SIRD Società Italiana di Ricerca Didattica
Giornale Italiano della Ricerca Educativa
Italian Journal of Educational Research

Improving Historical Thinking in Primary School Pupils: Results from a Quasi-Experimental Study

Sviluppare il pensiero storico nella scuola primaria: risultati di uno studio quasi-sperimentale

Valentina Della Gala
National Institute for Documentation, Innovation and Educational Research (Indire), Firenze (Italy)
Antonio Calvani
Society for Learning and Education informed by evidence (SapiE), Torino (Italy)

Double blind peer review

Citation: Della Gala, V., Calvani, A. (2025). Improving Historical Thinking in Primary School Pupils: Results from a Quasi-Experimental Study. *Italian Journal of Educational Research*, 34, 86-94 https://doi.org/10.7346/sird-012025-p86

Received: February 10, 2025
Accepted: May 7, 2024
Published: June 30, 2025

Pensa MultiMedia / ISSN 2038-9744
https://doi10.7346/sird-012025-p86

Abstract
Over the last decades, history education has been recognized for its potential to foster cognitive skills associated with critical thinking. For those involved in educational research it is important to ask whether and how these forms of thinking can be developed as early as primary school. To address these questions, a pedagogically significant model of historical thinking, sensitive to the specific challenges faced by children, was selected. This model formed the basis for a program that was tested in several 5th grade classes. The results indicate that even students at this educational level can, when supported by formative assessment, develop higher-order thinking skills connected to historical thinking and critical thinking. They also suggest the possibility of applying this teaching model in other contexts.

Keywords: Historical thinking, history education, historical knowledge, second-order concepts, historical competencies.

Riassunto
Negli ultimi decenni, si è sempre più riconosciuto all'apprendimento della storia la potenzialità di sviluppare abilità cognitive connesse all'esercizio del pensiero critico, essenziale affinché le nuove generazioni possano orientarsi in un mondo che si presenta sempre più complesso. Per chi si occupa di ricerca educativa è importante chiedersi se e con quali metodi queste forme di pensiero possano essere sviluppate sin dalla scuola primaria. Per rispondere a queste domande è stato selezionato un modello di pensiero storico pedagogicamente significativo e sensibile alle difficoltà specifiche dei bambini, da cui è nato un programma sperimentato in alcune classi di V primaria. I risultati indicano come anche alunni di questo livello scolare possano, se supportati con momenti di valutazione formativa, sviluppare rilevanti abilità cognitive connesse con il pensiero critico e storico e suggeriscono la possibilità di trasferire il modello didattico in altri contesti.

Parole chiave: Pensiero storico, didattica della storia, conoscenza storica, concetti di second'ordine, competenze storiche.

Credit author statement
Although the work was jointly drafted and revised, paragraphs 2 and 3 are attributable to Antonio Calvani and paragraphs 4, 5 and 6 to Valentina Della Gala. The Introduction and Conclusions are jointly drafted.

1. Introduction

History, among school subjects, often receives less attention compared to those assessed through large-scale assessments (such as PISA-OECD) or to trendy topics that have gained prominence in recent decades, such as digital literacy, social-emotional skills, or climate change. This lack of consideration is compounded by the common perception that learning history in school is primarily a matter of rote memorization, as well as by the persisting belief—prevailing at least until the late 1970s—that history is too complex for young pupils (Piaget, 1933; Ballard, 1970). In Italy, an examination of some widely used primary school history textbooks reveals that authors often restrict themselves to basic descriptions of civilizational frameworks, avoiding more complex or problematic aspects (Calvani & Della Gala, 2025). However, in recent decades, a growing body of research has demonstrated that history can play a pivotal role in fostering higher-order thinking skills. Over the past 40 years, particularly in Anglo-American and German contexts, scholars have investigated how to promote meaningful historical learning among students—from primary school through to university—through both empirical studies and theoretical reflections on the epistemology of history. Across different national research traditions, there has been a shared emphasis on moving beyond the perception of history as merely the memorization of facts embedded within a more or less canonical narrative. This shift is regarded as essential to enhancing the formative function of history education. Numerous studies have shown that understanding history as a discipline—and how historians reconstruct the past—involves complex knowledge that is not acquired intuitively, yet is essential for meaningful historical learning (Shemilt, 1983; Lee, 1983; Lee & Ashby, 2001). Building on these insights, researchers have examined the full range of skills and conceptual knowledge involved in historical reconstruction and understanding, with the goal of developing and testing instructional models that support their acquisition (Stearns et al., 2000; Wineburg, 2001; Barton & Levstik, 2004; VanSledright, 2004; Ashby et al., 2005). These competencies include formulating historical questions, adopting multiple perspectives, analyzing sources[1], and understanding the role of explanation and contextualization in constructing historical narratives. Collectively, these operations are referred to in history education as historical thinking—a construct that shares several features with critical thinking (Facione, 1990), including the analysis of information, the questioning of assumptions, and the drawing of evidence-based conclusions. This research has provided robust models and frameworks that can be effectively implemented even at the primary school level. In the first part of this paper, we describe a pedagogical intervention aimed at improving students' historical knowledge, as well as their historical and critical thinking skills. In the second part, we present the results of its implementation.

2. Principles and Objectives of the Pedagogical Intervention

This study is part of a long-term initiative by the SAPIE association aimed at systematically improving all subject areas in primary schools. This approach is informed by evidence-based educational models and references (Slavin, 2008; Hattie, 2009) and aligns with the EBID (Evidence-Based Improvement Design) methodology (Calvani & Marzano, 2020; Calvani, 2022) as applied to various disciplines within the Italian educational context. The intervention was preceded by an extensive review of international scientific literature, the findings of which have been presented in previous work (Calvani & Della Gala, 2024; Della Gala, 2024; Calvani & Della Gala, 2025). Below, we briefly summarize these findings, focusing on three key areas:

- Cognitive impediments that hinder pupils' understanding of historical processes. The most significant difficulties were identified in temporal orientation, constructing causal explanations, distinguishing between history and the past, and comprehending fundamental aspects of historians' work, particularly sources analysis (Wineburg, 1991; Carretero & Voss, 1994; Lee & Ashby, 2000; Berti, 2004).
- Over the past two decades, several models of historical thinking have been developed in North America and Europe. For a synthesis of the various national perspectives, see Levesque & Clark, 2018.

1 In this paper we will use the term "source" as equivalent to the term "evidence" for the sake of ease of exposition.

• Experimental studies aimed at enhancing cognitive skills related to historical thinking. Although the number of such studies is increasing, most focus on secondary school students and employ qualitative methods (Wilson et al., 2023).

Building on this foundation, we developed a taxonomy of historical thinking with three key dimensions and five corresponding learning objectives that address common difficulties faced by children. Table 1 outlines these dimensions and objectives, along with examples of assessment methods and learning activities.

Dimensions of Historical Thinking	Objectives	Examples of Assessment and Learning Activities
Knowledge of historical facts	Knowledge of key historical facts to establish a foundational framework, including spatio-temporal identifications.	*-Identify and position the given artifacts or information on a timeline or map ...*
Knowledge of historical categories[2]	Enhancing understanding of temporality beyond simple chronological sequencing, including temporal distance, duration, and the origins of present-day objects and behaviors.	*- Look at this image of an urban landscape: which buildings are the oldest?* *- How long does it typically last, on average…a battle, a siege, a war, or a myth?* *- How long have...telephones, cars, horoscopes, Carnival... existed?*
	3. Ability to recognize the role of causes and consequences in historical accounts.	*- What reasons might explain why prehistoric humans drew graffiti in caves?* *- In an agricultural society using wooden tools, what might happen if iron tools were introduced? What consequences could result from this change?* *- If Alexander the Great had lived longer, how might history have unfolded differently?*
	4.Ability to recognize the role and characteristics of sources analysis in historian's work.	*- What source might an archaeologist or a historian use to reconstruct a specific period, event, or phenomenon?* *- Read the following statements....* —A written source always tells the truth. —Two historians studying the same period and using the same sources will inevitably reach the same conclusions. *Do you agree (Yes or No)? Explain your reasoning.*
Historical competencies	5. Sources evaluation (Wineburg, 2001).	*-Carefully read or analyze this source (e.g., an episode of Roman history narrated by Titus Livius). Does the author present their own point of view? How can you identify it? Why might they have done so?* *-Examine this account written by an ancient historian. Modern historians question the accuracy of some of these events. Which events seem the most doubtful? Why?*

Tab 1: Learning objectives framework

In addition to the five learning objectives in Table 1, we added two more: knowledge of a basic historical lexicon and improved summarization skills. For the historical lexicon we provided a list of about one hundred historical terms of general value, which are not usually found in current textbooks.

3. The PS3c Program

The experimental variable is represented by the pedagogical intervention referred to as the PS3c program[3],

2 We have decided to refer to this dimension of historical thinking as knowledge of historical categories rather than second-order concepts, a term primarily used in history education research in England to refer to this aspect (Lee, 1983), to avoid potential confusion with "historical concepts", which are also objects of study.

3 PS stands for *Pensiero Storico*, meaning Historical Thinking; 3c refers to the Italian terms *Conoscenza Fattuale*, *Categorie Storiche*, *Competenze Storico-Critiche*, which translate to factual knowledge, categorical knowledge, and historical-critical competences.

which was implemented by experimental teachers in their classes between October 2023 and February 2024. This implementation followed a training course consisting of five online sessions, each lasting two hours.

The PS3c program has the following features:

- Explicit Objectives: The program explicitly aims to achieve the objectives outlined in Table 1, with two additional objectives included. Teachers were informed of these objectives and provided with examples of formative assessment tests to support their achievement. Students were also made aware of these objectives.
- Effective Lesson Methodology: The program adopts a methodology based on the principles of effective teaching (Rosenshine, 2010; Bell, 2020), with a focus on text comprehension, following the Reciprocal Teaching model (Palincsar & Brown, 1984; Rizzo et al., 2023).
- Cognitive Optimization: The program encourages a synthetic reframing of key knowledge, aligning with the ergonomic-cognitive suggestions derived from Cognitive Load Theory (Chandler & Sweller, 1991) and Cognitive Flexibility Theory (Spiro et al., 1995).

From a practical standpoint, the program was structured as follows:

- Duration and Content: The program spanned 30 hours over five months, covering the topics typically addressed in the 5th-grade curriculum (e.g., Greeks, Italic peoples, Romans, up to the foundation of the Empire). This timeline closely mirrors the average instructional time used in common teaching practices. Teachers were asked to conduct a prior assessment of the timeframe and to select only the most relevant content from the textbooks.
- Lesson Structure: Teachers used their textbooks, supplemented with additional guidance. The 30 hours were divided into two types of lessons: thematic lessons and frame lessons.
- Thematic Lessons: These focused on understanding textbook and followed the Reciprocal Teaching model with some enhancements. After the standard Reciprocal Teaching phases (questioning, clarifying, summarizing, and predicting), teachers were encouraged to engage students in further analysis, such as identifying opportunities to expand historical vocabulary, making temporal evaluations, hypothesizing causes and consequences, or reflecting on the methods historians use to construct historical accounts.
- Frame Lessons: These focused on timelines and aimed to gradually revisit and consolidate prior knowledge considering newly introduced concepts. This was achieved by constructing continuous synthesis frameworks and conceptual maps.

4. Testing the PS3c Program

4.1 The population

The program was implemented using a quasi-experimental design, which included a control group but did not rely on randomized samples (Ary e al., 2006). A total of 12 5th grade classes participated, belonging to three different schools, with a total of 165 pupils almost equally divided between males and females (Tab. 2).

Schools	Classes	Pupils	Female %	Male %
IC Pascoli (PZ)	4	61	56%	44%
IC Milani (SA)	1	16	56%	44%
IC Gavorrano e Scarlino (GR)	7	88	43%	57%
Total	12	165	51%	49%

Tab. 2: List of schools, classes and pupils who participated in the intervention.

Of the 12 participating classes, two are multi-grade classes: one includes six 5th grade pupils, and the other includes two. One class has more than 20 pupils, two classes have between seven and 10 pupils, and the remaining classes have between 11 and 19 pupils.

The data we present pertains to 125 pupils who participated in both testing phases, excluding certified pupils. Of these, 68 belong to the experimental group (EG).

4.2 Assesment tools

At both the pre-test and post-test phases, pupils completed three assessments: one investigating historical thinking skills and knowledge (HT), a test assessing knowledge of historical lexicon (HL), and a test evaluating the ability to summarize a text (ST) (2019). The first two tests were specifically designed for this study, while the third was standardized. The HT pre-test and post-test each consisted of 80 items divided into five sections corresponding to the program's learning objectives (see Tab. 3). The HL test included 22 items, each presenting a term and requiring pupils to select the correct synonym from four options. Lastly, the ST comprised 20 items aimed at assessing students' ability to summarize three short texts. During the analysis, one point was awarded for each correctly answered item. The two sets of tests were structurally identical but differed in content. For instance, the HT pre-test contained items related to history topics from the 4th-grade syllabus, while the HT post-test focused on 5th-grade syllabus topics. In addition, a questionnaire was administered to the experimental group teachers. This consisted of 32 questions, most of which addressed the progress of the instructional intervention (e.g., variations from the planned activities, frequency of using the suggested strategies, and the degree of attention to and attainment of the program's learning objectives). Nine additional questions invited teachers to provide feedback on the program's structure, objectives, and specific elements, as well as to suggest possible improvements. The entry testing phase was conducted in October 2023, while the exit testing took place in April 2024. In both cases, the tests were administered under the supervision of a supervisor. Some pupils completed the tests in classrooms equipped with computers connected to the Internet, while others used tablets connected to the institute's Wi-Fi network in their own classrooms. Access to the platform hosting the tests was managed individually through unique identification codes. These codes ensured both the privacy of the pupils and their recognizability for data analysis purposes.

Dimensions of Historical Thinking	Objectives/Sections	N° of questions
Knowledge of historical facts	Spatio-temporal identifications	36
Knowledge of historical categories	The temporal distance between different events or phenomena, their duration, and their permanence.	14
	Causes and consequences	10
	Sources	10
Historical competencies	Sourcing	10
Total		80

Tab. 3: Sections of the HT input and output tests.

5. Results

5.1 Test Results

When analyzing the test results, the initial levels of the EG and the control group (CG) were examined to ensure that they were sufficiently similar. The comparison revealed no statistically significant differences. Specifically, the initial differences remained below 25% of the standard deviation (SD), in line with the guidelines of the What Works Clearinghouse (2022). This alignment indicates that the two groups can

be considered well-matched overall. It is worth noting, however, that the initial scores on the ST were lower than the national benchmarks.

Tables 4, 5, and 6 present the entry and exit results for the three tests included in the two testing phases: the Historical Lexicon Test (HL), the Summarizing Test (ST), and the Historical Thinking Test (HT), the latter further divided into its respective sections. When comparing the entry and exit data, we observe that pupils in both groups show improvements. However, within this overall progress, the EG demonstrates a statistically significant advantage over the CG ($P < 0.001$ for both the HL and ST tests, with an ES[4] = 0.87 in the HL test and ES = 0.66 in the ST). Among these two variables, which were likely influenced by the classroom activities conducted during the thematic lessons, the most notable result is observed in the ST. This is because the lexicon test results may have been affected by the fact that the terms used were taken from the list provided exclusively to the EG teachers during their training.

Entry HL test (22 item)		Exit HL test (22 item)			
EG Mean (SD)	CG Mean (SD)	EG Mean (SD)	CG Mean (SD)	P<	ES Cohen's d
10.92 (4.32)	10.09 (2.96)	15.94 (4.49)	13.49 (3.61)	0,001	0,86

Tab. 4: Comparison of pre-test and post-test results in the HL.

Entry ST (20 item)		Exit ST (20 item)			
EG Mean (SD)	CG Mean (SD)	EG Mean (SD)	CG Mean (SD)	P <	ES Cohen's d
21.38 (4.47)	21.63 (4.48)	25.26 (5.68)	22.16 (7,55)	0,001	0,66

Tab. 5: Comparison of pre-test and post-test results in the ST.

	Sections/subsections	Entry HT Test		Exit HT Test			
		EG Mean (SD)	CG Mean (SD)	EG Mean (SD)	CG Mean (SD)	P <	ES Cohen's d
Knowledge of historical facts	Spatio-temporal identification 36 item	18,14 (4,51)	17,77 (4,07)	24,25 (7,66)	19,90 (6,68)	0,01	0,97
Knowledge of historical categories	Temporal distance, duration and permanence 14 item	5,31 (1,78)	5,13 (1,79)	7,05 (3,68)	5,44 (2,35)	0,02	0,60
	Causes and consequences 10 item	4,17 (1,19)	4,31 (1,21)	5,48 (2,53)	5,51 (2,00)	NS	
	Sources 10 item	4,00 (1,98)	3,43 (1,62)	4,72 (2,40)	4,05 (1,81)	NS	
	Total 34 item	8.17 (2,44)	7,74 (2,32)	10,20 (3,81)	9,56 (3,02)	NS	
Historical Competencies	Sourcing 10 item	3,30 (1,64)	3,30 (1,25)	5,29 (3,22)	4,92 (2,04)	0,01	0,77
	Total 80 item	34,92 (7, 9)	33.04 (6,57)	46, 79 (11,29)	39,82 (10,82)	0,02	0,56

Tab. 6: Comparison of pre-test and post-test results in the HT at both section and subsection levels.

4 The ES index we used is Cohen's d, which remains the most widely used, with the following thresholds: small (less than 0.2), medium (between 0.2 and 0.5), and large (0.8 or higher). For a critique of these values, which tend to decrease with larger sample sizes, see Kraft (2019).

The comparison of entry and exit data for the HT test also indicates overall improvement for both the EG and the CG. However, the exit results show a statistically significant difference in favor of the EG ($P < 0.02$, ES = 0.56). A deeper analysis of the results reveals a more nuanced picture. In the spatio-temporal identification section of the exit test, the EG achieved outstanding results, particularly in recognizing and describing artifacts (e.g., Greek temples, aqueducts, baths) and in reconstructing the chronological sequence of phenomena or historical periods, with an overall ES of 0.97. This indicates that even the limited time spent reorganizing information during the frame lessons was sufficient to enhance the retention of data and the connections between them. The EG also showed a high ES in historical competencies (ES = 0.77). However, item-level analysis in this category is not feasible, as the entry and exit tests differ not only in content but also in format, with the exit test featuring predominantly open-ended questions, whereas the entry test was primarily composed of closed-ended items. The only two comparable items within this section are modeled on the History Assessments of Thinking (HATs). These questions present students with one or two sources accompanied by a prompt designed to assess a specific dimension of historical thinking. Students are asked to agree or disagree with the statement and provide a brief explanation (Smith et al., 2018). In the entry test, students were shown Gustave Doré's illustration of Dante and Virgil encountering the Minotaur and asked whether the image could be considered useful source for understanding the origins of the Minotaur legend, along with reasons for their answer. In the exit test, a similar task was presented using a detail from Rubens' painting of Romulus and Remus being suckled by the she-wolf, this time relating to the origins of Rome. The EG data show a marked improvement: correct answers increased from 24% in the entry test to 43% in the exit test for the agree/disagree question, and from 6% to 59% for the free-response question. Regarding the section on causes and consequences, no statistically significant differences were observed.

5.2 The Teachers' Evaluation of the Intervention

The teachers involved were asked to complete a 32-item questionnaire, most of which focused on the implementation of the intervention (e.g., variations, use of recommended strategies, and achievement of target objectives). Nine items specifically assessed key elements of the program and suggested areas for improvement.

The most frequently used tools were conceptual maps, followed by timelines and activities to revisit and consolidate prior knowledge. All teachers adhered to the suggested thematic lesson plans and reviewed the entry test in class. Regarding the time allocated to each program objective, teachers reported spending an average of three hours on spatio-temporal identifications, as well as the analysis of the categories of duration and permanence. More time was devoted to reflecting on the causes of historical events and phenomena, as well as the nature of historical work. For activities involving reflection on evidence, two teachers spent no more than one hour, while two others dedicated five or more hours. However, the most time was devoted to reflection and learning the specific language of history.

With regard to the perception of the achievement of the objectives, the only critical element is the work carried out on the textbook: in only two cases did the teachers state that they were certain that the children changed the way they read and used the textbook. Regarding the assessment of the program's main features (theoretical approach, objectives, evaluation system, training received, pupil engagement, and program applicability), teachers were asked to provide their feedback using a 5-point scale. The results indicate that the theoretical approach and learning objectives were particularly well-received. High levels of satisfaction were also reported concerning the training provided. Similarly, good and very good ratings were given for the program's ability to engage learners. The lowest scores were attributed to the program's applicability, which was rated as acceptable in three cases and good in the others. This result should be considered in relation to the challenges that teachers encountered during the implementation of the program, particularly its applicability to multi-grade classes and issues related to timing.

Regarding the program's positive aspects, teachers reported that it was stimulating for both pupils and themselves. One teacher identified the adoption of a more intentional approach to the discipline as its most valuable feature. Others appreciated the summary nature of the frame lessons. Most importantly, they highlighted the streamlining of the curriculum, which allowed more time for operational activities such as cre-

ating conceptual maps, using timelines, and employing images for spatio-temporal identification. While the streamlining of the textbook was one of the most positively evaluated aspects, it also posed significant challenges, as it represented a departure from the traditional practice of studying the entire textbook. One teacher mentioned having to reassure parents who were concerned that their children might miss important elements of the curriculum. In response to the final question, «Can you give us any suggestions for the reapplication of the program?», the teachers suggested extending the time frame of its implementation. One teacher, whose pupils showed the highest improvement in the comparison between entry and exit tests, recommended starting the program earlier, specifically from the third year of primary school.

6. Discussion

The implementation of the program revealed several critical issues worth highlighting.

One potential criticism of programs that define operationalized objectives is the risk of promoting a "teaching-to-test" approach. While this concern is valid, it becomes less significant when high-quality cognitive processes are involved (Hattie, 2009, 2012). Hattie has conclusively demonstrated that awareness of learning objectives is a crucial factor in enhancing educational outcomes.

The sample size was relatively small, particularly given that this was an experimental design without random selection of schools or classes. Furthermore, the sample included atypical classes (e.g., multi-grade classes) and a notable proportion of students with language comprehension delays. These factors limit the generalizability of the findings. Both the EG and the CG showed significant improvement in the HT exit test. However, this improvement may partially stem from the characteristics of the test itself. Despite being formally almost identical, differences in the content of the entry and exit tests may have influenced the level of difficulty, potentially favoring the latter.

While the EG achieved significant gains compared to the CG in the HT test as a whole—particularly in the areas of spatio-temporal identification, durations, and historical-critical competencies—its progress was less marked in two key dimensions: hypothesizing about causes and consequences and understanding the concept of source. It is unclear whether this is due to insufficient instructional focus or limitations in the assessment instrument, as only a few items targeted these variables.

Despite these criticisms, the significant improvement shown by the EG in the summarizing and lexicon tests highlights the program's potential to enhance basic literacy skills. Additionally, the EG's outcomes in spatio-temporal identification, durations and permanence, and sourcing (historical competencies) demonstrate its substantial impact on developing historical thinking.

7. Conclusion and Future Research

Studies aimed at testing programs to enhance history learning are relatively scarce compared to those conducted in mathematics or science education and are predominantly qualitative. In this paper, we show an example of how also history education research can progress from theoretical debates and anecdotal experiences to experimental research, where well-defined instructional models and hypotheses are rigorously tested against control groups. In the intervention conducted, the experimental group significantly outperformed the control group in recognizing and understanding historical lexicon, synthesizing short texts, spatio-temporal identification, analyzing durations, and developing some historical competencies. Although some areas showed no significant improvements, there are sufficient reasons to continue refining the program. This includes addressing critical points through enhanced teacher training and increasing the reliability of the findings by testing the program on larger sample sizes.

References

Ary, D., Jacobs, L. C., & Sorensen, C. (2006). *Introduction to Research in Education.* Belmont: Wadsworth.

Ashby, R., Gordon, P., Lee, P. (2005). Understanding History — Recent Research. *International Review of History Education 4*. London-New York: Routledge.

Bell, M. (2020). *The Fundamentals of Teaching: A Five-Step Model to Put the Research Evidence into Practice*. London-New York: Routledge.
Ballard, M. (Ed.) (1970). *New Movements in the Study and Teaching of History*. London: Tample Smith.
Calvani, A., & Marzano, A. (2020). Progettare per un miglioramento basato su evidenze. Quale metodologia? *Italian Journal of Educational Research*, XIII, 24, 67-83. https://doi.org/10.7346/SIRD-012020-P67
Calvani, A. (2022). La ricerca didattica può diventare rilevante per la pratica? Se sì, in che modo? *Journal of Educational, Cultural and Psychological Studies*, 26, 143-62. https://doi.org/10.7358/ecps-2022-026-calv
Calvani, A., Della Gala, V. (2025). *Potenziare e valutare l'apprendimento della storia. Percorsi per la scuola primaria*. Roma: Carocci.
Chandler, P., & Sweller, J. (1991). Cognitive load theory and the format of instruction. *Cognition and Instruction*, *8*(4), pp. 293-332. https://doi.org/10.1207/s1532690xci0804_2
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). New York: Lawrence Erlbaum Associates.
Della Gala, V. (2024). Historical Consciousness and History Didactics in Jörn Rüsen's Theory of History. *Paideutika*, (40), pp. 183–200. https://doi.org/10.57609/paideutika.vi40.7783
Della Gala, V., Calvani, A. (2024). Potenziare il pensiero storico a scuola. Un modello integrato (contenuti, categorie, competenze) per individuare e valutare gli obiettivi didattici. *Journal of Theories and Research in Education*, 19(1), 45–63. https://doi.org/10.6092/issn.1970-2221/18575
Facione, P. (1990). *Critical Thinking: A Statement of Expert Consensus for Purposes of Educational Assessment and Instruction* (*The Delphi Report*). Educational Resources Information Center (ERIC).
Hattie, J. (2009). *Visible learning: A synthesis of over 800 meta-analyses relating to achievement*. London-New York: Routledge.
Hattie, J. (2012). *Visible learning for teachers.* London-New York: Routledge.
Kraft, M. (2019). *Interpreting Effect Sizes of Education Interventions*. Ed Working Paper, 19-10. Retrieved from Annenberg Institute at Brown University http://www.edworkingpapers.com/ai19-10
Lee, P. (1983). History Teaching and Philosophy of History. *History and Theory, 22*(4), 19-49. https://doi.org/10.2307/2505214
Lee, P., & Ashby, R. (2000). Progression in Historical Understanding among Students Ages 7-14. In P. Stearns, P. Seixas, S. Wineburg (Eds), *Knowing, Teaching, Learning. National and International Perspectives* (pp. 199-222). New York: New York University Press.
Lévesque, S., & Clark, P. (2018). Historical thinking: Definitions and educational applications. In Metzger S. A., McArthur, L. Harris (Eds.), *The Wiley International Handbook of History Teaching and Learning* (pp. 119-148). Hoboken: John Wiley & Sons. https://doi.org/10.1002/9781119100812.ch3_
Palincsar, A. S., & Brown, A. L. (1984). Reciprocal teaching of compréhension-fostering and comprehension-monitoring activities. *Cognition and instruction, 1*(2), pp. 117-175. https://doi.org/10.1207/s1532690xci0102_1
Piaget, J. (1933). Psychologie de l'enfant et enseignement de l'histoire. *Bulletin Trimestriel de la Conférence Internationale pour l'Enseignement de l'Histoire, 2*, pp. 1-18.
Rizzo, A. L., Traversetti, M., & Pellegrini, M. (2023). *Potenziare la comprensione del testo. Roma:* Carocci.
Rosenshine, B. (2010). *Principles of Instruction*. International Academy of Education (IAE). https://unesdoc.unesco.org/ark:/48223/pf0000190652
Shemilt, D. (1983). The Devil's Locomotive. *History and Theory, 22*(4), pp. 1-18. https://doi.org/10.2307/2505213
Slavin, R. E. (2008). What works? Issues in synthesizing educational program evaluations. *Educational Researcher, 37*(1), pp. 5-14.
Smith, M., Breakstone, J., & Wineburg, S. (2018). History Assessments of Thinking: A Validity Study. *Cognition and Instruction, 37*(1), pp. 118-144. https://doi.org/10.1080/07370008.2018.1499646
Spiro, R., Feltovich, P. J., Jacobson, M. J., & Coulson, R. L. (1995). Cognitive flexibility, constructivism and hypertext: Random access instruction for advanced knowledge acquisition in ill-structured domains. In L. P. Steffe, J. Gale (Eds.), *Constructivism in education* (pp. 85-107). Hillsdale, NJ: Lawrence Erlbaum Associates.
Wilson, K., Dudley, D. A., Dutton, J., Preval-Mann, R. & Paulsen, E. (2023). A Systematic Review of Pedagogical Interventions on the Learning of Historical Literacy in Schools. *History Education Research Journal, 20*(9), pp. 1-27. https://doi.org/10.14324/HERJ.20.1.09
Wineburg, S. (2001). *Historical Thinking and Other Unnatural Acts: Charting the Future of Teaching the Past (Critical Perspectives on The Past)*. Philadelphia, PA: Temple University Press.
What Works Clearinghouse. (2022). *What Works Clearinghouse procedures and standards handbook: Version 5.0*. U.S. Department of Education. https://ies.ed.gov/ncee/wwc/Docs/referenceresources/Final_WWC-HandbookVer5.0-0-508.pdf