections with what they normally study?
 7. Was the activity discussed again in class and/or did students talk about it among themselves afterwards?
 8. What is your overall impression of the activity?
 9. Do you think you will propose the activity again, possibly with other classes?
-

Table A.8 – Themes identified in students’ open-ended responses categorized according to the RFN systems, with frequencies for each group and phase.

Cognitive-epistemic system				
Theme	Experimental		Control	
	PRE	POST	PRE	POST
Understanding	2	0	0	1
Searching	3	3	0	0
Innovation	1	0	1	0
Predictions	2	0	0	0
Discovery	7	6	1	3
Knowledge	1	2	0	0
Laws	2	1	0	0
Analysis	5	2	1	0
Data analysis	9	3	0	2
Image analysis	0	3	0	0
Calculations	4	3	0	0
Experiments	1	1	0	2
Space exploration	2	1	0	0
Observation	26	33	2	7
Data collection	2	2	0	2
Research	10	15	2	3
Simulations	0	0	0	1
Instruments	1	0	2	5
Study	19	14	4	6
Telescope	0	4	1	1
Demonstration	2	0	0	0
Hypothesis	5	6	1	3
Models	1	0	0	0
Theories	4	2	0	0
Thesis	0	1	0	0
Superficial	19	16	20	11

Continued on next page

Theme	Experimental		Control	
	PRE	POST	PRE	POST
Social-institutional system				
Funding	0	3	0	0
Investments	1	0	0	0
Travelling	0	1	0	0
Collaboration	1	3	2	2
Competition	0	1	0	0
Conferences	0	1	0	0
Publications	2	4	0	3
Research institutions	0	1	0	1
Projects	0	3	0	0
Outreach	0	1	1	0
Progress	1	4	1	0

A.2 Quantitative data analysis

Before proceeding with the quantitative analysis of the three dimensions, it was verified that no items required reverse scoring, as all were formulated in a consistent direction – higher scores corresponded to higher levels of the construct being measured. Subsequently, for each dimension or subdimension, the response values were summed for each student.

Subsequently, three different types of statistical tests were performed on the dimensions (and on the subdimensions in the case of attitudes) using the SPSS software, widely used for statistical analysis in the social sciences. In addition to the division by group, the sample was further subdivided by gender. However, since the control group consisted of approximately 90% male students – a common situation in Italy, where vocational-technological schools are predominantly attended by males – it was decided not to further subdivide this group. Regardless of the subdivision into subsamples, for each dimension only the questionnaires containing responses to all the items within that specific dimension were included in the analysis. As a result, the sample size slightly differed across dimensions.

The first test conducted consisted of descriptive statistics of the scores, aimed at providing an overview of the distribution of results. The outputs of this analysis, presented in the tables in the following section, include the number of entries (N), the mean (M), the standard error of the mean (SE), the standard deviation (SD), and the minimum (Min) and maximum (Max) values of the sample.

Then, a paired t -test was used to assess whether the difference between the pre and post-test scores within each sample was statistically significant. The t -test is a statistical procedure used to compare the means of two related datasets – in this case, the same participants' scores before and after the intervention – to determine whether the observed differences are likely to have occurred by chance. The test produces a t -value, reported as $t(df)$, where $df = N - 1$ indicates the degrees of freedom of the sample. The t -value is calculated using the following formula:

$$t = \frac{M_{\text{post}} - M_{\text{pre}}}{SE_{\text{diff}}}$$

where M_{post} and M_{pre} are the means of post and pre-test scores, and SE_{diff} is the standard error of the difference between means. In other words, the t -value expresses the size of the observed difference relative to the variability of the data. Higher absolute values of t indicate larger differences between pre and post-test means compared to within-sample variability. The associated p -value expresses the probability that the observed difference occurred by chance: values lower than 0.05 (5% probability) are typically considered statistically significant (Fraenkel and Wallen, 1990). In addition, the effect size, expressed as Cohen's d , quantifies the magnitude of the difference, providing an estimate of how meaningful the observed change is independently of the sample size. It is calculated as:

$$d = \frac{M_{\text{post}} - M_{\text{pre}}}{SD_{\text{diff}}} = \frac{t}{\sqrt{N}}$$

where SD_{diff} is the standard deviation of the difference scores. According to Cohen (2013), values of d around 0.2 indicate a small effect, around 0.5 a medium effect, and around 0.8 or higher a large effect.

Finally, an Analysis of Covariance (ANCOVA) was performed to determine whether differences in post-test scores between groups or genders could be attributed to the intervention, while controlling for potential pre-test differences. ANCOVA combines features of Analysis of Variance (ANOVA) and linear regression, allowing the comparison of group means while statistically controlling for the influence of one or more continuous variables, called covariates (in this case, the pre-test scores). This procedure helps to remove the portion of variance in the dependent variable that is associated with the covariate, thereby providing a more accurate estimate of the effect of the independent variable (the intervention). The analysis produces an F -value, reported as $F(df_{\text{between}}, df_{\text{within}})$, where df_{between} represents the degrees of freedom associated with the independent variable (in this case, the group), and df_{within} refers to the degrees of freedom of the residual or error term. The F -value represents the ratio between the variance explained by the model and the unexplained variance, or error term. The error term corresponds to the portion of variability in the dependent variable that is not accounted for by the independent variable(s) or covariates – essentially, the residual variance within each group. A higher F -value therefore indicates that the variability explained by the model is substantially greater than the random variability within groups, suggesting that the observed differences are unlikely to be due to chance. The associated p -value is the same introduced before and indicates the probability that the observed differences occurred by chance, with values below 0.05 typically considered statistically significant. In addition, the coefficient of deter-

mination (R^2) indicates the proportion of variance in the dependent variable that is explained by the independent variable(s) included in the model. In other words, it reflects how well the model fits the data, with higher values representing a greater proportion of explained variance. However, since R^2 tends to increase with the number of predictors, the *adjusted R^2* provides a more accurate estimate by penalizing the inclusion of unnecessary variables. It thus offers a more reliable measure of model fit, particularly when comparing models with different numbers of predictors. In the context of ANCOVA, these values quantify the extent to which differences in post-test scores are explained by the model after accounting for the effect of the covariate (the pre-test scores). In conclusion, R^2 or *adjusted R^2* value represent the percentage of how well the model accounts for the variability observed in the data. Therefore, higher values indicate that the model explains a larger portion of the variability in post-test scores, whereas lower values suggest that most of the variance remains unexplained or is due to random factors.

The following tables present the results of the descriptive statistics for the three construct: SIQ, attitude (divided in intrinsic, extrinsic, utilitarian) and self-efficacy.

Group	Gender	N	Pre-test				Post-test			
			M	SE	SD	Min-Max	M	SE	SD	Min-Max
Exp	Male	35	58.40	0.81	4.79	46-68	59.69	0.68	4.03	50-68
	Female	33	58.33	0.83	4.76	47-68	58.00	0.80	4.62	49-68
	All	68	58.37	0.57	4.74	46-68	58.87	0.53	4.38	49-68
Con	All	30	56.47	1.10	6.03	44-69	58.67	0.89	4.87	48-72

Table A.9 – Descriptive statistics for pre-test and post-test scores on the SIQ, divided by group (Exp = experimental, Con = control) and gender.

Group	Gender	N	Pre-test				Post-test			
			M	SE	SD	Min-Max	M	SE	SD	Min-Max
Exp	Male	36	15.08	0.54	3.25	5-20	14.81	0.38	2.29	11-19
	Female	33	14.88	0.50	2.86	8-20	14.70	0.47	2.70	6-20
	All	69	14.99	0.37	3.05	5-20	14.75	0.30	2.48	6-20
Con	All	30	13.87	0.56	3.05	6-19	13.70	0.48	2.65	8-19

Table A.10 – Descriptive statistics for pre-test and post-test scores on the intrinsic dimension of attitude, divided by group (Exp = experimental, Con = control) and gender.

Group	Gender	N	Pre-test				Post-test			
			M	SE	SD	Min-Max	M	SE	SD	Min-Max
Exp	Male	36	9.25	0.57	3.43	4-16	9.22	0.45	2.67	5-14
	Female	33	8.15	0.69	3.96	4-18	8.30	0.63	3.60	4-18
	All	69	8.72	0.45	3.71	4-18	8.78	0.38	3.16	4-18
Con	All	29	8.59	0.60	3.22	4-15	8.10	0.52	2.79	4-13

Table A.11 – Descriptive statistics for pre-test and post-test scores on the extrinsic dimension of attitude, divided by group (Exp = experimental, Con = control) and gender.

Group	Gender	N	Pre-test				Post-test			
			M	SE	SD	Min-Max	M	SE	SD	Min-Max
Exp	Male	35	17.29	0.55	3.25	8-24	18.37	0.51	3.03	9-25
	Female	32	16.53	0.60	3.37	9-23	17.31	0.57	3.22	10-24
	All	67	16.93	0.40	3.30	8-24	17.87	0.38	3.14	9-25
Con	All	30	16.80	0.42	2.30	14-22	18.17	0.40	2.18	15-23

Table A.12 – Descriptive statistics for pre-test and post-test scores on the utilitarian dimension of attitude, divided by group (Exp = experimental, Con = control) and gender.

Group	Gender	N	Pre-test				Post-test			
			M	SE	SD	Min-Max	M	SE	SD	Min-Max
Exp	Male	37	10.95	0.41	2.50	6-15	10.43	0.40	2.44	5-15
	Female	33	10.42	0.44	2.50	5-15	10.24	0.50	2.85	3-15
	All	70	10.70	0.30	2.50	5-15	10.34	0.31	2.63	3-15
Con	All	31	8.90	0.59	3.30	3-15	9.45	0.47	2.63	6-15

Table A.13 – Descriptive statistics for pre-test and post-test scores on the self-efficacy dimension, divided by group (Exp = experimental, Con = control) and gender.

Appendix B

Cosmo Hunters supporting materials

B.1 Debriefing

During the debriefing, facilitators promote discussion among participants through the use of the following questions. We highlight also that during the debriefing we used the questions of the post-test to simultaneously collect our data and use them to promote discussion along with the students:

- Did you have fun? What did you like the most, and what did you like the least?
- What were the winning strategies? What were the losing ones? Who won, and why did they win?
- Quick presentation of the main frequencies of the electromagnetic spectrum and their use in our daily lives
- (Display 10 image objects) Which of these objects emit light? Why?
- (Display 6 types of electromagnetic frequencies) Arrange the following types of light in order of increasing energy
- (Display the three Observation cards representing the same celestial object seen through different frequencies) Why do these cards give more points when they are collected together?
- Why do you need a corresponding Telescope card to score points with the Observation cards? For the Visible cards, it wasn't necessary. Why?
- Brief explanation of the different type of telescopes.
- (Display 10 image instruments) Which of these instruments are telescopes?

B.2 Frontal lecture

The frontal lecture included the following arguments:

- what is astrophysics, and what does it mean to observe something (difference between light emitted and reflected)
- what is light (hints of dual nature)
- fundamental aspects of electromagnetic waves (wavelength, frequency, energy and their correlation)
- the electromagnetic spectrum (different kinds of light)
- different lights for different information (when studying a celestial object, it is important to look at every radiation emitted to obtain more details about its functioning)
- a detailed description of every frequency band (which kind of telescopes are needed to observe it, an example of a celestial object that emits it and how we use it in our daily lives)

B.3 Data

The following tables present the percentage frequencies of responses for engagement (B.1) and 2 (B.3), and descriptive statistics for Q1 (B.2), Q, Q3 (B.4), Q4 (B.5), which include the number of entries (N), the mean (M), the standard error of the mean (SE), the standard deviation (SD), and the minimum (Min) and maximum (Max) values of the sample.

Value	Game group	Lesson group
1	2.3%	0.0%
2	0.0%	1.5%
3	0.0%	5.9%
4	4.7%	7.4%
5	16.3%	36.8%
6	43.0%	44.1%
7	33.7%	4.4%

Table B.1 – Percentage frequency of the engagement question.

Group	N	Pre-test				Follow-up			
		M	SE	SD	Min–Max	M	SE	SD	Min–Max
Lesson	48	6.96	0.17	1.18	5-9	8.40	0.18	1.27	5–10
Game	75	7.11	0.12	1.07	4-10	8.45	0.15	1.30	5–10

Table B.2 – Descriptive statistics for responses to Q1: “Which of these objects emit light?”

Answers	Game group		Lesson group	
	Post	Follow-up	Post	Follow-up
Radio	47.7%	41.1%	30.9%	27.3%
Microwave	59.1%	51.1%	38.2%	29.4%
Infrared	77.3%	53.3%	32.4%	33.8%
Visible	46.6%	34.4%	19.1%	14.7%
Ultraviolet	72.7%	72.2%	47.1%	42.6%
X-rays	48.9%	41.1%	32.4%	27.9%
Gamma rays	38.6%	52.2%	26.5%	36.8%
Other (incorrect)	18.2%	21.1%	45.6%	50.0%

Table B.3 – Percentage frequency of Q2: “List the types of light you know”.

Group	Phase	M	SE	SD	Min–Max
Lesson (N=45)	Pre	6.36	0.20	1.33	3–9
	Post	8.73	0.16	1.10	6–10
	Follow-up	8.40	0.25	1.66	4–10
Game (N=66)	Pre	6.82	0.18	1.47	3–10
	Post	8.05	0.16	1.31	3–10
	Follow-up	8.59	0.17	1.38	5–10

Table B.4 – Descriptive statistics for responses to Q3: “Which of these instruments are telescopes?”.

Group	Phase	M	SE	SD
Lesson (N=44)	Pre	0.30	0.07	0.46
	Post	0.93	0.04	0.25
	Follow-up	0.80	0.06	0.41
Game (N=66)	Pre	0.53	0.06	0.50
	Post	0.83	0.05	0.38
	Follow-up	0.70	0.06	0.46

Table B.5 – Descriptive statistics for responses to Q4: “What do the three images shown here represent?”.

Appendix C

AMACA supporting materials

C.1 The educational path

The main core of AMACA is an educational path aimed at engaging with astrophysics in an atmosphere of cooperation and inclusion. The approach exploits Public Engagement with Science PES principles, transforming the traditional dynamic between science communicators and the target audience. Those who were initially considered the audience became active participants, contributing directly to the design and execution of playful activities.

The AMACA educational path consists of four phases, which are visually outlined in Figures C.1 and detailed below.

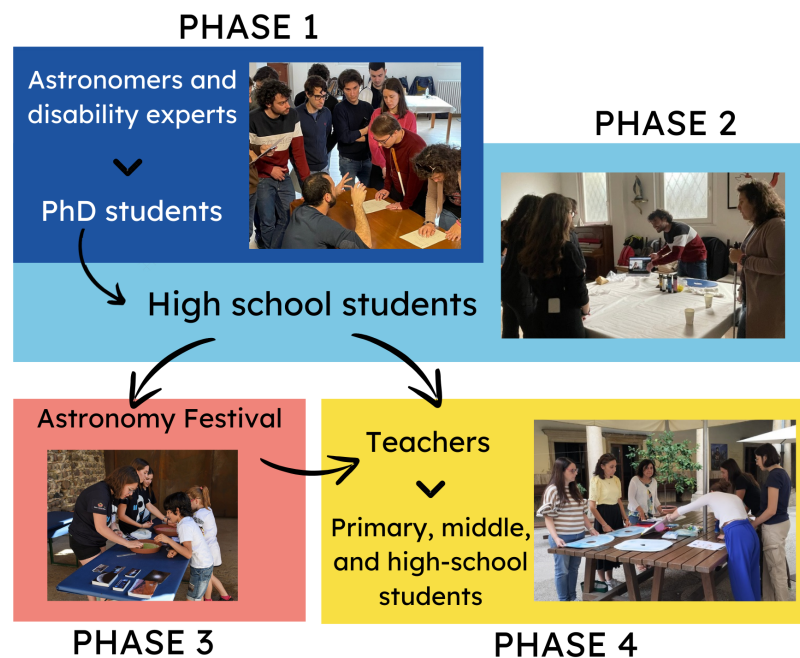


Figure C.1 – Overview of the AMACA project.

Phase 1 - From astronomers and disability experts to PhD candidates

- **Lecturers:** astronomers, psychologists, astronomy educators, disability experts, high-school science teachers
- **Target:** PhD candidates from the Physics and Astronomy Department of the Italian Universities of Padova and Bologna

A University course is held for Astrophysics PhD candidates, focusing on developing interactive and multisensory outreach activities. The lecturers are astronomers who work in the outreach and science education field at INAF, and as science teachers at high-school. Some of the lectures are held in collaboration with psychologists who work in the area of educational psychology and disabilities, and representatives of the Italian Union of the Blind and Visually Impaired (UICI - Unione Italiana Ciechi e Ipovedenti).

During the 18-hours course, the students are guided to develop playful activities related to astrophysics that involve at least two senses. Each year, the activities revolve around a different theme, the same that serves as the framework for the Astronomy Festival where they will later be presented. To guarantee accessibility, the activities are original and conceived from the start with this objective in mind, rather than taking already existing activities and modify them to add accessibility. This approach, in line with the principles of UDL, adopt a by design philosophy in which accessibility and inclusiveness are embedded in the creative process from the very beginning.

The course accommodates two cycles of evaluation and design. Once the activities are designed, they are tested with primary school students – the main age target – and people with sensory disabilities, mainly individuals with BVI. Then, the students improve the workshops according to the participants' feedback, and properly document the activities. At the end of the course, the PhD candidates provide high-school students with a training on the designed educational activities, with a careful focus on both the scientific content and its presentation mode with the public (see Phase 2). The activities are facilitated at the Astronomy Festival “The Universe in all senses” (yearly in June, see Phase 3) and delivered to school teachers in Fall (see Phase 4).

Phase 2 - From PhD candidates to high-school students

- **Lecturers:** PhD candidates and lecturers of Phase 1
- **Target:** high-school students

In this phase, the PhD candidates (assisted by the lecturers of Phase 1) become trainers and teach high-school students the activities they created and how to approach the public. The training is part of a 30-hours PCTO program that is split in two parts.

In the first part over a compressed two-day lecture (8 hours), high-school students receive an introduction on the basic astronomical concepts underlying the educational activities to learn. UICI and the Italian National Deaf Organization (ENS - Ente Nazionale Sordi) representatives also introduce best practices for an effective and respectful communication with people with sensory disabilities. High-school students are also trained on how to welcome and engage the public with the activities.

During the second part, high-school students, divided into small groups, are trained on the activities developed by PhD candidates. The overall approach exploits a playful approach, in which high-school students first experience the activity as public and then go deeper into the astronomy content with the PhD candidates. The aim is to give them the materials and resources to rehearse the activity on their own.

Phase 3 - From high-school students to the general public

- **Lecturers:** high-school students
- **Target:** visitors of the Astronomy Festival “The Universe in all senses”

The third phase happens in the context of the Astronomy Festival “The Universe in all Senses”, which takes place each year in June in Castellaro Lagusello, a small village in the North of Italy. During the three-day event, the public can attend for free various activities, including the playful educational workshops developed in previous phases, outreach conferences, sky observations, astronomy-related shows, poetry performances, exhibitions, and wall-mapping projections. High-school students are in charge of both leading the playful activities and welcoming the public at the Festival.

Phase 4 - From high-school students to teachers

- **Lecturers:** high-school students and lecturers of Phase 1
- **Target:** teachers with various educational backgrounds

The last phase of the educational path is a course for teachers of different school levels – from primary to high school – and disciplines, such as humanities, STEM, and special education. Their participation in the course is voluntary, yet recognized as part of the annual mandatory training. The aim is to promote the use of playful multisensory activities in schools, fostering the experimentation of the approach both in teachers and their students.

During the lessons, high-school students and lecturers of Phase 1, present the playful activities led during the Festival. Each teacher is trained on three different activities, which are chosen based on the age of the participants they address. The multi-disciplinarity and inclusiveness of the activities guarantee the participation of not only STEM teachers but also humanities and special education teachers.

The course is divided in three sessions: 1) high-school students demonstrate how the playful activities are conducted and convey the challenges they faced with the public during the Festival, and how they solved them; 2) an INAF astronomer holds one of the activities chosen by the teachers in their class. In this way, the teachers can assess how the students are engaged and how they react to such playful activity; 3) teachers run one activity of their choice, while one INAF astronomer is present for support and to give feedback at the end of the session.

At the end, there is one last online meeting with a final questions-and-answers session, where teachers can share doubts, concerns, report challenges, ask for help, and give feedback. The total duration of the course is 10 hours.