Journal of the Scholarship of Teaching and Learning, Vol. 11, No. 2, April 2011, pp. 34 - 51.

Evaluation of the ARDESOS program: An initiative to improve critical thinking skills

Carlos Saiz[1] and Silvia F. Rivas[2]

Abstract: It is desirable that reasoning, problem-solving and decision-making skills should form an integral part of our private and professional lives. Here we show how these skills can be improved through the use of the ARDESOS program. To test the effect of the program, we have also developed an assessment test (PENCRISAL). Our results are going in the desired direction. The ability to decide and make inductive inferences was improved, and this improvement was also seen in argumentation, although indirectly. In the future we must therefore improve our interventions in all factors, but in particular those referring to induction and problem-solving. Much remains to be done from the procedural point of view, but the preliminary results are very promising and we are convinced that our initiative has a good conceptual grounding.

Keywords: critical thinking, transference, assessment, instruction, reasoning, problem-solving, decision-making.

I. Introduction.

For some time we have been developing an intervention program with the aim of improving critical thinking skills. The first results of our efforts can be found in Nieto and Saiz (2008). As a result of the implications of those data, together with a profound theoretical analysis, we elaborated a first substantial conceptual modification of this intervention initiative, which henceforth will be referred to as *ARDESOS* (from the Spanish, equivalent to **Ar**gumentation, **De**cision, **So**lving of problems in daily **S**ituations) and which is described and discussed in Saiz and Rivas (2008a). However, this is only the first step in our journey, and it needs to be justified in order to be able to propose a solution to the important, still open, and unresolved problem of improving our capacity for critical reflection. Thus, in this Introduction section we shall proceed as follows. First, we shall briefly sketch a background of the field of enquiry, after which we shall delimit the sources of our work and justify it. Once we have justified our work from the viewpoint of intervention, we shall discuss the objectives of the present work, the problems addressed, and the solutions proposed.

The drive of human beings to improve their intellectual capacity is as old as the first cultures in which teaching played a role. Perhaps the place where this quest received the greatest attention, at least within Western tradition, was in Ancient Greece, with the first Pre-Socratic learned men. From these beginnings to the present day, important efforts have been made to improve our thinking skills, such as projects, involving *Instrumental Enrichment or Project Intelligence* (Nickerson, Perkins & Smith, 1985), among others. During the last two decades, ways of teaching students to think were developed, based on work addressing critical thinking, such as that of Ennis (1996). Currently, this line of critical thinking is probably the most fruitful

[1] Facultad de Psicología. Universidad de Salamanca. Avda. La Merced, 109. 37005 Salamanca. Spain. csaiz@usal.es
[2] Facultad de Psicología. Universidad de Salamanca. Avda. La Merced, 109. 37005 Salamanca. Spain. silviaferivas@usal.es

as regards initiatives of this kind (for a review of the justification, see Saiz, 2002a). Our work on instruction belongs to this tradition.

Critical thinking is a still heterogeneous concept and there are an excessive number of ideas about it (see Ennis, 1987; Lipman, 2003; McPeck, 1981, 1990; for a review of the concept sees Nieto & Saiz, 2008, and Saiz & Rivas, 2008a). Ours is explicit in the definition: we understand that "*Critical Thinking* is a process involving a search for knowledge through reasoning skills, problem-solving and decision-making that will allow us to achieve the desired results more efficiently" (Saiz and Rivas, 2008a, p. 131). Inference, or judgement, is what we essentially find behind the concept of thinking. However, is thinking only reasoning? Some authors believe so (Johnson, 2008), while others do not, assuming that solving problems and making decisions are activities that also form part of thought processes (Baron, 2005; Halpern, 1998, 2003; Mercier and Sperber, 2010). In this latter view, achieving our goals does not depend solely on one intellectual dimension. All three are important: not only reasoning, but also decision making and problem solving. From the viewpoint of psychology, these skills form part of our most valuable cognitive tools, something that is not contemplated in the more philosophical traditions. The difference between these two approaches is epistemological. Each responds differently to the following question: Should we have a theory about reasoning or about action? From the point of view of philosophy, we should work on a theory about reasoning, while from the psychological perspective the focus should also be on a theory about action (Saiz, 2009). Let us explore this issue further.

Normally, we think in order to solve problems or to achieve our goals. A problem can be solved by reasoning, but also by planning a course of action or selecting the most suitable strategy for the situation at hand. Thus, as well as reasoning we must also make decisions to solve our problems. Choosing is one of the most frequent and important activities that we engage in. Accordingly, we prefer to give it the importance it merits in a definition of thinking. Solving problems demands many intellectual activities, such as reasoning, deciding, planning … From this point of view, thinking is reflection and action; we can say that *thinking is reasoning and deciding in order to solve problems* (Saiz, 2009). However, the efficiency of our thinking, thinking critically, requires other components. In order to delimit the meaning of thinking efficiently, it is necessary to seek aspects outside the core, such as those depicted in Figure 1.

In Figure 1 we can find three concepts of the previous definition plus two other important components: motivation and meta-knowledge (attitudes are usually understood as dispositions, inclinations….; something close to motives but also to meta-knowledge). The fundamental nucleus of Critical Thinking continues to be that which has to do with skills, in our case reasoning, problem solving, and decision-making. But why introduce concepts of other types, such as motivation, in a description of Critical Thinking? Several years have passed since it was observed that, when addressing Critical Thinking, focusing only on skills does not allow all its complexity to be unveiled. The aim of the scheme in Figure 1 is to provide conceptual clarity to the adjective "critical" in the expression *Critical Thinking.* If we understand that *critical* refers to *efficacy*, we must also see that efficacy cannot be achieved merely with skills. Other protagonists must be brought into play, and at different times. Alone, intellectual capacities do not achieve the efficiency associated with the notion of "critical". First, for such capacities to be set in motion (for us to think) we must desire this to happen ("knowing begins with wanting", as one of our professors once said). Thus, motivation enters the game before skills; it sets them in motion. In turn, meta-knowledge allows us to direct, organize, and plan our skills in a profitable way, and it acts once skills have begun to function. The final goal must always be a desirable knowledge of reality; greater wisdom. The author who has best posed the role of these components is Halpern (1998, 2003), on whose work we based the development of our overall conception of what Critical Thinking is.

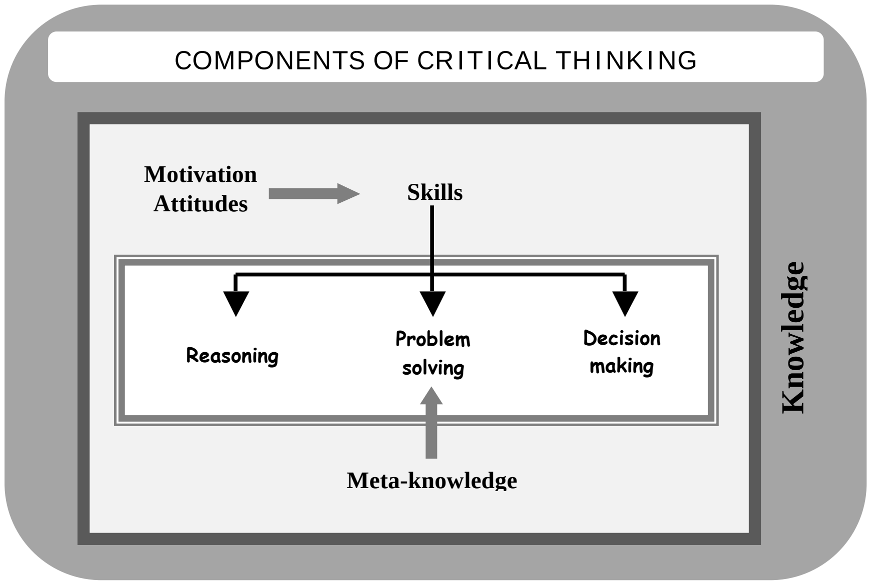

Figure 1. Components of Critical Thinking.

We believe that the fact of referring to the components of Critical Thinking, and time differentiating skills from motivation and meta-cognition, can help in the conceptual clarification we are seeking. On one hand, we specify which skills we are talking about, and on the other hand, which other components (other than *thinking*) are related to them, or even overlap them. We must be aware of the futility of the illusion of finding "pure" mental processes. Planning a course of action, an essential feature of meta-knowledge, demands reflection, prognosis, choice, comparison and assessment. … Is this not thinking? The different levels or dimensions of our mental activity must be related, or integrated. We believe that our avenue of enquiry will turn out better this way. Accordingly, our efforts towards conceptual clarification are directed to achieving that integration of the components of thinking. Our aim is to be able to identify what is substantial in thinking in order to determine what it is we can improve and assess.

Our initiative tries to overcome two drawbacks of other programs that we believe to be especially relevant. One of them is the time that many programs dedicate to intervention. Macro-programs (for example, the Instrumental Enrichment Program) aimed at teaching people how to think are limited in that they require many classroom hours for the development of intellectual skills. In most cases, similar lengths of time for working with our students are simply not available. Our instruction program can be completed in some sixty hours, which in most academic contexts is an attainable length of time. The other problem is the decontextualization of the programs designed to teach thinking, that is, the use of artificial activities. Most of the activities proposed in such programs are exercises and tasks that have little to do with the sphere of daily life. Such a departure from "reality" poses serious problems as regards instruction efficiency. One way of solving this is to propose a *problem-based learning approach*, employing tasks taken from daily situations, as we describe below.

The procedure used by us consists in directly teaching each of the three main skills mentioned above (see figure 1). These skills are essentially procedural knowledge, such that "doing" is more important than "describing how to do things". Also, since our aim is to generalize such skills to daily contexts, they should be practiced in different domains to increase the possibility of their use in any of them. However, although important, these two activities - practising and doing so in different contexts- are not as important as a third one. The most important terrain of our actions is the sphere of daily activities, common situations, and it is here where our main interest lies: ensuring that the main skills will be used in these situations. Thus, what we are seeking is above all for the transfer to materialize in daily life. If the difficulty in generalizing our intellectual skills lies in the huge difference between the field of acquisition and that of application, we should strive to eliminate or reduce such a difference to a minimum. This will be the core of our instruction, aimed at the greatest generalization possible of our essential skills to daily situations. Thus, the pillars of our intervention are *a lot of practice, interdomain practice, and tasks based on daily situations*, together with *biases* or *distortions*, which will be addressed later

To reduce the difference between the domain of intervention and that of application, it is necessary to use tasks or problems akin to those encountered in daily life. In many cases, the materials, tasks or problems used are of the following type: "If the card has an even number on one side, it will have a vowel on the other, and it is true that it does not have an even number on one side, hence it will not have a vowel on the other side"**.** Such exercises are too artificial. We can learn a form of conditional reasoning with the previous task (in this case, one of the most common fallacies, negation of the antecedent) but it will be very difficult to apply it in our daily lives. However, we use a daily problem (at least for those familiar with court juries) in which the same fallacy appears. It is very likely that in similar daily situations such a conditional error would be readily identified. Let us explore the following task (adapted from Halpern, 2003):

> **Example 1.** A jury must decide on the guilt or innocence of someone accused of murdering a young woman on March 18, studying the arguments and proofs of the prosecution and of the defence. The relevant data in the case are as follows. The accused has a perfect alibi as from 23:00 h for that 18th of March. In the trial, proof in favour of and against the accused is heard by the jury members. Also, all the witnesses related to the place of the crime are interrogated. However, as well as focusing on these data and testimonies, both lawyers make every effort to emphasize the actual time of death of the victim. Concerning this point, the police investigators establish that the murder occurred before 23:00 h. After deliberating, the jury emits a verdict of guilty. The main argument on which they base their decision states that the accused would be innocent if the crime had occurred after 23:00 h but since the crime took place before that time, the accused is clearly not innocent but guilty.
>
> Did the jury make a reasonable decision? Explain why or why not.

A task-problem such as Example 1, which simulates a daily situation, has at least two advantages; it may be interesting *per se*, and its context is similar to a real-life one. If we can manage to stimulate greater interest in the task, this will affect the efficiency of learning, and if we can manage to ensure that the distance between the academic context and the real world is minimum we can achieve a greater application of the acquired or improved abilities. Once the intervention has been posed as a procedure of simulation of our daily functioning, we must now detail it in terms of specific, not general, skills such as reasoning, problem solving and decision-making. Let us start with the first of these.

As mentioned, reasoning is an important mechanism of thinking. Nevertheless, there are many forms of reasoning. In Example 1 above, we illustrate one such form –conditional reasoning-, which is probably the most important of all, since explanations (causal reasoning) and the procedures of hypothesis testing (hypothetical reasoning), to mention just two, depend on it. However, this task of introducing reasoning into our daily functioning is much harder than what can be gathered from Example 1. Although we use specific daily situations for some of the types of reasoning in our instruction, we face the problem of argumentation -informal or practical reasoning- (Johnson, 2000, 2008; Govier, 2005; Saiz, 2002b; Walton, 2006). In our daily activities, we must assess or produce arguments to defend points of view, positions, theses, etc. Argumentation is possibly the most common and natural form of human reasoning. Its importance is such that it has been a focus of research along a large part of the tradition of critical thinking; that encompassed within informal or practical reasoning. In 1958, Toulmin (2003) proposed a model of argumentation that continues to be a reference for human reasoning today (see Blair, 2009). In the tradition of critical thinking, this scheme of Toulmin's has persisted and has become more understandable and applicable. However, what is missing is its use as an integrating framework of all modes of human reasoning. We have done this, in the way to be explained below. In our daily reflecting, when defining a point of view or explaining certain observation, we use analogic, causal or conditional arguments, to cite the most frequent. In teaching reasoning, what is the best way to proceed? Working separately with each form of reasoning, or integrating them in a general scheme? In most cases, we argue by integrating specific forms of reasoning within an argumentative or general explanatory line. Since this is a natural way of reflecting, let us proceed in the same way in our instruction. Other authors use another form of direct teaching of argumentation that is also efficient, although not so much (see Bensley et. al. 2010). We have opted for an argumentation task that includes different forms of reasoning, together with specific tasks for some of them that are difficult to integrate into an argumentative text. We have selected or drafted argumentative texts of some 2,000-3,000 words in length in which there are different argumentative structures: propositional, causal, analogical … In a task of this kind, we can explore different forms of argumentation in a single text in a natural way.

By integrating most of the reasoning within a general model of analysis we achieve a better understanding of the principles, and hence greater efficiency in the assessment of their soundness. However, it remains for us to describe the *problem solving* and *decision-making* tasks. We shall therefore recourse to Examples 2 and 3.

The tasks designed for these other basic skills involve situations common to many people. Again, we are attempting to simulate real problems within the academic context in order to facilitate transfer. In Example 2, we pose a common problem in which efficient strategies for solving problems must be brought into play. A general solution system, such as that of Bransford and Stein (1993) is perfectly applicable to situations such as that seen in the example.

> **Example 2.** Julia is 28 years old and only has primary education and she has been working for 10 years in a ceramics factory with three shifts (morning, afternoon, night) that rotate every 23 days. She earns 950 euros/month. She is tired of working so hard and hates the poor schedule and the low pay. She is disillusioned about her job prospects because she knows that with her academic background

> she is unable to aspire to a better job situation. She has decided to see how she can improve her professional status and to do so she has given herself some time to think about it. She has decided to go on the dole for a year and a half. Unfortunately, she has a 35-year mortgage to pay off and some money to pay for the car she has brought recently. Such debts really do not allow her to be out of work for any length of time.
> What would Julia's best plan of action be in these circumstances?

In Example 3, the problem is similar to that of Example 2, except that it focuses on the options of solution and hence on the task to be decided. In this way, our aim in the instruction is to stimulate the use of correct judgements about probability in order for sound decisions to be made. However, the use of general decision-making procedures is also fostered, with a view to boosting the necessary use of strategies for planning how to tackle a problem. This meta-knowledge factor is essential in all problem-solving tasks, together with "rethinking" the whole process of solution.

> **Example 3.** Julia is studying the profitability of setting up a business, such as a gift shop. At the Chamber of Commerce she is given information about how many establishments of this type there are in her city and how they are doing. She is told that there aren't many of them and that according to the protocols used to estimate the profitability of such businesses they do have a success rate -of working profitably- of 60%. She is also told that the success of this kind of business can be improved to a considerable extent if the proprietors specialise in 10 products representative of the area. In these cases, the profitability of the shop will rise to 90%. Julia doesn't know whether setting up a business like this will allow her to get by because she must take into account the investment she will need to launch such an enterprise. At the agency, she is given further details. A shop of this kind will have expenses of around 600 euros. This does not include the opening costs, since the Regional Administration is prepared to cover 100% of these. Another aspect to be taken into account is the profit margin over a month. She is told that she can easily make 3000/month.
> How should Julia proceed to assess the profitability of this business venture?

In the ARDESOS progam, we also attempt to foster attitudinal aspects through interest and *motivation* by using tasks that can be found in daily situations and that involve topics relevant to most people, such as education, health, leisure, etc. In our research, we are attempting to clarify what is understood by motivation or disposition with a view to incorporating such a distinction, more or less directly, into instruction. An excellent stance regarding this issue is that of Valenzuela and Nieto (2008). In their work, four motivational aspects were selected that in our opinion seem to be the most relevant to instruction; namely, *attainment*, *utility*, *cost* and *interest*. In their study, two of these aspects have proved to be especially relevant to Critical Thinking: *utility* and *interest*. In our research, *interest* is gathered under the type of task and the topics addressed. *Utility* involves posing the issue of whether there is anything more important than critical reflection and showing its goodness with results. A lot remains to be done in this field, although at least an important step in the right direction has been taken in recent years: the awareness of investigators of critical thinking that we should not only attend to skills but that we should incorporate crucial dimensions such as the motivational, attitudinal or meta-knowledge dimensions.

We have described the main aspects of our intervention with the exception of one, which we have left for the end of this section. Since the start of our applied research some time ago we have observed that the teaching of Critical Thinking is *biased*: students are instructed in good reasoning but not in preventing poor reasoning. We shall thus spend a little time on this discrepancy aspect, which we believe to be a limitation. Some time ago, in 1988, Baron (2008) pointed out that in order to improve thinking processes three aspects must be tackled: the descriptive, the normative, and the prescriptive, but little attention has been paid to descriptive issues from the Critical Thinking approach.

It was precisely a psychologist (Henle, 1962) who performed some very interesting descriptive studies in which she pointed out how poorly we reason. Henle posed daily problems in which, as a general conclusion, she found that we scarcely use formal logic and that above all we use our *personal* logic. In other words, our beliefs, our way of understanding reality, mark the course of reasoning, without taking into account essential aspects such as the relationship between the different affirmations of an argument. What is most important about Henle's work is that it is the first descriptive investigation to address how we reflect, and hence to ascertain which systematic errors we commit. The pioneering aspect of this work is that it calls attention to the limitations of our judgements and how important it is to be aware of these deficiencies in order to correct them. From a normative point of view, it is assumed that the idea is to teach students how to think correctly, but not that such teaching is harder if the biases and deficiencies in our way of thinking are not known.

Some time ago, in 1985, Nickerson (2008) differentiated *reasoning* from *rationalizing*. In the idea of *rationalizing,* the author was referring to many of the fundamental biases or errors in reasoning that have been identified since the work of Henle. In our daily activities, when we check an idea or a hypothesis we normally only focus on the information or data that confirm it, but never on those that refute it. This *confirmatory bias*, for example, is one of the most important ones in what Nickerson refers to as *rationalization*. The problem with these distortions, or errors, is that they cannot be corrected or eliminated merely through the acquisition of correct reasoning skills. Nickerson suggests a powerful reason for this. There is a certain automatic nature or unconscious functioning in our way of thinking, as is the case of confirmatory thinking, such that, for example, it cannot be corrected through a mastery of the scientific method, since when this is applied we continue to pay no heed to non-confirmatory data and again fall into the trap. These errors can only be eliminated by our becoming aware of them; becoming familiar with this way of proceeding with a view to avoiding it. The same occurs with fallacies. These cannot be prevented merely by applying criteria of soundness; we must have some knowledge of them, because the language and the way in which such pseudo-arguments are expressed are so subtle that they are able to confuse us much more easily than we would wish. However, since the *errors or distortions* of our way of thinking cannot be avoided through good judgement *they must be incorporated into instruction*; i.e., they deserve separate treatment. As regards reasoning, as well as addressing the most common fallacies, we naturally look at the confirmatory bias (with all its implications) as well as the errors of illicit conversions with universal or conditional propositions. We also address the error of confusing truth with validity and the error of using inductive strategies in deductive contexts, to cite some of the biases taught in our program (see Evans, 2007; Govier, 2005; Saiz, 2002c; 2002d). In sum, what we wish to show is the relevance of such descriptive issues in interventions and the need to incorporate them, as we are attempting to do here.

Having discussed the limitations of our thinking, we complete the description of the ARDESOS program. We have focused on the main pillars of the program: *a lot of practice, inter-domain practice, daily situations, and biases.* Procedural activity is a constant in all instruction initiatives and there is nothing new in incorporating a lot of practice in any program of this type. However, what is new is that those activities stem from different contexts, that they are posed as real problems, attending to the limitations of our minds to address them. This is because as far as we are aware such an approach has not been used previously. The aim of the present work is to check whether an intervention of this type will be efficient; that is, whether it will produce a reasonable improvement in Critical Thinking. Our final aim is to check whether such progress will become generalized. Our efforts are directed towards allowing the improvement in skills to be expressed in any personal or professional context. Let it not be forgotten that the tasks used in our interventions are simulated, suitably represented daily situations. If performance on them is good, it should also be good in reality, or at least we can hope that this will be the case, just as a flight simulator exercise is expected to provide the same responses as in a real airplane.

If our goal in this research is to develop our ability for critical reflection, it is because this ability is not manifested as much as it should be. We have already stated that when intellectual capacity is tested the results are much poorer than would be hoped for or expected. This is undoubtedly an important problem that merits future investigation. To achieve our aims, we developed the program described above (which we will detail in all its phases in the section addressing methodology) and we believe that this initiative has some features that could make it reasonably efficient. This is therefore our proposal for solving the limitations of or optimizing people's ability to engage in thinking properly. In simple words, our *working hypothesis* is that the performance of the participants in the ARDESOS program will be better than that of those who are not enrolled in it, but who nevertheless have received a classical instruction in thinking (based on decontextualized exercises of induction and deduction). Nevertheless, this must be confirmed, and to do so we carried out the study described in the following section.

II. Methodology.

A. Participants.

Initially we started out with a convenience sample of 199 students (85% women) from the fourth year of the Psychology degree at the University of Salamanca. As a control group, 114 students (84% women) from the fourth year of the Psychology degree at the University of Málaga were used. For different reasons (lack of information, incomplete tests, etc), the experimental loss was 22% in the intervention group and 18% in the control group. As a result, the final sample comprised 155 cases in the experimental group and 94 individuals in the control group. The equivalence of both groups as regards sex and age was analyzed. In the intervention group 84% were women while in the control group the figure was 87%. This difference is not statistically significant. ($\chi^2_{(1)}$=0.291; p=0.590). The mean of the individuals participating in the intervention was 22.77 years (s.d 1.09), while the corresponding age in the control group was 22.93 years (s.d. 1.20). This difference was not statistically significant either ($t_{(247)}$=1.06; p=0.289). Both tests confirmed the equivalence between groups with sufficient reliability.

B. Assessment materials and measurements.

PENCRISAL: test for the assessment of Critical Thinking skills.

As a measure of the magnitude of the effect of the intervention, and with a view to determining whether the intervention had afforded an improvement in Critical Thinking skills, the PENCRISAL test, explained below, was applied. A more detailed description of the test can be found in Saiz and Rivas (2008b).

PENCRISAL is a test comprising 35 problem-situation items offered in an open-response format. The statements are designed in such a way that they do not demand that the response should be elaborated and expressed in technical terms. Quite the opposite; they can be answered perfectly well in colloquial language. These 35 items are configured around 5 factors: deductive reasoning, inductive and practical reasoning, decision-making and problem-solving. In the distribution of the problem situations, in each factor the choice of the most representative structures of each of them was taken into account. These factors thus represent the fundamental skills of thinking and in each of them the most relevant forms of reflection and resolution in our daily functioning can be found. When PENCRISAL was applied, the order of presentation of the items was random, although care was taken to ensure that several situations belonging to the same factor would not appear consecutively.

PENCRISAL can be administered in written form or using a computerized version through the Internet. Also, it can be applied individually or collectively. In our study we chose the computerized, collective application owing to the advantages this offers. It offers the most advantages to the corrector by facilitating the tedious inputting of data, and all so for the person taking the test, since the programming system allows the test to be taken in several sessions, thereby reducing the possible effects of tiredness that it may elicit, especially as regards performance on the last items. The system also allows all the relevant aspects of the test to be controlled, such as preventing any item from not being answered, because the system will not pass to the next item until an answer has been given to the previous one, and preventing the subject from correcting previous answers or taking the test again once it has been completed. The Internet version allows students to take the test from any place where an Internet connection is available, such as at home. The collective administration, however, is carried out in a classroom with several computers (in our case, three classrooms with twenty computers each). The latter allows control over each of the subjects to ensure they are performing the test without any help, something that cannot be controlled when the test is taken alone, without supervision. We do believe these advantages are enough to choose the collective computerized application over the other possibilities.

The correction criteria used were established on the basis of three standard values:
0 points: when the answer given as the solution is **incorrect**.
1 point: when the solution is **correct,** but insufficient argumentation is given (the student only identifies and demonstrates an understanding of the basic concepts).
2 points: when as well as getting the correct answer the individual **justifies or explains** why s/he has arrived at that conclusion (where more complex processes involving real mechanisms of production are used).

Thus, a system of quantitative scaling was used, whose range of values was between 0 and 70 points as the maximum limit for the global score on the tests, and between 0 and 14 for each of the five scales.

Regarding the time during which the test should last, our test can be defined as a psychometric power test (addressing capacity); that is, with no limitation on time. Nevertheless, the mean duration estimated for completing the test is between 60 and 90 min.

Psychometric study of this scale was performed with the 313 university students described above. Factor analysis was used for construct validation. The conditions for its use were fulfilled satisfactorily (KMO=0.605 and p=0.000 in the Bartlett test). The results revealed a set of factors and subfactors that accounted for 59.35% of the variance. Most of the items (28; i.e., 80%) correctly demonstrated (with saturations > 0.500) that they belonged to the expected theoretical factors: 8 to the deductive factor; 4 to the inductive one; 7 to practical reasoning; 5 to decision making, and 4 to problem solving. Regarding reliability, this set of items attained an acceptable Cronbach alpha value (0.737; p<0.05). In general, the scale can be said to demonstrate its factor validity, and its reliability is satisfactory. Nevertheless, as a consequence of these observations, 7 items (20%) were modified or replaced by others and currently the second version of the test is in the validation phase.

C. Intervention program.

The aim of our investigation was to optimize the intellectual skills involved in Critical Thinking established above (*reasoning, problem solving,* and *decision making*).

Owing to the complexity of the skills addressed, the problem is only suitable for adult populations with at least an average intellectual level. Our work was carried out with university students since it was a convenient and available population.

Our intervention is designed for classroom application over 20-30 hours, distributed in 15-20 ninety-minute weekly sessions and a maximum time of 60 hours including the students' own work (see Appendix 1).

The name we used to designate this intervention is the "ARDESOS program for the development of Critical Thinking." This term covers the three large skills conforming our program -**AR**gumentation, **DE**cision and **So**lution- together with one of the main features of our intervention: the use of daily **S**ituations for the development of those skills.

The ARDESOS program is based on the direct teaching of thinking skills, since this type of instruction allows the transfer of knowledge; that is, teaching the skills that we wish to be mastered directly should allow them to be applied to any other context.

These skills are essentially procedural knowledge, and hence our intervention focuses more on process learning than on content learning. Contents are evidently necessary for all types of learning but these are rigid and static, while processes are flexible and allow us to create alternatives since each person can generate different ways to access the same information. These ways are transferable and, once acquired, they can be applied to any field of knowledge.

The teaching-learning strategy on which our intervention program is based is Problem-Based Learning (PBL). Activity revolves around the discussion of different problem situations designed in the program, and the learning of the skills of Critical Thinking arises from the experience of having worked with such situations. It is a method that stimulates metacognitive processes and allows students to practise by challenging them with real situations, where they must seek and investigate their own answers and solutions.

The ARDESOS program focuses on the teaching of skills that we consider to be essential for the development of Critical Thinking, and hence for good practices in people's daily activities. To do so, it is necessary to use reasoning and good strategies for solving problems and making decisions. As explained in the Introduction, these three skills are the basis of our intervention. Nevertheless, it should be noted that the intervention involves not only instruction in the skills used daily but also correction of the biases and errors committed when they are used.

The main procedures used in the different activities of the program are reflection and discussion, active participation by students, and training in the different skills of Critical Thinking.

The tasks used in the program are a simulation of daily situations in which problems are posed that must be solved with the skills of reasoning, problem-solving and decision-making. These problem situations allow the differences between the learning contexts and daily life to be minimized.

Our program was applied to reduced groups of students (not as reduced as we would have wished, owing to our student numbers) of 15-20 persons. We consider that the ideal number of participants would be 10-12, but this is not always possible to achieve. The length of the program is approximately 60 hours, which are distributed as follows: fifteen 90-minute sessions with 15-20 students (23 hours), ten 90-minute lectures with 50 students (15 hours) and seven 1-hour tutorials with 3-5 students groups. The remaining 15 hours are devoted to the solving of daily problems, carried out in the students' own time.

The procedure is as follows. The instructor begins a process of direct teaching of each skill, applying it in a practical way to specific examples. The emphasis of the teaching of each skill is placed on in the structural aspects of the different arguments, such that study of each of them does not depend on the content but on the structure. One aspect meriting attention is that the students must solve a series of problems before each of the sessions. This allows more time for the sessions and, additionally, it allows the students to become aware of the difficulties and to understand why they can solve some problems but not others. This in turn makes them aware of their own limitations so that in the practical classes they can explore them further. Moreover, since the students must attempt to solve the problems before the sessions they can compare the process they have followed with that of other students and that offered by the instructor. In this way, on one hand we are fostering meta-knowledge and, on the other, we are increasing practical activities.

In each session, the aim is for the students to tackle the problem situations actively. Performance is subject to continuous assessment with a view to stimulating the students to complete the activities before the sessions, which is crucial for the success of the program owing to the few hours available for direct contact. In this sense, all participants later received a detailed analysis and assessment of their work. Additionally, the evaluation of student performance was completed with classroom discussion by the instructor of all the difficulties and doubts that had emerged and a clarification of such problems. As stated earlier, we wish the students to become aware of their own thought processes in order to improve them.

The sessions revolved homogeneously around blocks of skills. Within the field of reasoning, argumentation was the main issue. In order to find intellectual tasks that could be applied in daily situations, we used a general model of argumentation, such as that of Toulmin (2003), which is followed by most authors (see, among others, Fisher, 2002; Govier, 2005; Johnson & Blair, 2006; Walton, 2006). Our contribution as regards the model of argumentation was to include all the forms of reasoning we were going to use in teaching it. The proposal of most authors is to separate argumentation (informal reasoning) from other forms of reasoning. We believe that this separation is not valid in daily life. When people defend a given stance or position, they argue making use of all the inference resources that they are able to, even though they are not aware of most of them. If we were to analyze argumentative texts produced by a person, different forms of reasoning would become apparent. The question that in due course emerged was that if in our daily use of reasoning we do not separate certain structures from others, since all of them are integrated in an argumentative text or discourse, why do this in instruction? Thus, we have developed a global focus about reasoning that has proved to be more efficient than studying the different types of argumentation separately. By using an integrated model, we facilitate the understanding and use of the different reasoning structures in any circumstance or situation. This allows us to achieve a better degree of skill in the domain of argumentation. The efforts to integrate these skills were also applied to decision-making and problem-solving. Here, within a general mechanism of problem solving we related and integrated the different decision strategies and the search for solutions. A large part of the materials used can be found at the following internet address:

http://www.pensamiento-critico.com/pensacono/prograpensa.htm#mat didac

D. Design.

In order to analyze the efficiency of the intervention, a quasi-experimental design was made of two groups with pre- and post-treatment measurements. The intervention (O_1 X O_2) and control (O_1-O_2) groups were formed and from these we first took a pre-treatment measurement. Then, after the program had been applied in the intervention group, we performed the post-treatment measurements.

E. Procedure.

Application of the ARDESOS program was carried out along one semester at the School of Psychology of the University of Salamanca. One week before the instruction we applied the PENCRISAL test to all the students (control and intervention groups) and one week after the end of the instruction the second measurement with PENCRISAL was implemented. The time elapsed between the pre- and post-treatment measurements was 4 months for both groups. The intervention was performed by a single instructor with good experience and training in the program.

F. Analysis of results.

To analyze the effect of the intervention, Student's *t* tests for independent samples with repeated measurements were implemented to check whether there were significant differences between the groups in the pre- and post- situations. Data treatment was accomplished using the SPPS package (v. 15.0).

III. Results.

As mentioned in the description of the PENCRISAL test, Critical Thinking was measured on the basis of five factors- Deduction, Induction, Practical Reasoning, decision-making (DM) and problem-solving (PS), and an overall score. Accordingly, the analysis was carried out attending to the performance observed on each of these 6 variables

First, we describe the results obtained in the *pre-post measurements in the control group*. As can be seen in table 1, no statistically significant differences were observed in four of the five factors of the scale: *deduction* ($t_{(79\ )}$=0.88; p=0.384), *induction* ($t_{(84)}$=0.00; p=1), *practical reasoning* ($t_{(81)}$=0.326; p=0.746) *and problem-solving* ($t_{(80)}$=0,00; p=1). Neither were there statistically significant differences in the overall scores of the test ($t_{(79)}$=1.25; p=0.218). Significant differences were only found for the *decision-making* factor ($t_{(\ 81)}$=3.43; p=0.001), with a mean of 5.73 on the pre-test and of 4.73 on the post-test measurement, from which a decrease in performance over time can be deduced. These data indicate that in general terms the group not receiving the treatment did not alter their skills during the four-month period between both measurements.

Regarding the *intervention* group, evidently it was expected that the *pre-post* measures would differ significantly. In table 1 it can also be seen that in the intervention group statistically significant differences were only observed for three factors. In *induction* ($t_{(92)}$=3.84; p=0.000), mean performance was higher at post-test (M=4.69) than at pre-test (M=3.74); in *decision-making* ($t_{(86)}$=2.08; p=0.040), an increase in performance also occurred after the intervention (M_{pre}=6.08; M_{post}=6.64). However, the significance reached on the *deduction* factor ($t_{(89)}$=3.83; p=0.000) was in this case the opposite of what was expected (M_{pre}=6.31; M_{post}=5.21), indicating that the students' performance on this skill was worse after the intervention. No significant differences were seen for the practical reasoning factor ($t_{(92)}$=0.332; p=0.741) or problem-solving factor ($t_{(92)}$=1.51; p=0.135). Regarding the total PENSCRISAL score, no significant differences were observed either between the pre- and post-treatment measurements ($t_{(86)}$=0.76; p=0.448). Taken together, these data suggest that the intervention group improved on some of the factors after the program had been applied.

In table 2, we describe the *pre-test measurements obtained in both groups* to see whether both groups were similar in their initial state as regards the PENCRISAL variables. In particular, the data show that the groups did not differ significantly in the following factors: *deduction* ($t_{(229)}$=1.69; p=0.092), *induction* ($t_{(231)}$=1.90; p=0.058), *decision-making* ($t_{(236)}$=1.42; p=0.156) and *problem-solving* ($t_{(236)}$=0.96; p=0.337). In contrast, statistically significant differences were seen in *practical reasoning* skills ($t_{(230)}$=6.29; p=0.000) between both groups, the intervention group obtaining better scores (M=6.47) than the controls (M=4.24). This could account for the significant differences also seen in the total mean of PENCRISAL ($t_{(226)}$=2.67; p=0.008), where the intervention groups maintained a higher score (M_{INT}= 26.36; M_{CONT}=23.81).

Finally, we analyzed the size of the effect observed in the PENCRISAL score after the intervention program. To accomplish this, we compared *both groups* as regards their *post-test* scores. Statistically significant differences were observed in the total score ($t_{(177)}$=2.71; p=0.008), with a higher performance mean in the intervention groups than in the control (M=26.63 and M=23.70, respectively), and also in three of the factors of the scale (see 2). Specifically, performance on *practical reasoning* was significantly better ($t_{(183)}$=5.02; p=0.000) in the intervention group (M=6.62) than in the control group (M=4.52); and the *decision-making* skill also underwent a significant improvement ($t_{(178)}$=7.27; p=0.000) in the intervention group (M=6.58) with respect to the controls (M=4.62). Nonetheless, the results concerning deduction show that the control group (M=6.03) was the one whose performance regarding this skill improved ($t_{(184)}$=2.25; p=0.026) with respect to the group that received the instruction (M=5.29). Finally, no significant changes were observed in the other two factors of the test: *induction* ($t_{(192)}$=21.35; p=0.179) and *problem-solving* ($t_{(186)}$=1.81; p=0.072). These data indicate the significant improvement due to the intervention in most of the factors with respect to the control group after application of the program.

Table 1. Means, standard deviations, and significance of the PENCRISAL measurements.

Comparison between pre-post-test measurements

	INTERVETION (n=155)			**CONTROL** (n=94)		
	PRE	POST	Difference	PRE	POST	Difference
	Mean (d.t.)	Mean (d.t.)	Dif. Between means p-sig n valid	Mean (d.t.)	Mean (d.t.)	Dif. Between means p-sig n valid
DED	6.31 (2.47)	5.21 (2.21)	1.10 ** **0.000** 97	5.97 (2.52)	6,32 (2.22)	-0,35 0.384 61
IND	3.74 (1.59)	4.69 2.20)	-0.95 ** **0.000** 99	4.49 (1.53)	4.49 (1.66)	0.00 1.00 72
RP	6.37 (2.69)	6.47 (2.74)	-0.10 0.741 97	4.42 (2.27)	4.53 (3.01)	-0.11 0.746 65
TD	6.08 (1.74)	6.64 (2.04)	-0.56 * **0.040** 88	5.73 (1.90)	4.73 (1.67)	1.00 ** **0.001** 63
SP	3.75 (1.32)	3.53 (1.23)	0.22 0.135 94	4.06 (1.21)	4.06 (1.39)	0.00 1.00 74
TOT	25.98 (6.27)	26.65 (7.35)	-0.67 0.448 70	24.69 (5.91)	23.36 (5.95)	1.33 0.218 55

* Significant at 5%   ** Significant at 1%

IV. Discussion and implications for future research.

Overall, it can be said that the results obtained with our ARDESOS program indicate efficiency in some of the factors, as seen from the significant changes in the right direction. However, it seems appropriate to spend some time exploring these results further. One very important observation is that the control group obtained the same scores at pre- and post test. Had this not been the case, we would be unable to say anything about the improvements obtained with the intervention. However, with this equality we can be reasonably sure that the changes achieved in the intervention group at post-test must have been due to application of our program. Overall performance was higher at post-test in the intervention group, which is what was expected. In sum, we seem to have achieved the ideal situation with this type of design: no differences in the control group and differences in the intervention group as regards their performance at pre- and post test, the latter values being higher. Nonetheless, we failed to achieve an improvement in all the skills taught. An improvement was observed in induction and decision-making, but not in deduction. We have no clear explanation for this, although the following could be advanced. In this study, we used the first version of PENCRISAL, in which we later detected certain deficiencies in the items; these have now been corrected. One of them could have been responsible for the anomaly. The level of difficulty of the test was high as regards situations of deduction. On working with the different types of reasoning with an integrated text, it is possible that -indirectly- more emphasis was being placed on seeking the elements of an argument, such as reasons and conclusions, than on formal structures. After the intervention, this -together with the difficulty of those items, could have led to a bias towards only argumentative forms (practical reasoning), sidestepping deductive forms too much. However, what we can explain is the improvement (although not significant) in deduction in the control group. This group received several hours of practice in deduction and a few practical sessions dealing with decision-making and induction. These activities clearly account for the improvement.

Another unexpected finding, which again we can account for, is the absence of before-after differences in practical reasoning. Application of the pre-post measurements was performed when the practical work in this area had already started, such that the gain on this factor was abolished by this lack of control. This is very patent in the measurements of the two groups. The intervention group started out from just over six (6.37) and the control groups from slightly more

Table 2. Means, standard deviations, and statistical significance of the PENCRISAL means.

Comparison between groups

	PRE-MEASUREMENT			**POST-MEASUREMENT**		
	Intervention (n=155)	Control (n=94)	Difference	Intervetion (n=155)	Control (n=94)	Difference
	Mean (d.t.) n valid	Mean (d.t.) n valid	Dif. Between means p-sig g.l.	Mean (d.t.) n valid	Mean (d.t.) n valid	Dif. Between means p-sig g.l.
DED	6,12 (2,41) 150	5,54 (2,56) 81	0,58 0,092 229	5,29 (2,30) 98	6,03 (2,22) 88	-0,74 ***** **0,026** 184
IND	3,88 (1,55) 149	4,29 (1,58) 84	-0,41 0,058 231	4,72 (2,27) 101	4,33 (1,68) 93	0,39 0,179 192
RP	6,47 (2,66) 147	4,24 (2,46) 85	2,23 ****** **0,000** 230	6,62 (2,77) 99	4,52 (2,88) 86	2,10 ****** **0,000** 183
TD	6,00 (1,88) 154	5,64 (1,79) 84	0,36 0,156 236	6,58 (2,00) 93	4,62 (1,57) 87	1,96 ****** **0,000** 178
SP	3,77 (1,28) 155	3,94 (1,25) 84	-0,17 0,337 237	3,50 (1,25) 97	3,85 (1,40) 91	-0,35 0,072 186
TOT	26,36 (6,45) 124	23,81 (6,05) 67	2,55 ****** **0,008** 189	26,63 (7,64) 82	23,70 (5,93) 80	2,93 ****** **0,008** 160

* Significant at 5%   ** Significant at 1%

than 4 (4.42). In the post- measurement, for the former we observed that this level persisted (6.47), as in the second case (4.53). However, it should be noted that that difference of two points between both groups is one third of the performance. If the intervention group had started out from four points, the difference would have been significant. Proof of this is that the mean between groups on the post- measurement was significant.

Neither did the students' performance on problem solving improve after the intervention. This would probably be due to the following reasons. Some problem-solving and decision-making items are general, and to be solved they demand procedures involving overall planning of the answer. It is possible that some interference might have arisen between both types of situation, preventing a treatment and differential solution for each of them. Finally, we failed to find significant differences between the groups on the post- measurements for induction. We believe that this can be explained in terms of the level of difficulty of those items, which produced the classic *floor effect*.

In our Critical Thinking evaluation test, we have detected a few limitations that need to be corrected. The first is its high level of difficulty. This characteristic might have prevented the detection of significant additional effects of the intervention. The difference in the number of items between some dimensions poses a second problem, and may affect the reliability of the test. These limitations, besides certain other minor problems, have been overcome in the current version of the test.

Globally, our program represents a very ambitious bet regarding the objectives it attempts to achieve. Such an instruction program requires a careful conceptual development and evolves along time as it achieves positive results. We are convinced that our intervention will provide these good results, but the path is still long. This work is the first to test the initiative and, as such, has yielded modest results; we are aware that these must be improved. We have indeed learnt a lot from what we have not achieved and we are currently putting our experience into practice and introducing modifications to the program. Our hope is to achieve a better efficiency in changing the skills of Critical Thinking, and we believe we are moving in the right direction.

References

Baron, J.B. (2005). *Rationality and intelligence* (2nd ed.). Cambridge. NY: Cambridge University Press.

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

Bensley, D. A. Crowe. D. S. Bernhanrdt. P. Buckner. C. & Allman. A. L. (2010). Teaching and assessing critical thinking skills for argument analysis in psychology. *Teaching of Psychology, 37* (2), 91-96.

Blair, J.A. (2009). The pertinence of Toulmin and Perelman/Olbrechts-Tyteca for informal logic. In H. Jales (Ed.). *Retoric and argumentation in the beginning of the XXIst century* (pp. 17-32). Coimbra. Portugal: Coimbra University Press.

Bransford, J.D., & Stein, B.S.J. (1993). *The IDEAL problem solver: A guide for improving thinking, learning, and creativity* (2nd ed.). San Francisco.CA: Freeman.

Ennis, R. H. (1987). A taxonomy of critical thinking dispositions and abilities. In J.B. Baron & R.J. Sternberg (Eds.). *Teaching thinking skills* (pp. 9-26). New York: Freeman.

Ennis, R. H. (1996). *Critical thinking.* Upper Saddle River. NJ: Prentice-Hall.

Evans, J.St.B.T. (2007). *Hypothetical thinking. Dual processes in reasoning and judgment*. Hove. UK: Psychology Press.

Fisher, R. (2003). *Teaching thinking: Philosophical enquiry in the classroom* (2nd ed.). London: Continuum

Govier, T. (2005). *A Practical Study of Argument* (6th ed.). Belmont. CA: Wadsworth.

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). *Thought and knowledge: An introduction to critical thinking* (4th ed.). Hillsdale, NJ: Erlbaum.

Henle, M. (1962). On the relation between logic and thinking. *Psychological Review*, 69, 366-378.

Johnson, R. H. (2000). *Manifest rationality: A pragmatic theory of argument.* Mahwah, NJ: Erlbaum.

Johnson, R. H. (2008). *Critical Thinking, Logic, and Argumentation.* Conferencia magistral. presentada en la Conferencia internacional: Lógica. Argumentación y Pensamiento Crítico. (Santiago de Chile. 8-11 de enero de 2008). Universidad Diego Portales. Santiago de Chile.

Johnson, R. H., & Blair. A. (2006). *Logical self-defensa.* New York: International Debate Education Association

Lipman, M. (2003). *Thinking in education* (2nd Ed.). Cambridge. MA: Cambridge University Press.

McPeck, J.E. (1981). *Critical thinking and education.* Oxford. UK: Martin Robinson.

McPeck, J.E. (1990). *Critical thinking: Dialogue and dialectic.* New York: Routledge.

Mercier, H., & Sperber, D. (2010). Why do humans reason? Arguments for an argumentative theory. *Behavioral and Brain Sciences,* 1-63.

Nickerson, R.S. (2008). *Aspects of rationality. Reflections on what it means to be rational and whether we are*. New York: Psychology Press.

Nickerson, R.S., Perkins, D. N., & Smith, E. E. (1985). *The teaching of the thinking*. Hillsdale, NJ: Erlbaum.

Nieto, A.M., & Saiz. C. (2008). Evaluation of Halpern's "Structural Component" for Improving Critical Thinking. *The Spanish Journal of Psychology*, *11* (1), 266-274. http://www.pensamiento-critico.com/pensacono/SJPS08.pdf

Saiz, C. (2002a). Enseñar o aprender a pensar. *Escritos de Psicología*, *6*, 53-72. http://www.pensamiento-critico.com/pensacono/escritosps.pdf

Saiz, C. (2002b). Razonamiento práctico: método de análisis. En C. Saiz (Ed.). *Pensamiento crítico: conceptos básicos y actividades prácticas* (p. 21-44). Madrid: Pirámide.

Saiz, C. (2002c). Razonamiento categórico. En C. Saiz (Ed.). *Pensamiento crítico: conceptos básicos y actividades prácticas* (p. 45-81). Madrid: Pirámide.

Saiz, C. (2002d). Razonamiento inductivo. En C. Saiz (Ed.). *Pensamiento crítico: conceptos básicos y actividades prácticas* (p. 126-182). Madrid: Pirámide.

Saiz, C. (2009). *Pensamiento crítico: ¿una teoría del razonamiento o de la acción? (Critical thinking: a theory of reasoning or a theory of action?* Salamanca: Universidad de Salamanca. http://www.pensamiento-critico.com/pensacono/index.htm
Registro Provisional de la Propiedad Intelectual nº: SA/19/10 -versión 2-. (intellectual property registration).

Saiz, C. & Nieto. A. M. (2002). Pensamiento crítico: capacidades y desarrollo. En C. Saiz (Ed.). *Pensamiento crítico: conceptos básicos y actividades prácticas* (p. 15-19). Madrid: Pirámide.

Saiz, C., & Rivas. S.F. (2008a). Intervenir para transferir en pensamiento crítico. *Praxis, 10* (13), 129-149. http://www.pensamiento-critico.com/pensacono/intertranspcpraxis.pdf

Saiz, C., & Rivas. S.F. (2008b). Evaluación en pensamiento crítico: una propuesta para diferenciar formas de pensar. *Ergo Nueva Época, 22/23, 25-66.*
http://www.pensamiento-critico.com/pensacono/evaluarpcergodf.pdf
English version: Assessment in critical thinking: a proposal for differentiating ways of thinking:http://www.pensamiento-critico.com/pensacono/evaluationCTergoENGLSH.pdf

Toulmin, S.E. (2003). *The uses of argument* (update edition). New York: Cambridge University Press.

Valenzuela, J., & Nieto. A.M. (2008). Motivación y Pensamiento Crítico: Aportes para el estudio de esta relación. *Revista Electrónica de Motivación y Emoción, XI*, 28. http://reme.uji.es/

Walton, D. (2006). *Fundamentals of critical argumentation.* New York: Cambridge Universitiy Press.