New Teaching Techniques to Improve Critical Thinking. The Diaprove Methodology

Carlos Saiz *Silvia F. Rivas*
University of Salamanca

The objective of this research is to ascertain whether new instructional techniques can improve critical thinking. To achieve this goal, two different instruction techniques (ARDESOS, -group 1- and DIAPROVE, -group 2-) were studied and a pre-post assessment of critical thinking in various dimensions such as argumentation, inductive reasoning, causal, analogic, and deductive reasoning, decision-making and problem-solving was conducted. The results show that both forms of instruction functioned as expected. Both groups improved in all the dimensions of critical thinking. However the group with the new technique was only better in two dimensions with respect to group 1, namely in deductive and inductive reasoning, but not in problem-solving and decision-making. One possible reason accounting for this discrepancy was our inability to implement the new methodology in all dimensions, such that it was applied where there was greater improvement in group 2 than in group 1. The implications of this study are important because they allow us to know which new variables are crucial for instructional methodology.

Introduction

Being able to think better is perhaps one of the most exhilarating projects in our lives, but at the same time one of the most disheartening ones. Exhilarating because we cannot think of anything more interesting to investigate - anything that would bring more benefits- than a mind with maximum capacity to interact with its surroundings in order to meet its goals as best as possible. However, such a project is also disheartening because it is subject to the inverse of the principle for achieving a good investment: "buy cheap and sell dear". By contrast, we "buy very dearly and sell very cheaply. Our achievements are few and costly with regards to time and effort, and if this were not enough we do not know whether these achievements will endure (that is, our endeavors will have been in vain).

We have been addressing this issue -i.e., checking which factors make us think better- for several years, (Saiz, 2002, 2011; Saiz & Rivas, 2008a, 2008b), and the results are indeed better than what we expected, although very costly (Olivares, Saiz & Rivas, 2013; Rivas, Morales & Saiz, 2014, Rivas & Saiz, 2012; Saiz & Rivas, 2011, 2012; Saiz, Rivas & Olivares, 2015). We have also managed to demonstrate that the effects of proper thinking persist for at least four years (Rivas & Saiz, 2015). The process followed in this study has been to continue building upon previous result. First, we worked on the design of a training methodology that would incorporate most of the teachable aspects that are known to be efficient (see Halpern, 1998, 2014) and that can be applied within a reasonable length of time. With this background, we developed and applied an instructional program –ARDESOS (from the Spanish, equivalent to ARgumentation, DEcision, SOlving of problems in daily Situations) (version 1, Saiz & Rivas, 2011, 2012; version 2, Saiz et al, 2015). The aim of the first version of this initiative to intervene in critical thinking was to test the whole program. To accomplish this, we took into account the three criteria of efficacy proposed by Perkins and Grotzer (1997); namely, effect size, persistence, and transfer. Evidently, in the case of persistence we have had to wait a few years to see whether the effect of our program persisted over time (Rivas & Saiz, 2015). We also used standardized procedures for the evaluation of critical thinking for assessing the changes produced by our teaching initiative as best as possible. After several failed attempts to address different ways of assessing effect size with sufficient psychometric guarantees as regards reliability and validity indices, we have finally developed and validated a test for

evaluating critical thinking, namely PENCRISAL (Rivas & Saiz, 2012; Rivas, Morales & Saiz, 2014; Saiz & Rivas, 2008b). However, quantifying these intellectual competencies properly is again a very costly task. There are many tests that measure critical thinking but they provide very limited information owing to the closed response format used (multiple choice answers, true/false…). Accordingly, we needed a measuring instrument that would offer us as much information as possible about the thinking processes that take place when the items of a test are being answered. However, no such test was available (Saiz y Rivas, 2008b) and hence we developed an instrument to meet our requirements; PENCRISAL. Thanks to our initiative, we have been able to detect limitations in our instruction program and implement improvements to increase its efficacy.

In the present work, aimed at increasing the efficacy of critical thinking, it is precisely this change in efficacy that we are seeking in our new methodology for teaching critical thinking. We support and justify this new method empirically as a new way of working in instruction. The process followed was to develop a program, apply it and, finally, assess it. In the first version of the program, our aim was to discern whether it would work as a whole or not; whether it produced changes in our students' way of thinking or not. Moreover, we needed to know whether the overall program was effective. In this first version of the program, we were able to show that a change did take place -that critical thinking was improved after the instruction- but we were unable determine which factors or variables of the program were the most important in the change or in the improvement of critical thinking after the instruction had been implemented. (Saiz & Rivas, 2011, 2012). We were aware that there are many factors involved in a program of this nature, all of which can contribute to the final result, but during this phase in the development of the first version of the program we were unable to pinpoint which ones were the most influential and how much. This issue has now been resolved. In the second version of the program we were interested in determining exactly which factors weighed the most or were responsible for the changes or improved critical thinking prompted by our program. With that study (Saiz, Rivas and Olivares. 2015) we were able to show that at least the factors or variables such as activity and specificity enabled critical thinking to be much more efficient. The attribution of greater emphasis to cooperative participation (activity) in PBL (problem-based learning) tasks, adapted to our ends, stimulated participants' involvement in the tasks designed (in the sense of using personal or professional problems or tasks that would stimulate greater involvement of the participants in such activities). With this type of task the aim is to make the activities to be developed as ecological as possible. Moreover, the PBL-based methodology is adapted to our wish to make the instructor more"helpful" or "directive". The tasks or problems posed were delivered via a guide (a rubric) or work scheme (specificity) designed to facilitate the problem/ task-solving process. We chose the rubrics technique to incorporate this second variable because it was the one best adapted to our aims, and was the one reported to have the best practical results in other studies. Rubrics allow all the relevant aspects of a complex problem to be related as simply as possible by means of a double-entry table. Furthermore, the cells of this table are weighted for a dual purpose: to quantify the importance of each part of the task and to evaluate this qualitative task quantitatively. The results obtained when these two factors were incorporated into the program (through modified PBL and rubrics) revealed a substantial improvement in learning and performance in comparison with the previous version of the program (Saiz, Rivas & Olivares, 2015).

At this juncture in our investigation, certain questions arose. Could our latest version of the program be considered stable? If the answer was no, then which improvements should we be looking at? As a research strategy, should we be exploring other important, decisive, variables in the program? If so, how could we identify them? What might help us to discover something more positive and slightly better than the previous version of the program than good, well oriented group practice, as we have seen up to now? To accomplish this we would need to move away from the context already covered and focus on the problem from the perspective of how critical thinking has really developed. We should try to see "the forest from above the canopy and not from the boles of the trees". It could be suggested that research in critical thinking has passed through two well differentiated periods, although there is also a third one that remains to be fully covered (Saiz, 2015). Metaphorically, it could be proposed that critical thinking has passed through its infancy/adolescence and youth but has still not reached the stage of maturity. The first step involved the establishment of its main elements: from informal logic we proposed a general model of argumentation as a way of understanding critical thinking. In a second step, this key element was expanded: argumentation as the integration of other no less important mechanisms, such as explanation, decision making, or problem solving. However, this period of "youth" in the development of the field of critical thinking, with an essentially applied view of those components (argumentation, explanation…), had the unequivocal aim of ensuring that those elements would function as best as possible in order to achieve our goals (for a better understanding of this description of the development of the field of critical thinking, see Saiz, 2015).

Although there are different ways of understanding what critical thinking is, there is a little consensus about how we should achieve our goals; namely, whether doing something critically is putting those mechanisms of thinking (argumentation, explanation, decision making and problem solving) into action in order to achieve our aims. Moreover, the developmental period in which the field now finds itself, i.e. seeking to attain the stage of maturity to which we wish to advance, is still not well delineated. We must therefore pass from words to deeds, from aims to actions, from good decisions to implementation. Thinking critically involves achieving our aims in the best way possible; not just anyhow but in the best of all the ways possible. Without this assumption it would make no sense to talk in terms of "critical"; the word "thinking" would suffice. The adjective critical can be understood in different ways, but there is one that is unequivocal and will always be held above the others: critical in the sense of efficacy. This idea has not been discussed for some time and is assumed, but it has not been possible to put it into practice, at least not as well as hoped. To speak of efficacy is to assume that there is one way of doing things that is better than the rest (for example, it is not possible to believe that one way of solving a problem is as good as another because both of them would then be equally efficacious and this is not possible; there will always be one solution that is better than the rest). Efficacy implies exclusivity. This viewpoint implies a normative model of performance, a bold bid that is clearly not always defended or supported since it is carried out inconsistently or in a contradictory fashion. Accordingly, it is not possible to say that one is thinking critically and at the same time say that it is an efficacious way of doing so, instead of understanding that it must be the most efficacious way of doing so. Efficacy should be the goal of critical thinking; i.e. it should be able isolate and discover the best way to reach our goals. If we do not accept that there may be a better way of thinking critically we must also accept the uselessness of the adjective critical.

Readers should not be alarmed by this; there is no conceptual totalitarianism or defense of some absolute truth, indeed far from it. What we defend is that it is possible, within a given period of time, or in other words within a given context, to be able to offer the best solution possible to a given problem, right here and right now. Defending this is to understand that the maturity of critical thinking involves an ineluctable desire to achieve efficacy. But how can we achieve this? The intent of our proposal is to change the center of gravity of critical thinking, moving it from argumentation towards explanation, that is, towards the best explanation. Efficacy can only be achieved if the best solution is reached. This proposal arises from the change in the focus of critical thinking. In turn, that focus gives maximum importance to the best explanation; it is the only one aimed at achieving maximum efficacy. This focus is the essence of the change we have introduced in our instruction program. What is new, then, is that we put efficacy achieved through the best explanation at the center of our program. This type of educational intervention, DIAPROVE, is what we propose as new in our program and is described below (for further information about this focus on efficacy based on the best explanation, see Saiz, 2015).

It should not be overlooked that instruction in critical thinking is still dominated by argumentation. Historically, good argumentation has been the basis or model of reference in critical thinking. A general model of argumentation, taken from Toulmin (1958/2003), and adapted by most researchers in the field, serves to integrate most of the skills associated with critical thinking, (such as the different forms of deductive, inductive, causal, analogic, probabilistic… reasoning) (Baron, 2008; Walton, 2006). There are different forms of reasoning that coincide with this integrating framework, proposed by Toulim and taken up by many researchers in the field of critical thinking. The ease with which most forms of argumentation can be integrated rests on the common objective of all arguments: soundness. From this viewpoint, all reasoning seeks to establish the solvency of an idea or conclusion from the proposal of other ideas, facts or reasons. There is always an idea to be supported by others, which results in judgments with a certain solidity. The great virtue of this model is that it allows all forms of human reflection possible to be addressed in a very simple way, beyond the formalism of logic or deduction. With just a few concepts, such as conclusion, reason, soundness and not much more we can account for nearly all the components of human thinking. Sound judgment would be the basis of a good decision and this, in turn, the best starting point for reaching a solution to the problem in hand. What we stress is the interrelationship between these fundamental components of critical thinking in the way proposed earlier. Thus, we reason and make decisions in order to solve problems. The problem with this reference model (argumentation as a guide for the other components of critical thinking) is not that it does not defend efficacy as what should be distinctive within critical thinking, because it does. The question revolves around how and not around what; i.e., not what is done but how it is done. Argumentation gives maximum importance to soundness and explanation serves to achieve this. What we propose is quite the opposite; that argumentation helps to reach a good explanation, the goal is different: explain reality in the best way. As discussed below, this change is radical and in instruction it has very different practical implications. Accordingly, it is not an issue of what to achieve but of how to achieve it.

We contend that to reach complete maturity we must replace this paradigm by one of explanation, or to be more precise, one offering the very best explanation. This new approach (seeking the best explanation, not the best argument) bears the germ of change within itself. Looking for the best explanation is not an essentially conceptual goal but

rather a practical one. We are not seeking to determine the why of a problem or phenomenon for pure intellectual pursuit, which of course is of interest, but also to see how solved. Actually, what we believe should be modified is the purely conceptual view of efficacy; what we seek are the best results that can be achieved through minimum use of the best resources. And this would indeed be a good "investment theory" (it invests little and achieves a lot), simply by reviewing the approaches of many researchers working in the field of critical thinking. Upon reviewing the different approaches of many authors such as Bassham, Irwin, Nardone & Wallace, 2012; Brookfield, 2013; Facione & Gittens, 2013; Fisher, 2011; Halpern, 2014; Moore & Parker, 2014; Paul, 2012; Paul & Elder, 2012, and Tittle, 2011, we see that this approach is only conceptual. We need to apply it and to do so we need to enter the terrain of instruction. The proposal offered here is a change in our instruction program aimed at achieving maximum efficacy and maximum improvement in critical thinking

How can we attain the best explanation or maximum efficacy? Until now, we have focused on developing a teaching method and perfecting it. In this method we introduced most of the factors that are known to improve critical thinking and, later, we brought two of them into the limelight: activity and specificity. The instrument used incorporates some metacomponents (non-cognitive components) of critical thinking, such as motivational and attitudinal metacomponents and metaknowledge (Olivares, Saiz & Rivas, 2013, 2015). PBL methodology itself demands cooperative work, the use of relevant problems, continuous feedback about the learning process, and the matching of efforts to the nature of the problem. However, in PBL nothing is mandated, imposed or suggested about the real nature of the results. Nothing is suggested about how to achieve the best results and no emphasis is placed on ensuring that something will function as best as possible. No emphasis is placed on the best results, on the best way of solving a problem; only on its solution. When an engineer is asked to solve a problem involving crossing a river, that engineer is not told just to build a bridge; s/he must ensure that it will not collapse. When a psychologist is asked to choose from among several people who are aspiring to adopt a child, that psychologist is not asked to choose between heterosexual, or homosexual couples, or single-parents … s/he is asked to decide which of them will provide the greatest wellbeing for the child. Accordingly, we believe that an essential point in all types of instruction should be to place maximum emphasis on solving problems in the best way possible. We should not insist merely on good form; we must seek the best. This is the first important gap that we encounter in all programs aimed at improving critical thinking. The second lacuna, which is evident because it derives from the first one, is how this is to be done. As mentioned above, to date we have been working conceptually and in an applied way with a model of argumentation, not one of explanation. We also stated that in conceptual terms we took into account casual reasoning, hypothetical reasoning…., but mainly in the service of a solid argumentation. The advantage of argumentation is that it is able to refer to values and realities, while explanations can only refer to realities. This is why a broader, less limited example is chosen. However, this is not quite true, at least regarding our desire to produce change. Values always affect our behaviour or actions and hence, indirectly, in the long run they are realities, at least for our goals. This is because what is important are the facts, not intentions, since we have no other way of knowing about their existence except through their manifestations. Once it has been established that the paradigm of explanation is not a drawback but instead the

most suitable one, we can then see how we are to go about achieving this change.

We mentioned above that in the context of instruction all the skills of critical thinking are taught, but not in an integral way. In our program we work them through the use of tasks integrating competencies. However, this is not sufficient. Let us return to the assertion made previously: "invest little and achieve a lot". To achieve maximum efficacy, it is not necessary to deal equally with, or give the same emphasis to, all the skills of critical thinking, or even to work all of them. For example, it is easy to understand that causality cannot be given the same emphasis as generalization, because technically the latter is a lower type of reasoning than the former, among other reasons because causality is the first step in establishing causal relations. Accordingly, a generalization alone would remain "incomplete". To seek a solution for this void, we propose that the method of instruction should be modified with a view to steering it towards the development of three fundamental aspects: learning to look, learning to combine conditional and causal structures and, finally, learning to rule out explanations or hypotheses. We refer to this instruction methodology as DIAPROVE (DIAgnosis, PROgnosis and VErification), and below we address some of its characteristics.

Some years ago, an experienced language teacher in secondary school speaking on a Spanish radio program said that we struggle valiantly to help our students learn to read and write but that we never teach them to listen. This indeed seems to be the case , perhaps because oral language is felt to be natural, close, and does not need prior teaching, and because we manage well with it; we understand each other, and we can communicate fairly easily. But what this wise teacher meant was just the opposite. Oral expression is in fact very elusive, imprecise, changing and unstable. This is not the case of written language because a written message is brought to life on paper or on a screen and we can revise it, go over it again, change or correct it. Oral language, however, does not have these benefits. It is like life; it does not allow replays, but instead advances in only one direction, and the best we can do is reconstruct it, if indeed we are lucky enough to be able to do so. Precisely because of this greater difficulty, from the point of view of learning more attention should be given to oral expression. However, the opposite occurs. We find the same situation in critical thinking. Most efforts are directed towards acquiring a good knowledge of the different skills of critical thinking. By contrast, among the different teaching initiatives we have yet to find a place devoted to learning how to observe facts in a contextualized way. Oral expression occurs in reality; there are interlocutors who interact and after this interaction we can imbue it with life, but when this happens it is fluid and very dynamic. When we convert such oral information into a written testimony we have lost its essence. Context, as happens with oral interaction, is what marks the essence of all thought processes. Outside context, everything is possible and nothing is real. The problems that we face daily occur in a given space and at a given times and this means that facts are not an abstraction. However, this tends to be forgotten in teaching. The facts are merely seen "on paper" such that it is almost impossible to look at them or observe them properly. Really important facts must be discovered; they must be sought out actively because we do not merely stumble across them accidentally. Often, what is evident or important passes unnoticed, or what we believe to be relevant is not, or small but crucial details escape our notice. Discovering the really important and decisive facts requires a skill that is not learned, a skill that is not taught. We do not know how to "listen" to the data, the true facts. Without this, reasoning or reflection will never reach port. From the point of view of instruction it is crucial to learn

how to see what is really important, to ensure that nothing escapes us. This is the first bid in our new trilogy of learning.

The second and third elements cannot be separated and consist of knowing how to combine the facts of contingency relations with disconfirmation techniques. For example, let us consider a daily situation involving a problem, unfortunately quite frequent. A family has a fixed income as the only source of income; there are no other earnings and family expenses surpass that income. In this circumstance the family would be expected to have debts, but it does not. This is the situation and those are the real data. The only explanation for this could be that it is not true that the family's expenses are above its earnings, but we know that this is not the case; we know that there are more expenses than income. How can we explain this? Perhaps that someone is earning money illegally or somebody is stealing things. According to this situation (taken from a real case) we only have these two possibilities; if there is no illegal income and someone must be stealing things and we can be sure that this can be demonstrated, as indeed was the particular case. With this procedure of disconfirming hypotheses, combining facts and principles, we aim to show that it is possible to render an explanation that is not only high likely but also completely authentic, although it should not be overlooked that in this case we are in a highly specific context. One of the supports for the efficacy and the best explanation can be found precisely in this. As long as we play our cards well, critical thinking allows us to convert what is probable into what is certain.

The modifications introduced in our program are consistent with the above; our being able to achieve the best results with the fewest resources. In comparison with other methods, seeking the best explanation is tantamount to saving time and effort. In another sense, our proposal aims to offer procedures that will allow us to demonstrate that the explanation of a given problem is certain and true. With accurate observation, a correct combination of facts and principles, and precise use of disconfirmation procedures, we achieve maximum efficacy in problem-solving. This is what we call critical thinking, at least as we understand it. We all know that a given science improves as it gradually comes to depend less on statistics, on probabilities. Mathematics and some branches of physics are at the pinnacle of this hierarchy since they demonstrate and predict with certainty, not with a certain degree of probability. When we say that critical thinking is achieving the best explanation of a fact or problem, we mean that there can be no other explanation within that context. Having achieved this, the solution or prognosis is algorithmic and certain.

We were convinced that our approach is sound, but we did not know whether this modification to our program would make the learning process and performance even better. This was an empirical issue that had to be subjected to testing. This, therefore, is the main goal of the present work: to see whether with these changes the improvements are better. In principle, it seem reasonable to conjecture that if a method provides such efficacy it can be expected to stimulate maximum interest; it will provide maximum incentives or motivation, and it will predispose students to make every effort to achieve the best results. In theory, it would appear that few things would arouse more interest than reaching our goals or solving our problems in the best way and with absolute certainty. However, we said "in theory"; in practice we must wait and see. As stated by Russell "Most people would rather die than think" (a very well-known and familiar adage). Our work seeks to determine what happens in practice, whether we really are incentivizing critical thinking or not, whether our work is a good bet or not. Below we describe the method used to check this.

As goals in the present work, we first try to improve the learning process in critical thinking, based on changes introduced in the instruction. A further aim is to assess the efficacy of the critical thinking instruction program, modified via the introduction of the DIAPROVE methodology. Note that our ARDESOS program has been tested previously on two occasions. Thus, according to our proposals, we aimed to see the following:

A better performance in the competencies of critical thinking in both instruction groups with the ARDESOS program, with and without DIAPROVE

A better performance in critical thinking by students who had worked with the DIAPROVE methodology in our program in comparison with those who had not.

Method

Participants

Initially, we started out with a group of 186 students enrolled in the first year of the psychology degree given at the University of Salamanca. Of these 95 belonged to intervention group 1 and the other 88 to intervention group 2. These groups were divided at the middle by surnames: group 1, A-L; group 2, M-Z. This administrative division was taken with no other modifications.

For different reasons, such as the lack of information, incomplete tests, etc, 20 participants were lost from the population. Accordingly, the definitive sample comprised 166 students, of which 87 belonged to instruction group 1 and the other 79 to instruction group 2.

We analyzed gender and age equivalence in both groups. Group 1 had 70% women and 17% men, while group 2 comprised 61% women and 19% men. This difference is not statistically significant ($\chi 2 = 0.262, p = .374$).

The mean age of the participants in group 1 was 19.17 years (SD, 2.805), while in group 2 it was 19.28 years (SD, 2.428). This difference is not statistically significant either ($t(164)=-0.259$, $p=.796$). Accordingly, the intervention groups can be said to be equivalent as regards gender and age.

Intervention programs:

ARDESOS v2 program for the development of critical thinking skills.

As commented previously, the ARDESOS program for the development of critical thinking skills implemented in the study was a revised version, ARDESOSv2 (Saiz, Rivas & Olivares 2015), of the initial version of the program (Saiz & Rivas, 2011, 2012).

The ARDESOS program is based on a method of the direct teaching of critical thinking skills that allows the transfer of such skills to all contexts. This transfer is achieved by using daily problems, both personal and professional, in the instruction. These skills essentially involve procedural knowledge and hence our intervention focuses more on the learning of processes than on contents. The latter (contents) are evidently necessary for all learning, but they are rigid and static, while processes are flexible and allow us to create alternatives, since each person can generate a different way to access the same information. These skills are transferrable and, once acquired, can be applied to any area of knowledge.

The learning-teaching strategy incorporated in our intervention program is Problem-Based Learning (PBL). This activity revolves around the discussion of the different problem situations designed in the program, and the learning of critical thinking skills emerges from the experience of having worked in these situations.

The ARDESOS program focuses on the teaching of the skills we feel to be indispensable for the development of critical thinking and hence for our daily lives. To accomplish this it is necessary to use relevant forms of argumentation, procedures for explaining and checking hypotheses, and good

strategies for problem solving and decision making. These skills are the basis of our intervention.

The program was implemented over 16 weeks and was designed for application in the classroom for a period of 55-60 hours, over a four-month period. Two sessions were held per week, each lasting an hour and a half. The program was applied in classes of 30-35 students, divided into groups of 4 for class work in collaborative groups. The sessions were organized in 4 activity blocks:

1. Argumentation – 4 weeks
2. Conditional and analogical reasoning – 4 weeks
3. Explanation and causality – 4 weeks
4. Decision making – 4 weeks.

The planning of the activities was initiated at the very start of the academic year and rubrics were provided for each of them. The class work focused on the development of these guided activities, with orientation from the instructor, whose only mission was to clarify any doubts that might have arisen while the task was being carried out; it was *not* the instructor's job to solve the problem. Assessment was carried out weekly, offering feed-back 2-3 days after the end of the sessions, suitably indicating the solution to each activity. This assessment was quantitative, as specified in each rubric.

The main characteristics of the program for instruction in critical thinking included:

Methods:

- Team work
- Direct teaching of skills
- PBL Methodology
- Learning from limitations and mistakes
- Integrated learning of skills
- Rubrics and portfolios.

Tasks:

- Comprehension tasks
- Production tasks

Materials:

- Daily problems
- Television series/films
- Professional problems
- Articles on opinion

ARDESOS v2. Program with DIAPROVE methodology

We stated above that critical thinking must replace the reference model of argumentation by one of explanation in order to achieve the best efficacy in problem-solving. This change imposes a different way of teaching in some aspects, although not in all of them. The program developed by us was modified by adding specific factors or variables, and this should not be overlooked. We did not remove anything that was already known to be functioning well, and evidently we added what we believed would strengthen the program. One factor hitherto not taken into account, at least, in the way done here, was to give maximum priority to achieving maximum efficacy. This made it necessary to seek a different mode of instruction strategy. Explaining is not the same as arguing, although epistemologically they may be hard to differentiate. However, this will not be of much use and hence the onus lies in achieving the best explanation. As stated earlier, being able to achieve maximum efficacy should be the most motivating part or at least would add a plus in this field. According to the idea of "investing exactly what is necessary in order to achieve the maximum benefit", it may be assumed that an important change should occur in our skills for solving problems and in the perseverance with which we use them. Regarding the instruction itself, the methodology included three new essential aspects: the development of observational powers, the combined use of facts and deduction, and efficacious management of disconfirmation procedures or the ruling out of hypotheses. These are the lynchpins of this new form of instruction, which requires more specific techniques for teaching or learning. A detailed explanation of these techniques would distract us from the main principles supporting them. What should be highlighted is that the methodology focuses on the three above aspects before other factors. And what we wished to know was whether this method would produce a greater change than its predecessor.

Instruments

PENCRISAL, the test of critical thinking.

As a measurement of effect size, and with a view to determining in which of the two groups the greatest increase in the improvement of critical thinking skills was produced, we applied the PENCRISAL test (Rivas & Saiz, 2012; Rivas, Morales & Saiz, 2014; Saiz & Rivas, 2008b).

PENCRISAL is a battery comprising 35 problem situations with an open-response format and is structured on the basis of five factors: *Deductive Reasoning, Inductive Reasoning, Practical Reasoning, Decision Making and Problem Solving,* with 7 items per factor. The items of each of the factors cover the most representative structures of the fundamental skills of critical thinking (Cronbach Alpha = .632; test-retest: r=.786, Rivas & Saiz, 2012)

The format of the items was open, such that the students had to answer a specific question, accounting for their answers. Accordingly, we used standardized correction criteria that assigned values between 0 and 2 points, depending on the quality of the answer. The test offered a total score for the skills of critical thinking and another 5 scores referring to the five factors. The range of values was between 0 and 72 points, as the maximum limit, for the global score on the test, and between 0 and 14 for each of the five scales.

The PENCRISAL test was administered in computerized form, through the Internet, using the SelectSurvey.NET assessment platform: http://survey .pensamiento-critico.com/Login.aspx

Procedure

The study was carried out during the 2013-2014 academic year with students from the first year of the degree in Psychology at the University of Salamanca. The instruction programs were applied in the second semester during the course entitled *The Psychology of Thinking*. In order to obtain a baseline of the performance of the students in critical thinking, we applied the PENCRISAL test one week before the start of instruction in both groups (pre-treatment measurement).

As previously indicated, our main aim was to demonstrate that the DIAPROVE methodology would increase the efficacy of the ARDESOS program. To accomplish this, here we used two groups of students, one of them receiving the ARDESOS program in its latest version (Saiz, et al., 2015) and the other one following the ARDESOS program, adding the DIAPROVE methodology.

We then applied the two instruction programs over a four-month period. Finally, one week after the end of the intervention the second measurement of critical thinking skills was taken (post-treatment), using the same test.

The program was implemented with the two groups at the same time and was delivered by two expert instructors with extensive experience and training in the program. The basis of the program (skills worked, the order of introduction of the skills, assessment systems, etc) was the same for both groups. What was different was the DIAPROVE methodology (G2) introduced in group 2 and the quasi-control groups of the ARDESOS program (G1).

Statistical analyses

For the statistical analysis we used the IBM SPSS Statistics 2.1 package. The statistical tools and techniques used were as follows: frequency tables and percentages for qualitative variables, with the Chi-squared homogeneity test; exploratory and descriptive analyses of quantitative variables with the test of goodness of fit to the normal gauss model and box diagrams for the detection of outliers; the usual descriptive statistical tests (means, standard deviations, etc) in numerical variables; *Students't* test for the significance of differences. Finally, in order to analyze the effect of the application of the programs we used a repeated measurements ANOVA test with an inter- and intra-treatment factor of the values of the PENCRISAL variable.

Results

Regarding the descriptive analysis of all the variables we were interested in, we observed that in the case of intervention group 1 only post-treatment inductive reasoning and post-treatment problem solving did not fit the normal curve model ($p > .050$) and pretreatment Induction and post-treatment Decision-making showed a slight deviation, with $p > .010$. The other variables fitted the normal curve model suitably, with $p < .050$. Regarding intervention group 2, only post-treatment Decision-making showed a slight deviation, although still with $p > .010$. Below we offer tables summarizing the descriptive statistics of the variables studied in both intervention groups. Next, we show the mean pre- values in group 1 and group 2 in order to check whether there are differences in the different factors measured between the two groups before the intervention. To do so, we performed a contrast of means for independent samples with Student's t test (see Table 3).

The results show that at the start intervention groups 1 and 2 only showed significant differences in the practical reasoning factor ($t(164)= 2.125$, $p=.035$). In the rest of the variables and in the total of the scale no statistically significant differences were seen. It may therefore be concluded that with the exception of this factor both groups are equivalent in the critical thinking variables. We then performed an ANOVA test, in which we compared the pre-post measurements and groups between each other. We also calculated the effect of the change in both cases.

Regarding the results obtained in the total and the five critical thinking factors according to the pre-post measurement time, significant differences were observed in the total and in all the factors: (F_{TOT} (1,164)= 130.213, p = .000; F_{DR} (1,164)= 54.706, p = .000; F_{IR} (1,164)= 30.638, p = .000; F_{PR} (1,164)= 75.592, p = .000) (F_{DM} (1,164)= 91.590, p = .000 and F_{PS} (1,164)= 4.150, p = .043). As expected, the scores obtained in the pre- condition were worse than those observed after the intervention (see Table 4).

In the comparison between groups, significant differences were only observed in the deductive reasoning factor (F_{DR} (1,164) = 4.339, p = .039) where the best performance was seen for intervention group 2 (M_{G2}=4.53;M_{G1}=3.08); this was consistent with what was expected. Finally, in the interaction of the two levels we only found significant differences in the inductive reasoning factor (F_{IR} (1,164)= 5.294, p = .023), group 2 being the one that showed the highest mean (M_{G2pre}=4.37, M_{G2post}=5.62; M_{G1pre}=4.75, M_{G1post}=5.26). According to this, in both groups the measurements varied in the same way, with the exception of Inductive Reasoning, where G2 began below and ended above G1 and hence improved more than the other group.

Table 1: Descriptive statistics of the pre- and post- PENCRISAL variables for intervention group 1

Variables	n	M	SD	95% CI	Range (min-max)	K-S p-sig
Pencrisal TOTAL-Pre-treatment	87	28.85	6.576	[27.45, 30.25]	14-46	.450
Deductive-Pre- treatment	87	3.39	1.882	[2.99, 3.79]	0-10	.088
Inductive-Pre- treatment	87	4.75	1.416	[4.45, 5.05]	2-8	.041
Practical-Pre- treatment	87	7.06	2.788	[6.46, 7.65]	1-13	.199
Decision-making-Pre- treatment	87	6.77	2.316	[6.28, 7.26]	2-12	.192
Problem-solving- Pre - treatment	87	6.89	2.254	[6.40, 7.37]	2-11	.073
Pencrisal TOTAL-Post- treatment	87	33.64	6.891	[32.17, 35.11]	16-50	.905
Deductive- Post - treatment	87	4.74	2.212	[4.26, 5.21]	0-10	.135
Inductive- Post - treatment	87	5.26	1.393	[4.97, 5.56]	1-8	.006
Practical- Post - treatment	87	8.82	2.326	[8.12, 9-12]	3-14	.170
Decision-making- Post - treatment	87	8.63	1.905	[8.23, 9.04]	4-12	.010
Problem solving- Post - treatment	87	6.39	2.581	[5.84, 6.94]	0-10	.001

Table 3: Comparison of mean pre- values in overall critical thinking and its corresponding factors as a function of the intervention groups.

Variables		n	M	SD	Difference between means	Student's t test T value	d.f.	p-sig. (bilateral)
Total	g.1. g.2.	87 79	28.85 27.23	6.576 5.267	1.623	1.743	164	.083 NS
Practical Reasoning	g.1. g.2.	87 79	3.39 2.78	1.882 1.781	.606	2.125	164	.035*
Deduction Reasoning	g.1. g.2.	87 79	4.75 4.37	1.416 1.360	.380	1.759	164	.080 NS
Induction Reasoning	g.1. g.2.	87 79	7.06 6.58	2.788 2.510	.475	1.150	164	.252 NS
Decision-making	g.1. g.2.	87 79	6.77 6.86	2.316 2.074	-.091	-.265	164	.792 NS
Problem-solving	g.1. g.2.	87 79	6.89 6.63	2.254 2.101	.252	.743	164	.458 NS

Table 2: Descriptive statistics of the pre- and post-´PENCRISAL variables for intervention group 2

Variables	n	M	SD	95%CI	Range (min-max)	K-S p-sig
Pencrisal TOTAL-Pre-treatment	79	27.23	5.267	[26.05, 28.41]	14-38	.687
Deductive-Pre- treatment	79	2.78	1.781	[2.39, 3.18]	0-7	.096
Inductive-Pre- treatment	79	4.37	1.360	[4.06, 4.67]	0-7	.064
Practical-Pre- treatment	79	6.58	2.510	[6.02, 7.14]	1-11	.333
Decision-making-Pre- treatment	79	6.86	2.074	[6.40, 7.33]	2-11	.233
Problem-solving- Pre - treatment	79	6.63	2.101	[6.16, 7.10]	2-10	.053
Pencrisal TOTAL-Post- treatment	79	33.78	6.887	[32.24, 35.33]	16-45	.473
Deductive- Post - treatment	79	4.34	2.062	[3.88, 4.80]	0-12	.193
Inductive- Post - treatment	79	5.62	1.876	[5.20, 6.04]	1-11	.061
Practical- Post - treatment	79	8.75	2.488	[8.19, 9.30]	2-14	.182
Decision-making- Post - treatment	79	8.71	2.119	[8.23, 9.18]	3-12	.036
Problem solving- Post - treatment	79	6.37	2.573	[5.79, 6.94]	1-11	.237

Table 4: Summary of the significance of the effects of both factors on the total variable, practical, deductive and inductive reasoning, decision making and problem solving

Variables	Means & SD		Factor	df	MC	F	p	Power	Eta2 partial
Total	Applic. pre. 28.08 (s.d. 6.02.) G.1. 28.03 (s.d. 4.33) G.1. pre. 28.85 (s.d. 6.57) G.2. pre. 27.23 (s.d. 5.26.)	Applic. post. 33.71 (s.d. 6.89) G.2. 33.71 (s.d. 4.75) G.1. post. 33.64 (s.d. 6.89) G.2. post. 33.78 (s.d. 6.88)	Applic. Pre/Post Group/ Group 2 Group /Applic.	1,164 1,164 1,164	2666.889 45.443 64.407	130.213 .723 3.145	.000** .396 NS .078 NS	1.000 .135 .422	.443 .004 .019
Deductive Reasoning	Applic. pre. 3.10 (s.d. 1.85) G.1. 3.08 (s.d. 1.33) G.1. pre. 3.39 (s.d. 1.88) G.2. pre. 2.78 (s.d. 1.78)	Applic. post. 4.55 (s.d. 2.14) G.2. 4.53 (s.d. 1.47) G.1. post. 4.74 (s.d. 2.12) G.2. post. 4.34 (s.d. 2.06)	Applic. Pre/Post Group 1/Group 2 Group /Aplic.	1,164 1;164 1 ,164	174.317 20.696 .932	54.706 4.339 .292	.000** .039* .589 NS	1,000 .544 ,084	.250 .026 .002

Variables	Means & SD		Factor	df	MC	F	p	Power	Eta2 partial
Inductive Reasoning	Applic. pre. 4.57 (s.d. 1.39) G.1. 4.55 (s.d. 1.01) G.1. pre. 4.75 (s.d. 1.41) G.2. pre. 4.37 (s.d. 1.36)	Applic. post. 5.43 (s.d. 1.64) G.2. 5.44 (s.d. 1.12) G.1. post. 5.26 (s.d. 1.39) G.2 post. 5.62 (s.d. 1.87)	Applic. Pre/Post Group 1/Group 2 Group/Applic.	1,164 1,164 1,164	64.886 .012 11.212	30.638 .005 5.294	.000** .945 NS .023*	1.000 .051 .628	.157 .000 .031
Practical Reasoning	Applic. pre. 6.83 (s.d. 2.66) G.1. 6.82 (s.d. 2.51) G.1. pre. 7.06 (s.d. 2.78) G.2. pre. 6.58 (s.d. 2.51)	Applic. post. 8.68 (s.d. 2.40) G.2. 8.68 (s.d. 1.66) G.1. post. 8.62 (s.d. 2.32) G.2. post. 8.75 (s.d. 2.48)	Applic. Pre/Post Group 1/Group 2 Group/Applic.	1,164 1,164 1;164	287.679 2.522 7.486	75.592 .279 1.967	.000** .598 NS .163 NS	1.000 .082 .286	.316 .002 .012
Decision-making	Applic. pre. 6.81 (s.d. 2.19) G.1. 6.81 (s.d. 1.59) G.1. pre. 6.77 (s.d. 2.31) G.2. pre. 6.86 (s.d. 2.07)	Applic. post. 8.67 (s.d. 2.00) G.2. 8.67 (s.d. 1.38) G.1. post. 8.63 (s.d. 1.91) G.2. post. 8.71 (s.d. 2.11)	Applic. Pre/Post Group 1/ Group 2 Group/Applic.	1,164 1,164 1,164	284.968 .580 .004	91.590 .100 .001	.000** .752 NS .971 NS	1.000 .061 .050	.358 .001 .000

Problem-solving	Applic. pre. 6.77 (d.t. 2.18) G.1. 6.75 (d.t. 1.58)	Applic. post. 6.38 (s.d. 2.56) G.2. 6.37 (s.d. 1.77)	Applic. Pre/Post	1,164	11.960	4.150	.043*	.526	.025
			Group 1/Group	1,164	1.575	.185	.668 NS	.071	.001
	G.1. pre. 6.89 (s.d. 2.25) G.2. pre. 6.63 (s.d. 2.10.)	G.1. post. 6.39 (s.d. 2.58) G.2. post. 6.37 (s.d. 2.57)	Group/Applic.	1,164	1.080	.375	.541 NS	.093	.002

NS= correlation not significant (*p*> .50) *Significant at the .05 level ** Highly significant at the .01 level

Discussion

One of the most important aspects in the context of instruction is the assessment of programs, since this allows us to check whether the initiative implemented is efficacious or not (Ennis, 2003; Fisher, 2001; Halpern, 2003; Lomask & Baron, 2003; Norris, 2003). In this research context, the present work aimed to improve instruction for teaching critical thinking, testing new instructional techniques in the ARDESOS program, which in previous studies has proved to be efficacious.

From the results obtained it may be seen that the effects of the ARDESOS program applied to both groups are significant. In this sense, the instruction program can be said to be efficacious since it elicits a significant improvement in competencies in critical thinking after the intervention.

Both group 1 and group 2 had similar performances prior to the intervention, except with regard to the skills of practical reasoning, in which the performance of group 1 was better than that of group 2. Thus, the efficacy of the ARDESOS program can be confirmed since both groups improved their skills in critical thinking, regarding both the whole test and the factors of deductive, inductive and practical reasoning, decision-making and problem-solving. These observations are consistent with the findings reported by other authors (Diestler, 2012, Luckey, 2003; Perkins, 2001; Swartz, 2003; Teays, 1996; Weinstein, 2003).

With respect to the DIAPROVE methodology introduced as a technique aimed at further improving the efficacy of the ARDESOS program, the results are not so promising. This methodology only improved our students' skills in deductive reasoning, since the participants receiving this type of instruction improved more than those who only received the ARDESOS program. Nevertheless, this methodology did not have a significant effect for the other factors. Regarding the interaction, this was likewise only significant for one factor, inductive reasoning, in which group 2 improved much more than group 1. In the other factors, the measurements of both groups varied similarly. These data can be accounted for by the fact that in group 2 the skills that could really be worked completely were precisely those of the deductive and inductive dimensions. Because they could not be worked within the dynamic of the groups, the other factors could only be partially developed. Accordingly, the data are only consistent with the intervention finally applied.

Although the results reported here are preliminary, we believe that in future studies, and when this methodology has been fully incorporated into the ARDESOS program, efficacy can be improved, since even though almost no significant differences emerge it is possible to note a trend in which the means of the students who received the methodology were higher than those who did not, although they do not reach significance.

In conclusion, it seems reasonable to suggest that these results are fairly promising and will prompt us to continue working in the direction of improvement and of the application of DIAPROVE in the ARDESOS program in order to further evaluate the efficacy of the intervention for the development of competencies in critical thinking.

References

Baron, J. (2008). *Thinking and deciding.* (4th ed.). Cambridge, UK: Cambridge University Press.

Bassham, G., Irwin W., Nardone H., Wallace J. M. (2012). *Critical thinking: A student's introduction* (5th ed.). New York: McGraw Hill.

Brookfield, S. D. (2013). *Powerful Techniques for teaching adults.* San Francisco: Wiley.

Diestler, S. (2012). *Becoming a critical thinker: A user friendly manual* (6th edition). Upper Saddle River, NJ: Pearson.

Ennis, R. H. (2003). Critical thinking assessment. In D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice* (pp. 293-313). Cresskill, NJ: Hampton Press.

Facione, P. A., & Gittens, C. (2013). *Think critically* (2nd ed.). New York: Pearson.

Fisher, A. (2011). *Critical thinking. An introduction* (2nd ed.). Cambridge, UK: Cambridge University Press.

Fisher, A. (2001). Assessing thinking skills. In A. L. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (3rd ed., pp. 541-545). Alexandria, VA: Association for Supervision & Curriculum Development.

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. (2003). The "how" and "why" of critical thinking assessment. In D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice* (pp. 355-366). Cresskill, NJ: Hampton Press.

Halpern, D. F. (2014). *Thought and knowledge. An introduction to critical thinking* (5th ed.). New York: Psychology Press.

Lomask, M.S., & Baron, J. B. (2003). What can performance-based assessment tell us about students' reasoning? In D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice* (pp. 331-354). Cresskill, NJ: Hampton Press.

Luckey, G. M. (2003). Critical thinking in colleges and universities: A model. In D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice* (pp. 253-271). Cresskill, NJ: Hampton Press.

Moore, B. N., & Parker, R. (2014). *Critical thinking.* (11th ed.). New York: McGraw-Hill.

Norris, S. P. (2003). The meaning of critical thinking test performance: The effects of abilities and dispositions on scores. In D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice* (pp. 315-329). Cresskill, NJ: Hampton Press.

Olivares, S., Saiz, C., & Rivas, S.F. (2013). Encouragement for thinking critically. *Electronic Journal of Research in Educational Psychology, 11* (2), 367-394. doi: 10.14204 /ejrep.30.12168_

Olivares, S., Saiz, C., & Rivas, S.F. (2015, unpublished manuscript). *To what extent do motivation and metacognition improve critical thinking?* Universidad de Salamanca, Salamanca.

Paul, R. (2012). *Critical thinking: What every person needs to survive in a rapidly changing world.* Tomales, CA: Foundation for Critical Thinking.

Paul, R., & Elder, L. (2012). *Critical thinking: Tools for taking charge of your learning & your life.* (3rd ed.). Upper Saddle River, NJ: Prentice-Hall.

Perkins, D. (2001). Thinking for understanding. In A. L. Costa (Ed.), *Developing minds: A resource book for teaching thinking* (3rd ed., pp. 446-450). Alexandria, VA: Association for Supervision & Curriculum Development.

Perkins, D. N., & Grotzer, T. A. (1997). Teaching intelligence. *American Psychologist, 52* (10), 1125-1133.

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. [Validation and psychometric properties of the PENCRISAL critical thinking test]. Retrieved from http://www.pensamiento-critico.com/archivos/ validacionpencrieng.pdf

Rivas, S.F., Morales, P. & Saiz, C. (2014). Propiedades psicométricas de la adaptación peruana de la prueba

de pensamiento crítico PENCRISAL. *Avaliação Psicológica, 13* (2), 257-268.

Rivas, S.F., & Saiz, C. (2015, manuscript submitted for publication). *Permanence of the effect of teaching critical thinking.* Universidad de Salamanca. Spain.

Saiz, C. (2002). Enseñar o aprender a pensar. *Escritos de Psicología, 6,* 53-72.

Saiz, C. (Ed., 2011). *Pensamiento crítico: conceptos básicos y actividades prácticas.* Madrid: Pirámide.

Saiz, C. (2015). Efficay, the heart of critical thinking. En Domínguez, C. (ed.). *Pensamento crítico na educação: Desafios atuais. (Critical thinking in education: Actual challenges).* Pp. 159-168. Vila Real: UTAD. Retrieved from http://www.pensamiento-critico.com/archivos /ebookutad.pdf

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. [Assessment in critical thinking: a proposal for differentiating ways of thinking]. *Ergo, Nueva Época, 22-23* (marzo-septiembre), 25-66. Retrieved from http://www.pensamientocritico.com/archivos/evaluationCTergoENGLSH.pdf

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., & Olivares, S. (2015). Collaborative learning supported by rubrics improves critical thinking. *Journal of the Scholarship of Teaching and*

Learning. 15 (1), 10-19. doi: 10.14434/josotl.v15i1.12905

Swartz, R. J. (2003). Infusing critical and creative thinking into instruction in high school classrooms. In D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice* (pp. 207-251). Cresskill, NJ: Hampton Press.

Teays, W. (1996). *Second thoughts: Critical thinking from a multicultural perspective.* Mountain View, Calif.: Mayfield Publishing.

Tittle, P. (2011). *Critical thinking. An appeal to reason.* New York: Routledge.

Toulmin, S. E. (1958/2003). *The uses of argument* (updated ed.). Cambridge, MA: Cambridge University Press.

Walton, D. (2006). *Fundamentals of critical argumentation.* Cambridge, UK: Cambridge University Press.

Weinstein, M. (2003). Some foundational thoughts on critical thinking practice. In D. Fasko (Ed.), *Critical thinking and reasoning. Current research, theory, and practice* (pp. 273-289). Cresskill, NJ: Hampton Press.