Thinking Skills and Creativity 48 (2023) 101306

Contents lists available at ScienceDirect

Thinking Skills and Creativity

journal homepage: www.elsevier.com/locate/tsc

Evaluation of the effectiveness of the ARDESOS-DIAPROVE critical thinking training programme

Silvia F. Rivas *, Carlos Saiz

Facultad de Psicología, Universidad de Salamanca, Avda. de la Merced, 109-131, Salamanca 37005, Spain

A R T I C L E I N F O

Keywords:
Critical thinking
Argumentation
Explanation
Decision-making
Problem-solving
Training and evaluation

A B S T R A C T

The changes of the 21st century arising from globalisation and technology have led to a transformation both in the professional and the personal environment. Today's highly complex world demands a more in-depth way of thinking in order to adapt to these changes. In this paper, a training system has been proposed to address these issues. In recent years, our team has progressively developed a programme for teaching critical thinking to university students, obtaining very positive results in the improvement of these skills. Our experience over the years has enabled us to modify and improve the ARDESOS programme. These modifications include the addition of a new methodology called DIAPROVE. The new approach is based on primarily working on the diagnostic, prognostic, and verification procedures. Our general objective in this study was to improve critical thinking skills using the latest version of our training programme. We have implemented this programme to see if it produced the desired changes in terms of the magnitude of effect and generalisation of skills. The effectiveness of previous versions had already been demonstrated in different studies, and our aim in this study was to show whether this latest version of the programme (DIAPROVE) substantially improved the effectiveness of the teaching of transversal thinking skills. The PENCRISAL test was the tool used to evaluate the level of improvement. To verify our hypotheses, we used a quasi-experimental design involving a control and an experimental group. Pre/post intervention measures were taken in order to determine the changes in learning. The results show that the experimental group significantly improved its performance in critical thinking for the control group, both globally and in each of the dimensions; in addition, a significant interaction effect was obtained, which further supports our starting hypothesis.

1. Introduction

The social and technological advances that are taking place in our society are giving rise to several abrupt changes. These transformations impose new demands to which we must adapt. Hence, it is important to analyse and understand these changes and the challenges they pose to education. As a result, critical thinking skills have become more crucial than ever to achieve a minimum of personal well-being and adequate professional competence in any field. Therefore, incorporating the development of critical thinking skills in higher education has become an essential need in any career profile, as they are key tools for dealing with the modern-day challenges we face at present. The acquisition of methods and ways of thinking must be prioritised in higher education, promoting

* Corresponding author.
E-mail address: silviaferivas@usal.es (S.F. Rivas).

https://doi.org/10.1016/j.tsc.2023.101306
Received 21 April 2022; Received in revised form 24 April 2023; Accepted 29 April 2023
Available online 2 May 2023

knowledge acquisition through a range of skills such as reasoning, explanation and causality, decision-making and problem-solving, and encouraging critical reflection on today's vast volume of information.

For these reasons, we propose to address such demands through the development of critical thinking. To achieve this, we need a training approach that focuses mainly on *procedural skills* and *learning outcomes*. The development of a procedural and outcome-based intervention is based on our conception of critical thinking. Previous works (Saiz, 2017, 2020) has widely elaborated on this vision, which we will outline below. In the Toulmin model (Toulmin, 1958, 2003), *argumentation*, in its different modalities, was the foundation of the field and, consequently, the skill to be improved. From the mid-1980s and the 1990s onwards, this monopoly of argumentation started to weaken, thanks to a number of landmark publications, in which the importance of considering other processes of our cognitive machinery within critical thinking (CT), such as the *skills of deciding* and *problem-solving* (Ennis, 1996; Facione, 1990; Halpern, 1989; Paul, 1995). These competencies were difficult to integrate within a theory of argumentation, at least for authors in the philosophical tradition (Govier, 2014; Johnson, 2008; Johnson & Hamby, 2015; Walton, 2006). However, most authors in the field now understand that they must address all fundamental cognitive competencies such as reasoning, explaining, deciding or problem-solving (Bassham et al., 2023; Facione, 2011; Foresman et al., 2016; Halpern & Dunn, 2023; Moore & Parker, 2021). Moreover, today this field still maintains its applied nature, whereby authors always seek to develop and improve these fundamental thinking skills (Brookfield, 2013; Halpern, 1998; Saiz, 2017; Sternberg & Halpern, 2020).

If processes such as decision-making and problem-solving are to be considered, the Toulmin model is not enough; this requires or implies the implementation of mechanisms for *action or execution* and, therefore, for the modification or changes of said reality. In recent years, further progress has been made in this direction, at least by some authors, in the sense of understanding intelligence or critical thinking, particularly as a set of skills that allow us to solve complex real-world problems, which is not possible with other ways of understanding intelligence (Halpern & Dunn, 2021). In our opinion, these recent developments in CT represent a decisive contribution, which we support. We are committed to this approach, albeit integrating critical thinking skills in a more meaningful way and giving greater prominence to explanation and causality, as the core of problem-solving in everyday life (Pearl & Mackenzie, 2018; Saiz, 2020).

We understand that "to think critically is to reach the best explanation for a fact, phenomenon or problem, in order to gain insight and solve it effectively" (Saiz, 2020, p. 27). Knowing or solving a problem requires knowing what causes are responsible said certain events or problems. If the explanation is sound, we can choose the best course of action or the best option to resolve the situation effectively by bringing about the desired change. Therefore, it is the explanation that determines the decision and the solution and, finally, the change and the well-being or achievement (Saiz, 2020). If improving CT is our *goal*, we believe that this is the best way to achieve it. Therefore, our intervention *aims* to achieve change through the development of the aforementioned skills. This is the intervention framework for the development of such skills. Let us describe in more detail how this approach shapes training.

As mentioned above, to solve a problem we need a good causal explanation (Pearl, 2009; Pearl & Mackenzie, 2018), in fact, the best possible explanation (Azar, 2019; Saiz, 2020). Once we have an explanation that provides a complete and unambiguous sense of the situation, we can safely decide on the option that will most effectively solve the problem, thus shifting the situation in the intended direction. Deciding and problem-solving strategies are difficult to distinguish, because ultimately, by reaching the best explanation, we know which course of action or option will solve the problem. Therefore, and again, the causal explanation enables us to decide or reach a solution in the best possible way, being the central mechanism allowing the solution to be reached.

In our intervention programme, we learn to solve problems according to this approach, which is summarized and specified later in Fig. 2. However, it is worthwhile mentioning some specific aspects now. In this process (Fig. 2), providing one single and full sense or fully explaining events or problems is carried out by using certain deductive principles, or rules of disconfirmation that allow causality to be established; these discarding procedures must be applied to reality, and they must contemplate what the facts reveal for each of the proposed explanations, about each event. By discarding and retaining the best explanation in each case, we can move with certainty towards the construction of a single and full sense of the situation. This enables us to predict what will happen, avoiding the possibility of being deceived as this prediction relates to what is going to happen. And if that prediction comes true, it is reasonable to think that we have found the best explanation (see the steps in Fig. 2). In this context, one learns to think critically and work with important personal and professional problem-solving projects, i.e. you work in the real world, not in the ideal of mental representation alone. Getting the best explanation needs the help of certain logical principles or certain formally equivalent rules of causality, in order to disconfirm or maintain that explanation. We must understand here that these deductive principles do not determine reality but guide us in explaining the events that take place in reality, by discarding the causal relations that the facts do not support or, on the contrary, maintain. The role of this kind of deductive inference is to test our hypotheses with the facts. The essential difference is that we do not consider the ideal world of deduction, but the real world of causality (see this development in Saiz, 2020). Charles Sanders Peirce was one of the pioneers of this integration of deduction and behaviours or facts, and of the accurate predictions that can be made from here about what will happen in a highly probable, almost certain way. After all, we argue that CT is about *solving problems effectively*, and this needs causality and explanation.

In recent years we have developed an ambitious instructional programme, in which we have made important modifications, resulting in the *third version* of the programme. As will be described below, in the *first version* we developed a programme that selectively incorporated the variables that we knew had worked best in previous studies by other authors. One of the most systematic works in this regard is that of Halpern (1998) who subsequently tested some of these variables, such as specificity, with a 15-hour instructional programme (Marin & Halpern, 2011). Unfortunately, as these authors also point out, there are few studies aimed at the development of CT in higher education. At the time, we have extensively reviewed these initiatives (Saiz, 2017) and, indeed, there are hardly any works of these characteristics (Brookfield, 2012, 2013; Butler & Halpern, 2020; Saiz, 2017). Moreover, we have not found any programmes incorporating different techniques and with classroom sessions lasting close to 60 h, as is the case of our programme. Continuing with the description of this version, we used what was already supported in the literature, but which had not been tested together, as there have been studies with interventions that tested specific variables separately and implemented a treatment that involved fewer hours (10–15 h of face-to-face practice). However, in this first version, we have incorporated all the relevant variables already tested in previous studies (Brookfield, 2012; Butler & Halpern, 2020; Halpern, 1998; Marin & Halpern, 2011; Saiz, 2017), within a classroom intervention of about 60 h. What studies use these variables together and with practical sessions of so many hours? As mentioned, we are unaware of the existence of studies that use these variables together and have practical sessions of so many hours. Some believe that thinking skills can be improved with a few hours of practice; to think that it is possible to improve CT with a few hours of intervention only shows that either they have a naive view of our higher-order cognitive machinery, or they are unaware of the difficulties in improving the fundamental and complex skills of this machinery.

After the efforts made to make this first version of our programme work but observing that these results could be improved (Saiz & Rivas, 2011, 2012), in the *second version*, as will be specified a little later, we set out to identify and test the most decisive variables of our intervention within the entire programme, something that we do not know had been addressed in the literature previously. As we said, we knew which variables were good for an intervention, and they were identified in the literature, but not which ones were the best. Therefore, our objective here was to find out which variables work best in order to give them greater prominence in our programme as a whole (Saiz et al., 2015). After achieving this goal, in the *third and final version* of the programme (DIAPROVE - see significance below), we developed an innovative instructional system. Fig. 2 below summarises, and we believe helps to understand the methodology employed. It summarises 8 steps that have proven to be very effective in problem-solving tasks. In our conception, it is argued that at the core of CT we seek to solve important (personal and professional) problems. This idea points out the importance of the context of each problem for its resolution, and how decisive the context of the problem is for its solution. What we want to clarify by referring to the context is that in most of the problems we face, when we can reasonably well delimit their context, it is possible to face their solution in a highly probable way, or with enough certainty, as long as we consider the 8 steps described in Fig. 2 (see in Saiz, 2020; Saiz & Rivas, 2016).

Solving a problem forces us to change reality, so we must correctly identify the causal relations at work in order to be able to modify them. The deductive principles used as a procedure for disconfirming reality are formally equivalent to the rules of the sufficient and necessary condition of the genius John Stuart Mill (this logical justification can be found in Saiz, 2020). What happens is that, for many, if not for most important everyday problems, a principle such as the negation of the consequent is more operative and easier to use than the rule of the necessary condition, such as disconfirmation procedures. We must not forget that our conception is based on reaching the best explanation in order to solve a problem; therefore, employing discarding procedures is essential to discover the real causes behind a given problem. However, our procedure is not only based on this, but we also need to discover all the whys and incorporate them into a system such as the construction of a causal scenario that provides a single and full sense to the problem we are trying to solve. If we achieve this, we can make predictions about future actions or behaviours, which will allow us to verify the explanation we have arrived at and solve the situation or problem effectively.

This entire methodology is set out in full in Saiz (2020), which for reasons of space, we can only summarise here; however, there is *a preview of this approach* in Saiz and Rivas (2016), where *we partially tested this third version of the programme*. This whole learning process requires an open mind and contemplating different options and tentative hypotheses; but we must bear in mind that when explanations or hypotheses are discarded by comparing them with facts, not all options are valid, only those that uniquely account for the situation or problem. In this sense, it is worth clarifying that this procedure only partly coincides with learning techniques of the scientific method. We should not forget that verification procedures in science are based on hypothetical reasoning, which is nothing more than causal reasoning plus verification (Saiz, 2020).

Of course, CT contemplates hypothetical and causal reasoning, proper to science, and all the more so when we defend that we seek the best explanation by causality for the problems that CT tries to solve. What we provide is a methodology of application to everyday contexts, without which it is difficult to use the methods of science; and this is because it is necessary to take into account more things, as has already been said, and a very important one that is not directly contemplated in the procedures of hypothesis testing, is to consider the *limitations, deficiencies or biases* of our thinking mechanisms. This Achilles' heel of our cognitive machinery must be addressed in detail if we are to claim that CT is to effectively solve the problems of everyday life; without addressing these deficiencies, it is not possible to be effective. Failure to consider these limitations of our mind centrally in the intervention prevents them from disappearing.

Biases or shortcomings occur unconsciously because they are very consolidated automatisms resulting from experience and practice. The only way to avoid these difficulties is with many practical activities, which allow us to be aware of these automatisms and avoid them. Procedural and declarative knowledge are very different in nature, and everything that has to do with inference or judgement is procedural because we have to apply or use it. Perkins (2009) reminded us some time ago that we can only say that we know something when we are able to apply it. This is the difference, we would argue, between knowing and know how to apply (based on the distinction between knowledge and wisdom; see in Saiz, 2020). Our programme in its latest version maintains these principles. This distinction probably helps to understand the difference and importance of working in instruction with comprehension tasks and production tasks. It is not possible to think well or critically if one only works with the former. For example, in a comprehension task, such as the structural analysis and evaluation of an argument given to us, we must identify the components, their relationships and whether these are correct; however, in a production task, we must look for the components and relate them, because we do not have them. The cognitive mechanisms involved in both tasks are distinct, complementary and necessary.

So far, we have made the effort to justify the conceptual basis of our programme; next, we must describe how it has been applied, in order to clearly specify our instructional methodology.

As we have already said, in previous years, we have developed a programme designed to teach critical thinking amongst university students, obtaining outstanding results in the improvement of these skills. As a result of the work carried out in the classroom, together with in-depth theoretical analysis, the *first* conceptualisation of this intervention initiative was developed: the ARDESOS programme (**AR**gumentation, **DE**cision, problem-**SO**lving in everyday **S**ituations), which was described and analysed in Saiz and Rivas (2008a). In subsequent studies (Saiz & Rivas, 2011, 2012), this instructional programme for critical thinking (ARDESOS) was further developed and evaluated. The positive performance of this didactic methodology was the stimulus that led us to complete a *second* version of the programme (Saiz et al., 2015). Following the review and improvement of this version, we introduced the new instructional methodology, ARDESOS-DIAPROVE (*version 3*), which was perfected to achieve the highest effectiveness in the teaching of critical thinking. This method yielded very positive results in the initial trials we carried out, as it substantially improved the programme and, as a consequence, the performance in critical thinking skills (Saiz & Rivas, 2016). Fig. 1 provides a complete description of this methodology.

The design of the programme is intended for implementation over a maximum of sixty in-person hours and a minimum of forty (these are the hours of face-to-face work in the classroom with the trainer), backed by the virtual support on our platform (https://www.pensamiento-critico.com/cursos); and within one of the courses, the topic of *critical thinking*. The minimum age level to which the programme can be applied is 16-17 years old.

This program is innovative in that it jointly uses the following aspects: (1) incorporation of a widely contrasted teamwork methodology, such as PBL (Problem-Based Learning; Morales et al., 2015; Saiz & Rivas, 2012); (2) incorporation of motivational techniques to reach a commitment to a specific work system and adherence to an activity schedule (Rivas & Saiz, 2016b; Olivares et al., 2013; Valenzuela et al., 2011); (3) in-depth work based on the biases, limitations and deficiencies of our cognitive system; (4) integrated learning to ensure activities are meaningful; (5) use of comprehension and production tasks, as both involve different thinking mechanisms and both should therefore be encouraged; and (6) inclusion of a wide range of problems, using different materials (see Fig. 1).

Our programme was implemented and consolidated at a university level following the above guidelines, with the aim of testing if the methodology works. However, a conceptual and empirical evaluation of the programme necessarily had to be developed. Since there was no adequate test in Spanish, we had to develop and validate one of our own, the PENCRISAL critical thinking test (Rivas & Saiz, 2012; Saiz & Rivas, 2008b). Once the psychometrically validated tool was available, the programme was tested in all its versions, producing acceptable but improvable results (Rivas & Saiz, 2016a; Rivas et al., 2016; Saiz & Rivas, 2011, 2012, 2016; Saiz et al., 2015).

However, the experience of these years of programme development led us to conclude that our intervention procedure could yield even better results. Hence, the *third version* of the programme was developed: ARDESOS using DIAPROVE methodology. Once again, the first step was to introduce changes in the conceptual model, as we have explained above.

The foundations of the third version of our instructional methodology called DIAPROVE (**DIA**gnosis, **PRO**gnosis and **VE**rification) are found in Saiz (2017, 2020), and Saiz and Rivas (2016). With the help of Fig. 2, and what is described in previous paragraphs, this methodology has been conceptually substantiated, and now a general description of its characteristics is provided. Our approach to improving critical thinking skills in the classroom consists of a general system of action that was developed in the form of a guide,

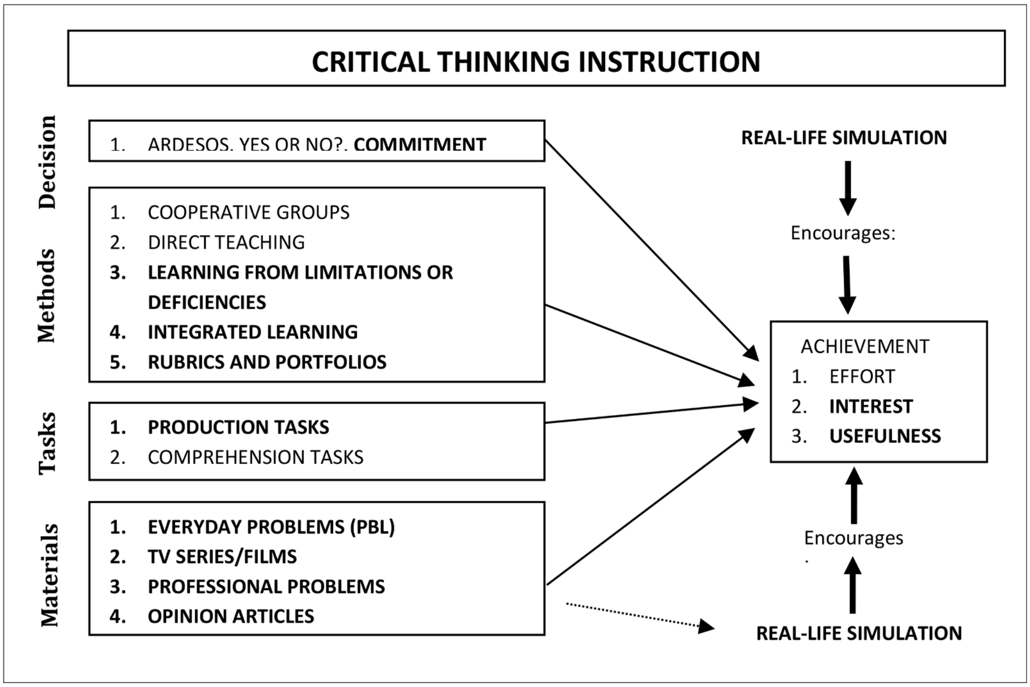

Fig. 1. Characteristics of the ARDESOS programme (modified from Saiz et al., 2015,).

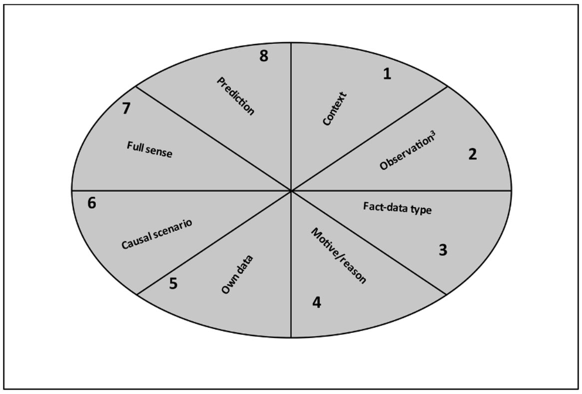

Fig. 2. CT Learning Process (Saiz, 2017, p.81).

which can be applied to any type of problem (see Fig. 2).

In the guide shown in the figure above, the steps that should be taken to achieve our aims as effectively as possible have been didactically outlined. In this process, some steps are particularly important; the first of these is the proper observation of the unambiguous facts. Another key point is to determine the fundamental reason or motive for every act or behaviour. It is necessary to causally simulate events to find meaning behind the problem situation, our third key step. The fourth and final main point of these eight is to reach a reliable prognosis. Once we reach this point with certainty, we can say that we will be able to verify our explanation of reality. This entire method ensured that our training procedure was relatively easy to apply. The specific details can be found in the already published works supporting our methodology (Saiz, 2017, 2020; Saiz & Rivas, 2016).

According to the previous approach, the objectives proposed were: (1) improving the process of learning critical thinking skills through instruction, and (2) testing the effectiveness of the critical thinking training programme (ARDESOS-DIAPROVE). In this sense, our working hypothesis was that students who had participated in the ARDESOS programme with the DIAPROVE methodology were expected to perform better in CT than those who did not participate in the programme.

Before describing the methodology used in our work, we summarize the important aspects of our research. The general objective of this work is to improve critical thinking skills through a modified version of an instructional program. Over the past few years, we have developed and tested a training program, which we have improved in three versions. The latest version (DIAPROVE) is the subject of the present study. This version was partially tested in Saiz and Rivas (2016). In that research, we only managed to improve a part of critical thinking skills, with a very modest effect on the magnitude of the improvement. For this reason, we have retested the instructional program, aiming for improvement in all fundamental critical thinking skills, and seeking an effect on the magnitude of that further improvement, in order to be able to affirm that said improvement is relevant and compensates the effort made in the intervention program. If we achieve these results, which we will present below, we can clearly state that we have an instructional program to improve critical thinking skills, which is novel, relevant, and which produces important changes in our cognitive system. We are not aware of any program as ambitious as the one we are presenting here, and we are not aware of others that offer results in the magnitude of the effect like the ones we seek with our program and that we will verify later.

Finally, the version of the program tested here differs from that of Saiz and Rivas (2016) in two fundamental aspects: (a) the fundamental skills worked on in the instruction have been better specified and have been further orientated towards explanation and causality as competencies central to problem solving; and (b) the face-to-face work in the classroom has been more directive and orienting than in the 2016 work-tested program. These changes have been introduced because our conception of the relationships that exist between the fundamental skills of critical thinking is much clearer and more precise, as is the case of the explanation, which we have already exposed. From a practical point of view, the methodology employed in the classroom by the instructor has been much more directive than in the test of prior instruction. We have seen that students need more support from the instructor, due to the high level of difficulty that critical thinking skills have. These improvements make the present work very different from the previous version of the program. Thus, our work faces a relevant problem in the field, that is, how to improve critical thinking? how to achieve this? We are not aware of any program that sets out to improve fundamental critical thinking skills, with such comprehensive instruction and inter- and intra-domain practice. The solution or answer that we give to the problem is the instruction through the DIAPROVE program. Therefore, our goals are: (1) to improve fundamental critical thinking skills through instruction, and (2) to test the effectiveness of the DIAPROVE critical thinking training program. In this sense, our working hypothesis is simple; we expect that the students who participate in the DIAPROVE program have better performance in critical thinking than those who did not participate in the program.

Next, we will present the methodology used.

2. Method

2.1. Participants

In the present study, we used a total sample of 220 university students. The experimental group included 110 first-year Psychology students from the University of Salamanca. The control group consisted of 110 first- and second-year students of the Education Bachelor's Degree at the University of Oviedo.

Both groups were analysed for gender and age equivalence. In the experimental group, 86% were women, while in the control group women accounted for 91% of the students. This difference is not statistically significant ($\chi2_{(1)} = 2.791$; $p = .71$). However, the mean age of the intervention group was 18.93 years (SD 1.775), while the age of the control group was 20.14 years (SD 2.929). This difference was statistically significant ($t_{(218)} = -0.3703$; $p = .000$). Therefore, we can state that the intervention groups were equivalent in terms of sex, but not in terms of age.

2.2. Instruments

Critical Thinking Test: to measure the magnitude of the effect and thus determine which of the two groups had a greater improvement in critical thinking skills, we applied the *PENCRISAL* test (Rivas & Saiz, 2012; Saiz & Rivas, 2008b). The *PENCRISAL* is a battery of 35 open-ended production problem situations, consisting of five factors: *Deductive Reasoning, Inductive Reasoning, Practical Reasoning, Decision-making* and *Problem-solving*, at a rate of 7 items per factor. The items in each of the factors reflected the most representative structures of the fundamental skills of critical thinking.

The items had an open-ended format where the participants had to answer a specific question by justifying their answers. Standardised correction criteria were set up, assigning values between 0 and 2 points, depending on the quality of the response. The test provided a total score for critical thinking skills and five other scores relating to the five factors. The range of values was between 0 and 72 points as the maximum for the overall test score, and between 0 and 14 for each of the five dimensions. The reliability measurements show adequate levels of precision according to the scoring methods, with the lowest Cronbach Alpha values at 0.632, and the test-retest correlation at 0.786 (Rivas & Saiz, 2012). PENCRISAL was a computer-based test administered via the Internet through the following evaluation platform: SelectSurvey.NET V5 (https://www.classapps.com/product_ssv5.aspx)

2.3. Intervention programme

As mentioned above, the present study applied the third version of the programme: ARDESOS_DIAPROVE (Saiz, 2017, 2020; Saiz & Rivas, 2016). This programme focuses on the direct teaching of skills that have been considered essential for the development of critical thinking and successful performance in daily life. To this end, we must use reasoning, explanation and the appropriate decision-making and problem-solving strategies, one of the fundamental characteristics of our intervention being the use of everyday situations for the development of these skills.

The DIAPROVE methodology incorporates three new and essential aspects: the development of observation, the combined use of facts and deduction, the effective management of disconfirmation processes, and the discarding of hypotheses. These are the foundations of our training, which requires specific teaching-learning techniques.

The intervention in the experimental group took place over 16 weeks and was designed for a total in-class time of 55–60 h. The programme was implemented in classes of approximately 30–35 students, divided into groups of four to work collaboratively in the classroom. The programme was organised into six blocks of activities: (1) nature of critical thinking, (2) problem-solving and effectiveness, (3) explanation and causality, (4) deduction and explanation, (5) argumentation and deduction and 6) decision-making and problem-solving. These blocks were homogenously configured to facilitate a global and integrated approach to the skills, thus ensuring that the different structures could be understood and used in any situation, while also facilitating a higher degree of proficiency in the mastery of each skill.

Our programme used problem-based learning (PBL) and cooperative learning (CL) as integrated didactic strategies in the teaching and learning of the critical thinking programme. The combination of these methodologies had a positive influence on students, allowing them to participate more actively in the learning process, contextualise the content, develop their problem-solving skills and abilities and increase their motivation.

To implement our methodology in the classroom, we designed a teaching system following said guidelines. Two types of tasks were performed: (1) comprehension and (2) production. The materials used for these activities were the same for all the modules of the programme. Our priority when teaching critical thinking must be making it useful for our students, and this can only be achieved through their application. Therefore, the importance of contextualising the reasoning approaches within everyday problems or situations ensures that students use them regularly and that they understand their usefulness. What we aimed to do with the materials used is to address the issues of applicability and usefulness. Accordingly, *the materials used* for the different tasks were: (1) everyday situations and (2) professional or personal problems.

The duration of the tasks was one week and consisted of cooperative group work in class, followed by revision, correction, and clarification by means of a shared discussion. In this way, the student obtained the feedback required to acquire the fundamental procedural content. In this way, we intended to make the students aware of their own thinking processes to improve them. In order to carry out the tasks, they were provided with rubrics for each one of them, which acted as a reference for the development of these tasks.

The control group consisted of students at the University of Oviedo, of the same age, academic year and university level, who were taking the subjects of psychology of learning and psychology of education. In these subjects, they developed learning, comprehension and knowledge acquisition strategies. In these subjects, the main developed skills were comprehension and acquisition. These skills were addressed in the classroom for approximately the same number of hours as the intervention of the experimental group. Similarly, the control group had virtual support in the aforementioned subjects, in which activities complemented those carried out in the classroom, to which a similar amount of time was devoted. The control group did not receive any instruction in critical thinking. This group developed and improved their learning, comprehension and knowledge acquisition skills, giving us a reasonably qualified group to compare with our intervention group.

2.4. Procedure

The application of the ARDESOS-DIAPROVE programme was carried out over a semester at the Faculty of Psychology of the University of Salamanca (experimental group). One week before the start of the programme, the PENCRISAL test was administered to all students (control and experimental group). Likewise, one week after the end of the programme, the results of the second PENCRISAL test were collected. The time between the pre-treatment and the post-treatment measure was four months for both groups. The intervention was delivered by instructors with sound experience and training in the programme.

2.5. Design

To analyse the effectiveness of the intervention, a quasi-experimental design was carried out, with pre-treatment and post-treatment measures taken by an intervention group and a control group. The statistical package IBM SPSS (version 25.0) was used for the statistical analysis of the data. Analyses were performed using the non-parametric Kruskal-Wallis alternative, as the assumptions of normality were not satisfied. As the results were similar in the p-values, we decided to use the ANOVA results as this model has greater statistical power.

3. Results

First, a descriptive analysis of critical thinking variables was carried out. Table 1 shows a descriptive summary of the scores obtained by the students in the sample, as well as the coefficients of skewness and kurtosis of their distribution.

All the variables evaluated show a high dispersion of values between the minimum and maximum scores, approaching the mean of an intermediate value within this range of values. The values of skewness and kurtosis are very low, which indicates the absence of a normal distribution of the scores in the sample. As noted in the design section, the corresponding non-parametric analyses performed gave similar values to those obtained in the ANOVA.

In the following, pre-treatment measures obtained in the control and experimental groups are described so as to determine whether the two groups had a similar initial performance on the PENCRISAL test. The data show that the groups did not significantly differ in either the total test score or in any of the 5 test factors. These data provide evidence that the groups were equivalent in terms of critical thinking performance prior to the implementation of the intervention programme (see Table 2).

Next, an ANOVA test was performed to compare the pre-post measurements and both groups for all the critical thinking variables analysed. The ANOVA results show that, in relation to the total Critical Thinking variable, there were significant differences in the pre-post measurement, with $p < .01$ (F_{CT} (1, 218) = 112.189, $p = .000$), where the post condition obtained a higher performance (M_{pre} = 23.74; M_{post} = 27.65). The same happened at an intergroup level $p < .01$ (F_{CT} (1, 218) = 96.665, $p = .000$), where the experimental group had a significantly higher mean (M_{eg} = 28.60) compared to the control group (M_{cg} = 22.79). On the other hand, the intersection between the group and the time of application also showed significant differences $p > .01$ (F_{CT} (1, 218) = 154.073, $p = .000$) (see Table 3). The results suggest that there was an interaction between the time of measurement (pre/post) and the group (experimental/control). Thus, having a similar level in the pre-test as the control group, the experimental group showed a significantly better performance in the post-treatment measure than the control group.

The results obtained in the five Critical Thinking factors according to the pre-post measurement time suggest that there were significant differences in all factors (F_{DR} (1, 218) = 43.918, p =,000; F_{IR}(1, 218) = 17.540, $p = .000$; F_{PR} (1, 218)= 65.432, $p = .000$; F_{DM} (1, 218) = 40.039, $p = .000$) except for problem-solving (F_{PS}(1, 218) = 0.828, $p = .364$). As expected, the scores obtained in the pre-intervention were lower than those obtained after the intervention. Problem-solving was the exception, and although the means increased in the post, the difference was not significant and the performance was practically the same. When comparing the groups, we

Table 1
Description of the tests and performance measures used in the study*.

Variables	Min.	Max.	Mean	SD	SK	K
PENCRISAL total pre-test	11	37	23.74	5.167	−0.099	−0.282
Pre-test deductive reasoning	0	8	2.88	1.649	.260	−0.364
Pre-test inductive reasoning	1	14	4.18	1.366	1.923	11.468
Pre-test practical reasoning	1	11	5.18	2.206	.324	−0.375
Pre-test decision-making	0	11	5.69	2.169	−0.047	.196
Pre-test problem-solving	0	11	5.82	2.169	−0.197	−0.334
PENCRISAL total post-test	12	42	27.65	7.354	.020	−0.948
Post-test deductive reasoning	0	10	3.72	2.078	.474	−0.232
Post-test inductive reasoning	0	9	4.70	1.710	.212	−0.043
Post-test practical reasoning	0	12	6.61	2.630	−0.096	−0.605
Post-test decision-making	1	11	6.66	2.119	−0.199	−0.291
Post-test problem-solving	0	11	5.97	2.219	−0.306	−0.292

* N = 220; Min: Minimum, Max: Maximum, SK: Skewness; K: Kurtosis.

Table 2
Comparison of the pre-test means according to the intervention groups for the Critical Thinking variable (PENCRISAL).

Variables		n	Mean	Standard Deviation	Mean difference	Student's *t*-test T-value	d.f.	p-sig. (bilateral)
Total Critical Thinking	e.g.	110	24.35	5.235	1.227	1.770	118	.078[NS]
	c.g.	110	23.13	5.047				
Deductive reasoning	e.g.	110	2.87	1.751	−0.009	−0.041	118	.968[NS]
	c.g.	110	2.88	1.549				
Inductive reasoning	e.g.	110	4.27	1.544	.182	.987	118	.325[NS]
	c.g.	110	4.07	1.162				
Practical reasoning	e.g.	110	5.36	2.071	.373	1.254	118	.211[NS]
	c.g.	110	4.99	2.329				
Decision-making	e.g.	110	5.92	1.772	.464	1.736	118	.084[NS]
	c.g.	110	5.45	2.170				
Problem-solving	e.g.	110	5.93	2.003	.218	.745	118	.457[NS]
	c.g.	110	5.71	2.328				

NS = non-significant ($p > .50$).

Table 3
Summary of the significance of the effects of both factors on the total Critical Thinking variable.

Variable	Means and SD		FACTOR	d.f.	Mean Square	F	p	Power	Partial Eta Squared
Total Critical Thinking	Pre-test app. 23.74 (SD 5.167)	Post-test app. 27.65 (SD 7.354)	Pre/Post App.	1 and 218	1684.820	112.189	0.000**	-	.340
	E.G. 28.60 (SD 4.38)	C.G. 22.79 (SD 4.38)	Exp/ Con Group	1 and 218	3717.820	96.665	0.000**	-	.307
	E.G. Pre 24.35 (SD 5.23)	E.G. Post 32.85 (SD 5.61)	Group/ Application	1 and 132	2313.820	154.073	0.000**	–	.414
	C.G. Pre 23.13 (SD 5.04)	C.G. Post 22.45 (SD 4.75)							

also found significant differences in line with the intervention in the five critical thinking factors (F_{DR} (1, 218) = 26.982, p =,000; F_{IR}(1, 218) = 30.116, p = .000; F_{PR}(1, 218) = 61.750, p = .000; F_{DM} (1, 218) = 59.791, p = .000; F_{PS} (1, 218) = 7.284, p = .008). *The interaction* of the two levels also showed significant differences in all factors (F_{DR} (1, 218) = 64.383, p = .000; F_{IR} (1, 218) = 28.880, p = .000; F_{PR} (1, 218)= 64.604, p = .000; F_{DM} (1, 218)= 46.250, p = .000; F_{PS} (1, 218) = 6.863, p = .009) where the experimental group had a higher performance in the post-measurement than the control group (see Table 4).

Fig. 3 reflects these interactions in detail, showing the mean values of the PENCRISAL test scores, both in total and in the five factors, according to the pre- and post-treatment levels. As shown in Fig. 3, the highest performance took place in the post-treatment condition and this increase was higher for the experimental group. The interaction in the total score and each dimension was highly significant, which further supports the influence of our intervention; moreover, the improvement in performance in the post-measure was very high compared to the pre-measure in the experimental group; on the contrary, in the control group, there was a decrease in performance in the post measure, and it only remained the same in the practical reasoning dimension, albeit at a very low level. Thus, these unambiguous and powerful interaction effects support our intervention proposal, providing clear empirical evidence of its effectiveness.

4. Discussion and conclusions

In recent years, we have developed a critical thinking intervention programme that has been modified to improve it. In general, the creation of an instructional initiative is always a lengthy process, requiring an enormous effort in terms of time and dedication, and modest in terms of its achievements. This is particularly true if the aim is to improve more complex mechanisms of our cognitive

Table 4
Summary of the significance of the effects of both factors on the practical, deductive and inductive reasoning, decision-making and problem-solving variables.

Variables	Means and SD		FACTOR	d.f.	Mean Square	F	P	Power	Partial Eta Squared
Deductive reasoning	Pre-test app. 2.88 (SD 1.64)	Post-test app. 3.72 (SD 2.07)	App. Pre/Post	1 and 218	78.627	43.918	.000**	–	.168
	E.G. 3.80 (SD 1.45)	C.G. 2.79 (SD 1.45)	Exp/Con Group	1 and 218	114.036	26.982	.000**	.999	.110
	E.G. Pre 2.87 (SD 1.75)	E.G. Post 4.75 (SD 2.17)	Group/ Application	1 and 218	116.082	64.838	.000**	.249	.229
	C.G. Pre. 2.88 (SD 1.54)	C.G. Post 2.70 (SD 1.35)							
Inductive reasoning	Pre-test app. 4.18 (SD 1.36)	Post-test app. 4.70 (SD 1.71)	App. Pre/Post	1 and 218	29.020	17.540	.000**	.986	.074
	E.G. 4.85 (SD 1.13)	C.G. 4.01 (SD 1.13)	Exp/Con Group	1 and 218	77.784	30.116	.000**	.423	.121
	E.G. Pre 4.27 (SD 1.54)	E.G. Post 5.45 (SD 1.59)	Group/ Application	1 and 218	47.784	28.880	.000**	–	.117
	C.G. Pre. 4.09 (SD 1.16)	C.G. Post 3.95 (SD 1.48)							
Practical reasoning	Pre-test app. 5.18 (SD 2.20)	Post-test app. 6.61 (SD 2.63)	App. Pre/Post	1 and 218	225.511	65.432	.000**	–	.231
	E.G. 6.79 (SD 1.69)	C.G. 4.99 (SD 1.69)	Exp/Con Group	1 and 218	354.602	61.750	.000**	.338	.221
	E.G. Pre 5.36 (SD 2.07)	E.G. Post 8.22 (SD 2.25)	Group/ Application	1 and 218	222.657	64.604	.000**	.174	.229
	C.G. Pre. 4.99 (SD 2.32)	C.G. Post 5.00 (SD 1.89)							
Decision-making	Pre-test app. 5.69 (SD 1.99)	Post-test app. 6.66 (SD 2.11)	App. Pre/Post	1 and 218	104.082	40.039	.000**	–	.155
	E.G. 6.92 (SD 1.44)	C.G. 5.41 (SD 1.44)	Exp/Con Group	1 and 218	250.509	59.791	.000**	–	.215
	E.G. Pre 5.92 (SD 1.77)	E.G. Post 7.94 (SD 1.70)	Group/ Application	1 and 218	120.227	46.250	.000**	–	.175
	C.G. Pre. 5.45 (SD 2.17)	C.G. Post 5.38 (SD 1.68)							
Problem-solving	Pre-test app. 5.82 (SD 2.16)	Post-test app. 5.97 (SD 2.21)	App. Pre/Post	1 and 218	2.475	.828	.364[NS]	.148	.004
	E.G. 6.21 (SD 1.78)	C.G. 5.56 (SD 1.78)	Exp/Con Group	1 and 218	46.475	7.284	.008**	.766	.032
	E.G. Pre 5.93 (SD 2.00)	E.G. Post 6.51 (SD 2.11)	Group/ Application	1 and 218	20.511	6.863	.009**	.742	.031
	C.G. Pre. 5.71 (SD 2.32)	C.G. Post 5.43 (SD 2.20)							

NS= non-significant correlaction (p> .50).
*Significant at 0.05.
** Highly significant 0.01.

machinery, such as those of thinking, where the difficulty is even greater. However, from our experience and the results obtained, we believe these efforts are worthwhile. For this reason, we have been engaged in this project for an extended period, and this paper describes the latest version of our critical thinking training programme, empirically demonstrating that its performance has been reasonably successful.

Our aim in the present study was to test the effectiveness of the third version of the ARDESOS-DIAPROVE programme. Our hypothesis was clearly stated: we expected the programme to improve all critical thinking skills significantly. In addition, a significant improvement in the generalisability of the competencies being trained was also expected. To achieve these objectives, we had to learn from the achievements of the previous versions of our programme. First, we had to demonstrate its overall effectiveness, and then identify the key factors that were responsible for the change achieved through the intervention. Finally, we concluded that our conception of critical thinking had to be modified to give greater prominence to the role of action and, therefore, to emphasise the importance of causality; in this way, explanation became the fundamental mechanism in problem-solving, that is, in the achievement of our goals or, in other words, in effectively producing the necessary changes to materialise those goals.

The results obtained, which we have presented above, show the expected improvements after the application of our programme. Results show that the programme obtained significant differences in overall performance and, in addition, a significant interaction effect in all dimensions and total performance. These interactions are particularly relevant to our data, as they are rarely achieved in this type of study. The first finding supports our hypotheses, and the second shows a significant improvement in the experimental group compared to the control group. Thus, not only is it positive that there is a change, but also the magnitude of this change. This significant improvement can also be observed in all but one of the dimensions, namely problem-solving. One possible interpretation of

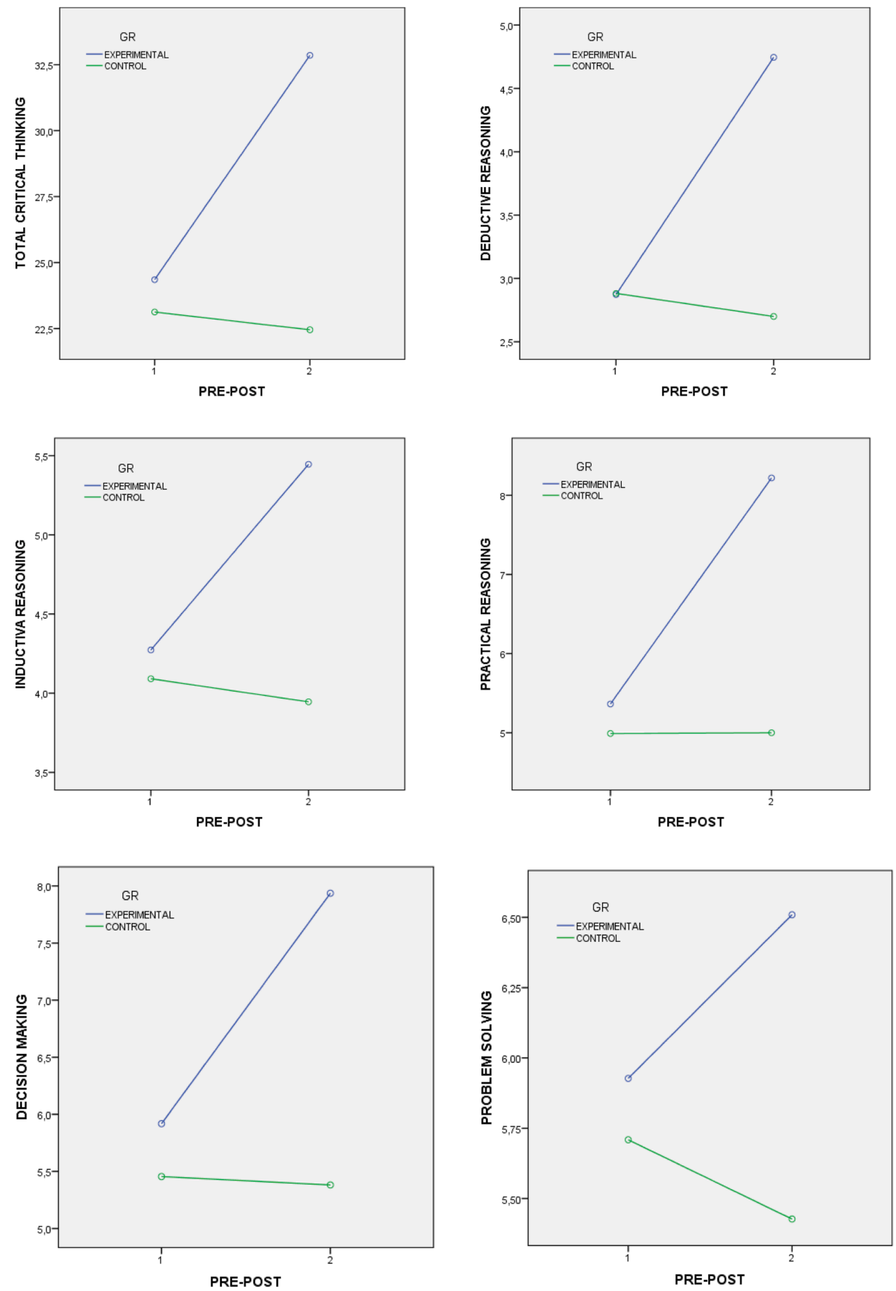

Fig. 3. Interaction effect of the PENCRISAL variable in the pre-treatment and post-treatment measures.

this lack of effect is that in our PENCRISAL test, decision-making and problem-solving included general items, which may be more difficult to distinguish in problem-solving. Therefore, these items may be less responsive to the training than those in decision-making. However, this lack of effect may also be due to the fact that the practical activities related to these two dimensions could not be differentiated. At this point, there is an intrinsic difficulty between decision-making and problem-solving, as it seems that the process in both is only differentiated by the availability or lack of options or courses of action and decisions; in the first case we are talking about decision-making, and in the second case about problem-solving. Conceptually, this distinction is not an essential problem, since this is a mere reflection of the continuum in the functioning of our cognitive machinery. However, from an empirical point of view, this distinction becomes a problem, as the measurement of each of these two dimensions cannot be properly differentiated.

It should be noted that the programme we have tested involved many hours of practice and training, between 40 and 60, which should raise awareness of how costly it is to achieve important changes or establish the right-thinking patterns. Cross-domain and intra-domain practice require many hours of dedication in order to consolidate the fundamental competencies of critical thinking. The main implication of our intervention is that nothing can be achieved without a minimum time of dedication; it may seem excessive, but it hardly compares to the macro-programmes that were developed in the 1980s (see Saiz, 2017). In those outstanding initiatives, the programmes had an average duration of 200 to 300 h, nowhere near the 40–60 that we have invested in our initiative; this is twenty per cent of the time spent in those programmes. Therefore, we can say that, if we achieved such improvements by dedicating a fifth of our time, this should also be considered an important achievement.

Naturally, there is no end to an intervention initiative, but in light of our results, we can state that there has been reasonably good progress in strengthening our programme. In terms of future implications, the foundations of the instructional system have been set. Now, the materials used in the programme need to be further developed to become more interesting and appealing. Another major implication is the need of studying how to better differentiate problem-solving and decision-making activities in the instruction. By achieving this, we will be able to obtain the significant results that we were unable to achieve in the present study. Hence, further studies could show the influence of the programme across all the dimensions, and in particular, on one that plays a key role in our approach.

Authors statement

The submission this article entitled:

Evaluation of the effectiveness of the ARDESOS-DIAPROVE critical thinking training programme implies that the work described has not been published previously, that it is not under consideration for publication elsewhere, that its publication is approved by all authors and tacitly or explicitly by the responsible authorities where the work was carried out, and that, if accepted, it will not be published elsewhere in the same form, in English or in any other language, including electronically without the written consent of the copyright-holder.

The authors assume the aforementioned statements.

Data availability

The data that has been used is confidential.

References

Azar, R. M. (2019). *La inferencia a la mejor explicación revisitada*. Alcances y límites. Universidad de Buenos Aires. http://repositorio.filo.uba.ar/bitstream/handle/filodigital/11273/uba_ffyl_t_2019_33637.pdf?sequence=1&isAllowed=y.

Bassham, G., Irwin, W., Nardone, H., & Wallace, J. M. (2023). *Critical thinking: A student's introduction* (7th ed.). McGraw Hill.

Brookfield, S. D. (2012). *Teaching for critical thinking: Tools and techniques to help students question their assumptions*. Wiley.

Brookfield, S. D. (2013). *Powerful techniques for teaching adults*. Wiley.

Butler, H. A., & Halpern, D. F. (2020). Critical thinking impacts our everyday lives. In R. J. Sternberg, & D. F. Halpern (Eds.), *Critical thinking in psychology* (2nd ed, pp. 287–324). Cambridge University Press.

Ennis, R. H. (1996). *Critical thinking*. Prentice-Hall.

Facione, P.A. (1990). Critical thinking: A statement of expert consensus for purposes of educacional assessment and instruction. "The delphi report". The California Academic Press.

Facione, P. A. (2011). *Think critically*. Prentice Hall.

Foresman, G. A., Fosl, P. S., & Watson, J. C. (2016). *The critical thinking toolkit*. Wiley.

Govier, T. (2014). *A practical study of argument* (7th ed.). Cengage Learning.

Halpern, D. F. (1989). Thought and knowledge. *An introduction to critical thinking* (2nd edition). Erlbaum.

Halpern, D. F. (1998). Teaching critical thinking for transfer across domains - dispositions, skills, structure training, and metacognitive monitoring. *American Psychologist, 53*(4), 449–455.

Halpern, D. F., & Dunn, D. S. (2021). Critical thinking: A model of intelligence for solving real-world problems. *Journal of Intelligence, 9*(2), 22. https://doi.org/10.3390/jintelligence9020022

Halpern, D. F., & Dunn, D. S. (2023). *Thought and knowledge. An introduction to critical thinking* (6th Edition). Taylor & Francis.

Johnson, R.H. (2008). Critical thinking, logic and argumentation. paper presented at the conferencia internacional: Lógica, argumentación y pensamiento crítico. Santiago de Chile, January 8-11.

Johnson, R. H., & Hamby, B. (2015). A meta-level approach to the problem of defining "critical thinking". *Argumentation, 29*, 417–430.

Marin, L. M., & Halpern, D. F. (2011). Pedagogy for developing critical thinking in adolescents: Explicit instruction produces greatest gains. *Thinking Skills and Creativity, 6*, 1–13.

Moore, B. N., & Parker, R. (2021). *Critical thinking* (13th Edition). McGraw-Hill.

Morales, P., Rivas, S. F., & Saiz, C. (2015). Critical thinking skills assessment with pencrisal test in a hybrid approach to PBL. In E. En, A. de Graaff, A. Guerra, N. A. Kolmos, & Arexolaleiba (Eds.), *Global research community: Collaboration and developments* (pp. 213–222). Aalborg University Press.

Olivares, S., Saiz, C., & Rivas, S. F. (2013). Encouragement for thinking critically. *Electronic Journal of Research in Educational Psychology, 11*(2), 367–394.

Paul, R. (1995). Critical thinking: How to prepare students for a rapidly changing world. Foundation for Critical Thinking.

Pearl, J. (2009). *Causality: Models, reasoning, and inference* (2nd ed). Cambridge University Press.

Pearl, J., & Mackenzie, D. (2018). *The book of why. the new science of cause and effect*. Basic Book.

Perkins, D. N. (2009). *Making learning whole: How seven principles of teaching can transform education*. Jossey-Bass.

Rivas, S. F., & Saiz, C. (2012). Validación y propiedades psicométricas de la prueba de pensamiento crítico PENCRISAL. *Revista Electrónica de Metodología Aplicada, 17*(1), 18–34.

Rivas, S. F., & Saiz, C. (2016a). The effects of teaching critical thinking persist over time. *Journal of Education and Human Development, 5*(1), 240–248.

Rivas, S. F., & Saiz, C. (2016b). Instrucción en pensamiento crítico: Influencia de los materiales en la motivación y el rendimiento. *Revista Latinoamericana de Estudios Educativos, 12*(1), 91–106.

Rivas, S. F., Saiz, C., & Olivares, S. (2016). Increasing critical thinking through motivation and metacognition intervention. *Revista Psicologia, Educação e Cultura, 10*(1), 304–328.

Saiz, C. (2017). *Pensamiento crítico y cambio*. Pirámide.

Saiz, C. (2020). *Pensamiento crítico y eficacia* (2a ed). Pirámide.

Saiz, C., & Rivas, S. F. (2008a). Intervenir para transferir en pensamiento crítico. *Praxis, 10*(13), 129–149.

Saiz, C., & Rivas, S. F. (2008b). Evaluación del pensamiento crítico: Una propuesta para diferenciar formas de pensar. *Ergo, Nueva Época, 22-23*, 25–66.

Saiz, C., & Rivas, S. F. (2011). Evaluation of the ARDESOS program: An initiative to improve critical thinking skills. *Journal of the Scholarship of Teaching and Learning, 11*(2), 34–51.
Saiz, C., & Rivas, S. F. (2012). Pensamiento crítico y aprendizaje basado en problemas. *Revista de Docencia Universitaria, 10*(3), 325–346.
Saiz, C., & Rivas, S. F. (2016). New teaching techniques to improve critical thinking. The DIAPROVE methodology. *Educational Research Quarterly, 40*(1), 3–36.
Saiz, C., Rivas, S. F., & Olivares, S. (2015). Collaborative learning supported by rubrics improves critical thinking. *Journal of the Scholarship of Teaching and Learning, 15*(1), 10–19.
Sternberg, R. J., & Halpern, D. F. (2020). *Critical thinking in psychology* (2nd ed.). Cambridge University Press.
Toulmin, S. E. (1958). *The uses of argument*. Cambridge University Press.
Toulmin, S. E. (2003). *The uses of argument* (updated edition). Cambridge University Press.
Valenzuela, J., Nieto, A. M., & Saiz, C. (2011). Critical thinking motivational scale: A contribution to the study of relationship between critical thinking and motivation. *Electronic Journal of Research in Educational Psychology*, 823–848. September.
Walton, D. (2006). *Fundamentals of critical argumentation*. Cambridge University Press.