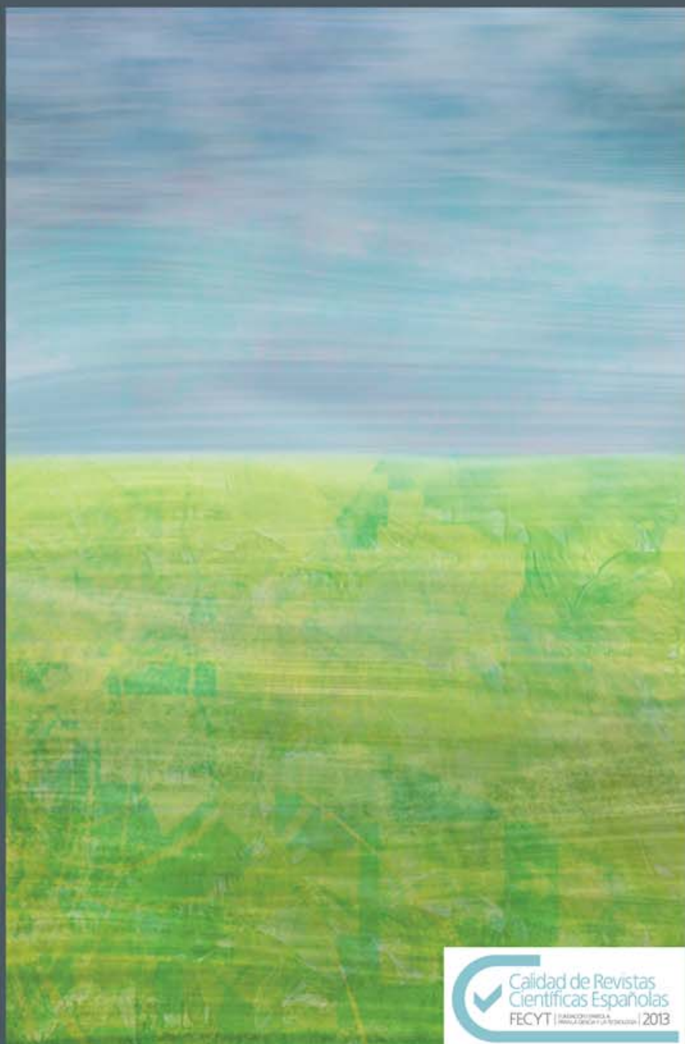
revista de EDUCACIÓN
Nº 371 ENERO-MARZO 2016

Razonamiento y heurísticas en pruebas de comprensión lectora

Reasoning and heuristics in reading comprehension tests

Alejandra Platas-García
José M. Castro-Manzano
Verónica Reyes-Meza



GOBIERNO DE ESPAÑA
MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

Calidad de Revistas Científicas Españolas
FECYT 2013

Reasoning and heuristics in reading comprehension tests

Razonamiento y heurísticas en pruebas de comprensión lectora

DOI: 10.4438/1988-592X-RE-2015-371-312

Alejandra Platas-García
Benemérita Universidad Autónoma de Puebla
José Martín Castro-Manzano
Verónica Reyes-Meza
UPAEP

Abstract

One of the problems that foreign language teachers face is recognizing the student's degree of inferential reading comprehension. A possible way to approach this problem is through the design of comprehension tests and the analysis of heuristics. In this paper we suggest that it is possible to find a homomorphism between an inferential model and the abstract structure of the multiple-choice tests that require inferential comprehension. This homomorphism may be analyzed from two standpoints: design (top-down) and resolution (bottom-up). From the former it is possible to design multiple-choice tests by using a taxonomy of reasoning (deductive, inductive, and abductive); from the latter it is possible to record the used heuristics in the resolution of such tests (coincidence, elimination, association, randomness, and relevance). Since these points of view can be studied experimentally, we have performed a series of observations by applying a test on 32 students of Italian as a foreign language from different academic departments. The results we obtained showed statistical difference between the number of right answers and the language (p=0,0002), and the type of reasoning both in Italian (p=0,0068) and Spanish (p=0,0005); but they did not show differences with respect to academic programs in deductive (p=0,6178), inductive (p=0,1702), and abductive (p=0,9412) reasoning; nor with respect to the length of the body of the text in Italian (p=0,6121) or Spanish (p=0,9194). With respect to the employed heuristics we found differences in the

Received: 13-08-2015 Accepted: 13-11-2015

frequency of use with respect to the type of reasoning (p=0,0001). Given these results we suggest the importance of identifying the types of reasoning and the heuristics used in the resolution of inferential reading comprehension tests in foreign language with emphasis in the relevance heuristic, since it contributes to the evaluation of reading comprehension at an inferential level.

Keywords: foreign language, reading comprehension, deduction, induction, abduction, relevance.

Resumen

Uno de los problemas con los que se enfrentan los docentes de lenguas extranjeras consiste en reconocer el grado de comprensión lectora inferencial de los estudiantes. Una posible forma de abordar este problema es a través del diseño de pruebas de comprensión y el análisis de heurísticas. En este artículo sugerimos que es posible encontrar un homomorfismo entre el modelo inferencial y la estructura abstracta de las pruebas de opción múltiple en las que se solicita comprensión inferencial. Este homomorfismo puede analizarse desde dos puntos de vista: el del diseño (top-down) y el de la resolución (bottom-up). Desde el primero es posible diseñar pruebas de opción múltiple usando una taxonomía de razonamientos (deductivo, inductivo y abductivo); desde el segundo es posible registrar las heurísticas empleadas en la resolución de estas pruebas (coincidencia, eliminación, asociación, azar y relevancia). Como estos puntos de vista pueden estudiarse experimentalmente, hemos llevado a cabo una serie de mediciones mediante la aplicación de una prueba a 32 estudiantes de italiano como lengua extranjera provenientes de diferentes áreas académicas. Los resultados que obtuvimos mostraron diferencias significativas entre el número de respuestas correctas, el idioma (p=0,0002) y el tipo de razonamiento tanto en italiano (p=0,0068) como en español (p=0,0005); pero no mostraron diferencias con el área académica en razonamientos deductivos (p=0,6178), inductivos (p=0,1702) y abductivos (p=0,9412); tampoco con la longitud del cuerpo del texto en italiano (p=0,6121) o en español (p=0,9194). Con respecto a las heurísticas empleadas encontramos diferencias significativas en la frecuencia de uso por tipo de razonamiento (p=0,0001). A partir de estos resultados sugerimos la importancia de identificar los tipos de razonamiento y las heurísticas en la resolución de pruebas de comprensión lectora inferencial en lengua extranjera con énfasis en la heurística de relevancia, porque posibilita una evaluación de la comprensión lectora en un nivel inferencial.

Palabras clave: lengua extranjera, comprensión lectora, deducción, inducción, abducción, relevancia.

Introduction

Let us consider a *gedankenexperiment*. Suppose that four students take a multiple-choice test as part of a reading comprehension evaluation. And suppose, further, that such test is valid (instructions are clear, all possible answers are justified), reliable (for each question one and only one possible answer is the right one), and viable (conditions for its correct application exist). Finally, suppose that, after finishing the test, all four students obtain the highest grade.

Does this mean that the four students *comprehended* the text? Is it possible for a student to answer correctly without gaining reading *comprehension*? Consider the next alternatives. Say the first student chooses her answers based upon her previous knowledge (coincidence); the second one selects her answers by identifying similar expressions occurring in both the text and the possible answers (association); the third one picks her answers after discarding some choices (elimination); and the fourth chooses her answers randomly, say, because she had no enough time to finish the test (randomness).

What we want to point out with this thought experiment is that, indeed, it is possible for a student to answer a reading comprehension test correctly without having reading comprehension. The heuristics of coincidence, association, elimination, and randomness are useful strategies for multiple-choice test resolution, but they do not necessarily reflect reading comprehension at an inferential level. The motivation behind this research comes from our teaching experience regarding this issue with students of Italian as a foreign language.

Usually, the evaluation of reading comprehension is accomplished through different exams from which we can highlight multiple-choice tests due to the next features: validity (they measure what they pretend to measure), reliability (the provide consistency in the results), and viability (they are easy to apply) (Palencia del Burgo, 1990, p.225; Bachman 1990, p.25). For further explanation on the notions of validity see chapters VIII in Alderson, Clapham & Wall (1995) and IV in Hughes (2003). However, given that reading comprehension cannot be directly observed (inferential comprehension particularly) and we have to ask the students to achieve some goals in order to indicate a level of comprehension (Pérez Zorrilla, 2005, p.128), it follows that producing the above features in an inferential reading comprehension test is not an evident task.

Revista de Educación, 371. January-April 2016, pp. 000-000
Received: 13-08-2015 Accepted: 13-11-2015

In this contribution we study types of reasoning and heuristics with the purpose of developing strategies to produce inferential reading comprehension tests that guarantee, as much as possible, effective evaluation of inferential reading comprehension skills.

We have organized this paper in four sections. In the first one we present our proposal by giving some details about the types of reasoning and the heuristics that we have used in this research. In the second one we describe the methodology we employed to obtain the results that we display on the third section. Finally, in the fourth section, we show the conclusions of this study.

Reasoning and heuristics

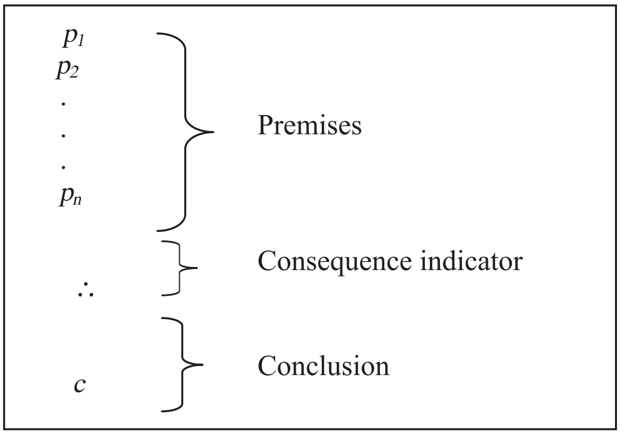
Previous researchç has approached the problem we described above by designing multiple-choice reading comprehension tests with inferential questions for undergraduate students (Velásquez, Cornejo & Roco, 2008, p.124). Following this methodology we studied the types of reasoning and the heuristics involved in these kinds of tests.

When we talk about inferential reading comprehension we appeal to the next hierarchy (Elosúa & García, 1993):

- *Decoding*: to decipher a code, give meaning to printed letters, associate written words with meanings available in memory, transform printed letters into syllables and sound in order to activate a meaning.
- *Literal comprehension*: to combine the meanings of several words in order to produce propositions given the explicit information of a text.
- *Inferential comprehension*: to produce more integrated and schematic mental representations involving the explicit information of the text and the previous knowledge of the agent in order to go beyond the information of the text.
- *Metacomprehension*: to establish reading goals, and verify if they a reached and rectify opportunely; it requires self-control over the process of comprehension.

Hence, when a reader has inferential comprehension is because, assuming the previous levels, it is able to perform inferences. An *inference* is a process in which a reasoner uses a finite set of data (usually represented by premises, $p_1, \ldots, p_n$) to reach new data (known as conclusion, *c*) (Cook, 2009, p.151). A visual representation of this inferential model is the next one:

FIGURE I. Visual representation of the inferential model



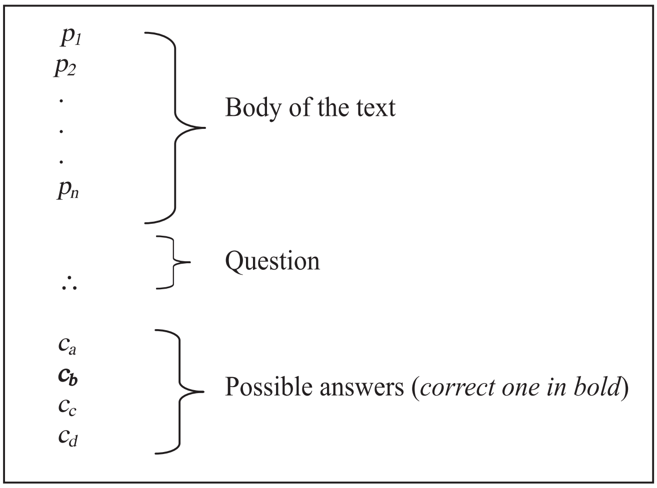
Source: Personal elaboration

Assuming this description, our first proposal consists in defining a homomorphism between the abstract structure of multiple-choice tests that demand inferential comprehension and the inferential model described above.

We define this homomorphism in the next way: the *premises* of the inferential model are represented by the *body of the text* in the reading comprehension test, the *consequence indicator* is represented by the *question*, and the *conclusion* corresponds to the *correct answer* within a set of possible answers. A visual representation of this homomorphism is the next one:

Received: 13-08-2015 Accepted: 13-11-2015

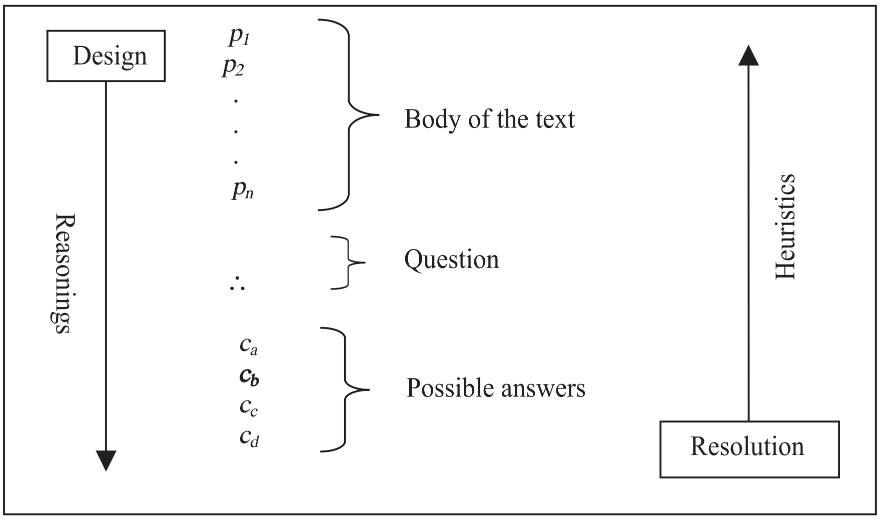
FIGURE II. Visual representation of the homomorphism



Source: Personal elaboration

The analysis of this homomorphism may be accomplished with two strategies: *top-down* and *bottom-up*. The first one would consist in the study of such structure from the point of view of the *design* of the reading comprehension test; the second one would consist in the revision of such structure from the *resolution* perspective. With respect to the first strategy we expound the *types of reasoning*; with respect to the second strategy we show a taxonomy of *heuristics* employed in the resolution of such tests (Figure III).

FIGURE III. Analysis of the homomorphism



Source: Personal elaboration

Types of reasoning

Reasoning is an inferential process that may be divided into three categories: deduction, induction, and abduction.

Deductive reasoning

Informally, in a deduction the result of an inference is necessarily true if the truth of the premises warrants the truth of the conclusion (Douven, 2011). More precisely, a deduction is defined as a finite sequence of statements within a formal system in which every statement in the sequence is an axiom, an assumption, or the result of applying an inference rule to one or several previous statements. The final statement of such sequence, the argument's conclusion, is derived and every assumption is a premise of the derived argument (Cook, 2009, p.88).

Received: 13-08-2015 Accepted: 13-11-2015

Deduction supposes, therefore, that information is complete and that, hence, the conclusion does not add new information to what is already said in the premises (Rodríguez Rodríguez, 2005, p.90). Thus, in deductive reasoning it is impossible for the premises to be true and the conclusion to be false (Cook, 2009, p.81).

Deductive reasoning may be valid or invalid depending on the compliance or violation of certain structural norms that we do not explain in this work due to reasons of space, but that the interested reader may find in classical works like (Copi, 1979) and (Mates, 1972), or more recent ones like (Enderton, 2001) and (van Dalen, 2004).

A typical example of a valid deductive reasoning would be the next one:

p_1 All men are mortal.
p_2 Socrates is a man. (1)
c Therefore, Socrates is mortal.

In (1) the truth of the conclusion is necessarily inferred from the truth of the premises because the form of such reasoning is correct (i.e., it follows the rules of classical logic). To clarify this distinction let us see an example of an invalid deductive reasoning:

p_1 Some cats are mammals.
p_2 Some mammals are felines. (2)
c Therefore, some cats are felines.

If we let *C* stand for *cats*, *M* for *mammals*, and *F* for *felines*, we can obtain the next representation of (2):

p_1 Some *C* are *M*.
p_2 Some *M* are F. (2')
c Therefore, some *C* are *F*.

Now suppose (2) is valid. If this is the case, then we can substitute the elements of (2') and preserve the truth of the conclusion. When we let *C* stand for *even numbers*, *M* for *natural numbers*, and *F* for *odd numbers* we obtain:

p_1 Some even numbers are natural numbers.
p_2 Some natural numbers are odd numbers. (2")
c Therefore, some even numbers are odd numbers.

Hence, since the substitution in (2') produces a reasoning with true premises and false conclusion we can conclude that (2) is an instance of an invalid reasoning. This invalidity is due to the formal structure of this deductive reasoning.

Inductive reasoning

The most common concept of induction says that it is a reasoning process that goes from the particular to the general. This concept, however, is quite restricted, and so, a much wider concept has been developed, one that regards induction as an inference that allows us to extract conclusions with certain degree of support (Hawthorne, 2014). The *degree of support* that the premises provide offers the conceptual base that sustains the conclusion, but this degree of support is not deductive since it does not involve necessity, but probability. The interested reader may find more attributes of this type of reasoning in traditional studies like (Cohen & Nagel, 1934), specially chapters VIII, XVI, and XVII; or in more advanced works like (Pearl, 2000).

In contrast with deductive reasoning, inductive reasoning may start with true premises and reach false conclusions (Cook, 2009, p.150). Consider an example of a correct inductive reasoning, adapted from Okasha (2002, p.19):

p_1 The first five eggs in this box are rotten.
p_2 All the eggs in this box have the same expiration date. (3)
c Therefore, it is very likely that the sixth egg will be rotten too.

(3) may start from true premises and, nevertheless, it may reach a false conclusion; however, although what is stated in the premises does not necessarily guarantee the truth of the conclusion, it is more likely that it is true, for the degree of support to infer such conclusion is higher than the degree of support to infer its negation.

Received: 13-08-2015 Accepted: 13-11-2015

An example of an incorrect inductive reasoning would be the next one:

p_1 Socrates was a philosopher and he was Greek.
p_2 Plato was a philosopher and he was Greek. (4)
p_3 Aristotle was a philosopher and he was Greek.
c Therefore, all philosophers are Greek.

The conclusion in (4) cannot be correctly inferred because its degree of support is null: there is at least one philosopher that is not Greek, for example, Ortega y Gasset.

Abductive reasoning

This type of reasoning is the less considered in the typical introductions to logic given that, since 1865, inferences were divided into two classes: induction and deduction. However, Peirce (mentioned in Aliseda 1997, p.10) maintained that there are other types of probable inferences besides induction, namely, abductions.

According to Aliseda (1997, 1998), the difference between induction and abduction can be explained as follows: while abduction is an inference that looks for explanations, induction starts from a series of observations to reach general statements; in this sense, an induction results in predictions, while abduction does not account for further observations (Aliseda, 1997, p.9).

As induction, abduction has no necessity but probability criteria and it is also a synthetic kind of reasoning. Thus, an abduction is a reasoning process that produces explanation with certain inferential structure that is triggered by a fact that needs an explanation (Aliseda, 1998, p.10-11). That is why this inference is also known as *inference to the best explanation*.

Up next we show an example of abduction adapted from Okasha (2002, p.29):

p_1 The cheese in the larder has disappeared, apart from a few crumbs.
p_2 Scratching noises were heard coming from the larder last night. (5)
c Therefore, the cheese was eaten by a mouse.

Received: 13-08-2015 Accepted: 13-11-2015

To reach the previous conclusion several hypothesis may be assumed; however, the mouse hypothesis provides the best explanation given the available data. To illustrate why this is so, let us consider an example with an alternative hypothesis:

p_1 The cheese in the larder has disappeared, apart from a few crumbs.
p_2 Scratching noises were heard coming from the larder last night. (6)
c Therefore, the cheese was eaten by the newborn baby.

The newborn baby hypothesis does not seem to be the better explanation because it needs a set of sub-hypothesis that would require corresponding explanations, for example, that newborns have the actual skills to reach the cheese in the larder and make the cheese disappear; but this explanation is not parsimonious because our experience base allows us to justify that a person of such characteristics is not able to perform such task.

Heuristics

Heuristics are cognitive processes or principles that, generally, promote fast and efficient coding, inference, information retrieval, and production (Morado & Savion, 2002). Heuristics are, therefore, inferential strategies that we use to solve problems. Morado & Savion (2002) explain that some heuristics are learned by experience and by the repetition of successful executions; this concept includes any inferential strategy, whether automatic or deliberately adopted.

There are heuristic taxonomies such as those of Tversky & Kahneman (1974), who expounded three kinds of heuristics employed under uncertainty:

- *Representativeness*: when judgment is guided by similarity.
- *Availability*: when judgment about the probability of an event is evaluated according to the ease with which examples come to mind.
- *Anchoring and adjustment*: when judgment relies on the first piece of information given.

This taxonomy is foundational and has several instances, such as that of Nevo (mentioned in Cohen, 1991), who proposed a taxonomy of strategies used to answer multiple-choice reading comprehension tests (Table I).

Received: 13-08-2015 Accepted: 13-11-2015

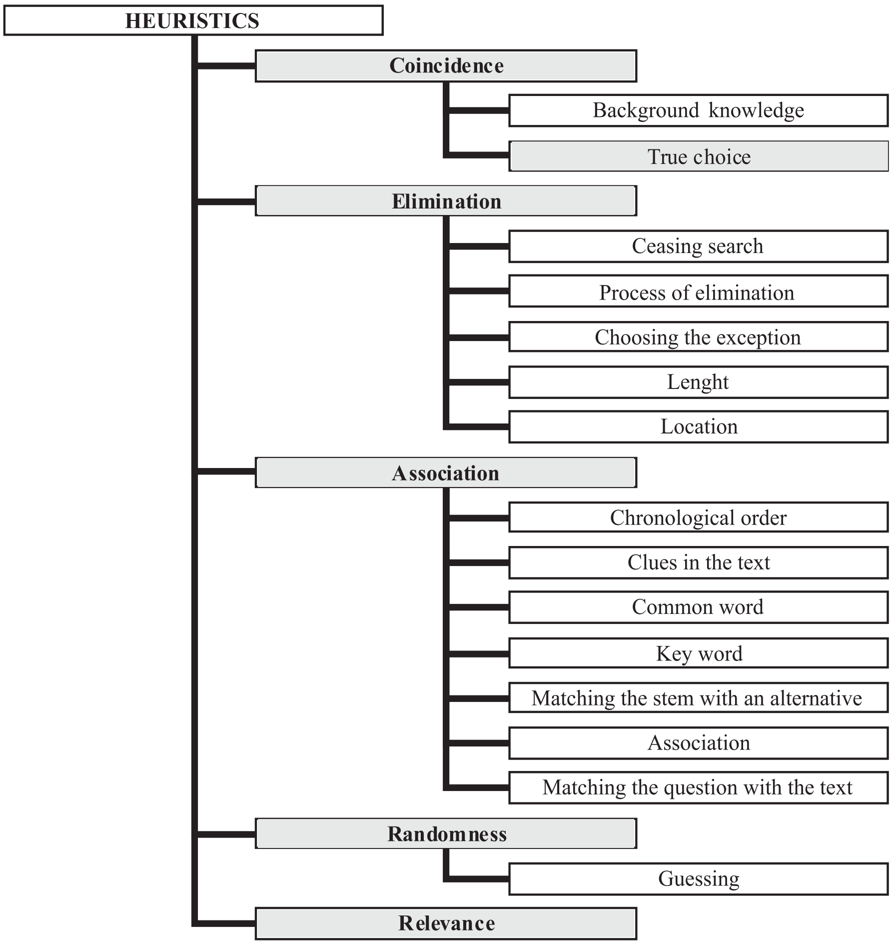
TABLE I. Nevo's taxonomy of heuristics as appears in Cohen (1991)

Strategie	Description
1. *Backgroundknowledge*	General knowledge outside the text.
2.*Guessing*	Guessing without any particular considerations.
3. *Returningtothepassage*	Returning to the text to look for the correct answer, after reading the questions and multiple-choice alternatives.
4. *Chronologicalorder*	Looking for the answer in chronological order in the passage.
5. *Clues in thetext*	Locating the area in the text that the question referred to and then looking for clues to the answer in that context.
6. *Ceasing search at plausible choice*	Reading the alternative choices until reaching one that was thought to be correct. Notcontinuingtoreadtherest of thechoices.
7. *Process of elimination*	Selecting an alternative not because it was thought to be correct but because the others did not seem reasonable, seemed similar, or were not understandable.
8. *Choosingtheexception*	Suspecting a choice to be the correct answer because it constituted an exception or had something different about it.
9. *Length*	Being drawn to an alternative because it was longer/shorter.
10. *Location*	Being influenced by the location of the alternative within the set of alternatives.
11. *Common Word*	Choosing an alternative because it had in it a word that was common that was heard all the time.
12. *Key Word*	Arriving at an alternative because it had in it a word that appeared to be a key word.
13. *Matching the stem with an alternative*	Selecting an alternative because it had in it a word/words that appeared in the item stem as well.
14. *Association*	Selecting the alternative because it had a word in it that evoked an association with a word in the first language or in another language.
15. *Matching the question with the text*	Selecting an alternative because it had a word/words that also appeared in the text, because it had words similar in sound, meaning, or belonged to the same word family or because it just seemed to be related.
16. *Otherstrategy.*	

Source: Cohen, 1991

As we can see, Nevo's taxonomy could be more synthetic without loss of specificity, given that some of those strategies may be related by inclusion. So, under the heuristic of coincidence we include strategy 1 plus another one that we call *true choice*. Under the elimination heuristic we include strategies 6, 7, 8, 9, and 10. Under association we include 4, 5, 11, 12, 13, 14, and 15. Under randomness we include strategy number 2. We have not included strategies 3 and 16 because they are not well defined. In Figure IV we can appreciate this new taxonomy enriched with the heuristic we have suggested, the heuristic of relevance.

FIGURE IV. Taxonomy of heuristics



Source: Personal elaboration

Received: 13-08-2015 Accepted: 13-11-2015

Up next we expound our taxonomy and we emphasize the heuristic of relevance.

Coincidence

The heuristic of coincidence consists in solving problems by matching possible solutions with knowledge or beliefs we already have. In particular, when we talk about a multiple-choice reading comprehension test coincidence is a strategy that pretends to answer the test by looking for matches between our sets of beliefs and the possible answers. In this sense, Hughes (2003) advises not to include items that could be answered by using previous knowledge or beliefs that would not require the students read the body of the text (p.155).

Certainly, coincidence may be successful in solving inferential reading comprehension tests because it is easier to comprehend a text when the topic is already known (Elosúa & García, 1993; Bachman, 1990, p.273), but it also may be an impediment since what these tests pretend to measure is reading comprehension, not whether our base of beliefs agrees with the body of the text (previous knowledge) nor whether the possible answers match our previous knowledge (true choice).

Elimination

Elimination pretends to solve problems by rejecting some solutions until finding out one that seems to be right one. In the context of a multiple-choice reading comprehension test the heuristic of elimination attempts to answer the test by discarding hypothesis by following an inferential process similar to that of the *modus tollens* in which by rejecting certain consequences, certain possible answers get deleted.

This heuristic may be efficient because it is similar to a falsification process, which is a process that in spite of being unable to secure the truth of a hypothesis, it offers the conditions to know when it is false (Popper, 1972); however, it may also be inadequate given that multiple-choice reading comprehension tests pretend to measure reading comprehension, not the skill to reject possible choices.

Received: 13-08-2015 Accepted: 13-11-2015

Association

The heuristic of association tries to solve problems by looking for similarities with previous solutions. In the case of inferential reading comprehension tests this heuristic allows us to look for textual similarities between the body of the text and the possible answers.

This heuristic may be efficient because it is similar to the decision making processes we perform when we confront novel situations in which we look for similarities with previous situations (Klein, 1998). However, when we require an inferential reading, such as the one we require in this study, this heuristic is not always the fittest, given that in these tests we look forward for comprehension, not for the ability to identify similarities.

Randomness

The heuristic of randomness pretends to solve problems by appealing to random choices. In particular, when we talk about inferential reading comprehension tests this heuristic is usually applied when the body of the text is not understood or when there is no enough time to employ another strategy. This last situation agrees with results from Farr, Pritchard, and Smitten (1990), who found that it is very common to answer these tests with as little time as possible.

This heuristic may be useful when we are in situations in which we have no preference between one thing over another (*equidesiderability*), as in Buridan cases (Bratman, 1999, p.11); nevertheless, the use of this heuristic, evidently, does not entail inferential reading comprehension because these tests attempt to measure comprehension, not our guessing skills.

Relevance

We introduce the logical criteria of relevance as a legitimate heuristic for the general resolution of problems and as a solution candidate for the issue of inferential reading comprehension.

In strict sense, relevance is described as the requirement that the premises in a reasoning be *actually used* to reach the conclusion (Mares,

Revista de Educación, 371. January-March 2016, pp. 150-179
Received: 13-08-2015 Accepted: 13-11-2015

2004, p.6). In terms of our homomorphism this means that the body of the text has to be actually used to find the right answer among the possible choices. The logical standard of relevance originated as a demand of rationality after the problem of irrelevance, which consists, *grosso modo*, in that it is possible to find formally correct reasonings with true content that, nevertheless, are irrelevant. An example that illustrates this problem is the next one (adapted from Mares, 2014):

p_1 The moon is made of green cheese. (7)
c Therefore, either it is raining in England right now or it is not.

The reasoning in (7) is correct, in spite of being perplexing, because it is impossible for the premises to be true and the conclusion to be false; the problem, however, is that, despite correct, the premises do not seem to have any connection with the conclusion (Mares, 2004, p.6).

The logical standard of relevance developed to solve this problem requires the antecedent to be actually used to prove the consequent in a given conditional proposition; in other words, the antecedent and consequent must have a common variable in order to share semantic content (Méndez, 1995, p.242-243). The interested reader may find a deeper explanation of these facts in the foundational work of Anderson & Belnap (1975) or in the recent works of Mares (2004) and Priest (2008).

Here we propose this standard of relevance not only as a logical requirement, but as a legitimate heuristic for the resolution of inferential reading comprehension tests. That this is possible should be clear because there is a homomorphism between the structure of the inferential model and the structure of inferential reading comprehension tests.

To illustrate the specificity and usefulness of relevance let us go back to our thought experiment and consider the next example:

(8)

READ THE NEXT PASSAGE AND CHOOSE THE RIGHT ANSWER:

"I still remember that afternoon when I was walking along the levee and discovered a shiny object in a small trash pile. With a curiosity easily explained by my collector's temperament, I bent over to pick it up, and rubbed it against my coat sleeve. I saw that it was a tiny, silver pin engraved with symbols that seemed meaningless to me at the time. I put it in my pocket, and without giving it another thought, I went home." (Ribeyro, 1993)

According to the text, what follows necessarily?
☐ A) the protagonist is a temperamental collector.
☐ B) the protagonist had lost a silver pin in the levee.
☐ C) the pin is now one of the protagonist's belongings.
☐ D) the protagonist will decide to use the pin.

Let us suppose that one of our students, call it the *associative agent*, picks her answers by using *association*. She could pick the possible answer *A* given the similarity between the expressions "collector's temperament" and "temperamental collector"; however, this similarity is by no means an equivalence; just as the similar expressions "the job of my life" and "the life of Job" do not represent the same. She could also pick option *B* given the occurrence of the words "silver pin" and "levee" also present in the body of the text, but that option would also be incorrect because the text describes that the protagonist found the pin, not that he lost it.

Suppose that another student, the *eliminative agent*, discards options *B, C*, *D* because all of them include the word "pin", while option *A* is the only one that misses it (in Nevo's terms we would say that she is using strategy number 8). Imagine that a third agent, the *coincidence agent*, picks option *D* because she has previously read Ribeyro's short story *The insignia*. The choice of option *D* is incorrect, in spite of being true (heuristic of *true choice*). And suppose the fourth student, the *random agent*, picks her answer randomly: she would have a rather low chance of choosing the right answer: 25%.

Finally, let us suppose that we add a new student, a *relevant agent*, that selects her answers by using the heuristic of *relevance*. She would have used the body of the text in order to prove the correctness of option *C*. If we represent the process performed by a relevant agent we could find out the next representation:

p_1 The protagonist noticed and picked up a silver pin that was in trash pile.
p_2 The protagonist put the pin in his pocket (which is one of his belongings) and then went back home.
$\therefore$ By deduction.
c_c The pin is now one of the protagonist's belongings .

To approximate the problem of inferential reading comprehension we have developed a series of tests that relate the types of reasoning and the heuristics previously explained in order to describe their influence in the students' performance.

Method

A transversal study was performed on a random sample of undergraduate students from a center of foreign languages.

Sample

32 undergraduate students (22 female, 10 male) of Italian as a foreign language at an intermediate level (2.5 years of study) whose mother tongue is Spanish. Mean age of the participants was 23.6 years. They were students from different academic areas: Social Sciences and Humanities, Administration, and Sciences. All of them gave informed consent through a letter.

We considered language, type of reasoning, academic area, and length of the body of the text as independent variables; and the number of (in)correct answers, and heuristics as the dependent variables.

Instrument

To design our test we followed the structure of some sections of the tests that evaluate reading comprehension of Italian, such as the *Certificazione di Italiano come Lingua Straniera* (CILS), and we organized the questions according to the types of reasoning.

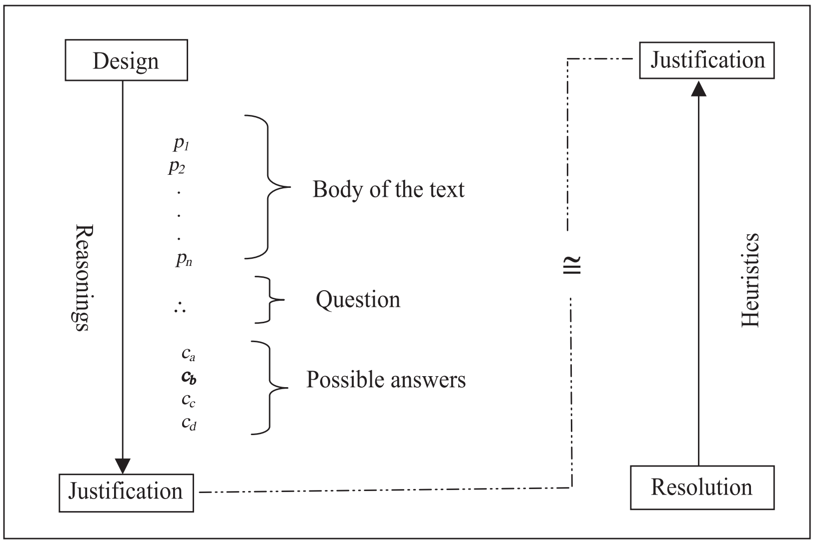
The test consisted of two parts, the first one in Spanish, the second one in Italian, each one with six questions and four possible answers (each answer has a logical justification according to its type of reasoning). The six questions were divided according to the type of reasoning and we registered the heuristics employed for the resolution of each question with an interview that allowed us to extract the justification behind each student's resolution.

We have used Spanish and Italian for two reasons: i) because it is our interest, as teachers of foreign language, to recognize the degree of inferential reading comprehension and Spanish works as a reference language; and ii) because, at the moment, we were working with students that were learning Italian.

Up next we present, as an example, one of the questions we used in our test. We do this as a means to show our methodology. We show the question (body of the text, question, possible answers) with its corresponding *justification* (*top-down* approach) and the heuristic employed in the *justification* of the resolution (*bottom-up* approach), in such a way that is possible to assume some degree of reading comprehension by following the next procedure: if the justification behind the question (*top-down*) is equivalent to the justification of the resolution (*bottom-up*), then we can suppose some degree of inferential reading comprehension. Figure V represents this procedure.

Received: 13-08-2015 Accepted: 13-11-2015

FIGURE V. Representation of the procedure



Source: Personal elaboration

(9)

READ THE NEXT PASSAGE AND CHOOSE THE RIGHT ANSWER:

"A lot can be learned from what surrounds us without having anyone to teach us about it either directly or indirectly but, in contrast, we always have to ask our fellows for the key to enter the symbolic garden of meaning" (Adapted from Savater, 1997, p.31).

According to the text, what follows necessarily?

☐ A) The will to learn is necessary for the existence of learning.
☐ B) Meaning cannot be learned in isolation.
☐ C) The garden of meaning is a metaphor to explain independent learning.
☐ D) The relation with our fellows is fundamental for all us.

Notice that the question in example (9) requires us to find which option is inferred *necessarily* from the given information. This means that the question in turn requires a deduction as a justification. Consider, further, that from the fact that *some* options are wrong does not necessarily follows that the remaining option is the right one: they could all be wrong or there could be more than one right answer, and that is the reason why we need justifications (Table II).

TABLE II. Answers' justifications for example (9)

Choice	Justification
A	If *A* is the right answer, then *A* must the conclusion of a deductive reasoning whose premises are encoded in the body of the text. But if we codify such information we obtain the next configuration of premises: p_1 It is possible to learn many thing without having someone teaching us such things. p_2 If there is meaning comprehension, then one is not isolated. It is evident that option *A* does not necessarily follows from p_1 and p_2 because *A* has information that is not included in $p_1$nor in p_2, namely, "the will to learn".
B	Suppose B is the wring answer. If *B* was the wrong answer, it would have no relation with p_1 or p_2. Now, if we analyze p_2 we notice that it asserts that: p_2 If there is meaning comprehension, then one is not isolated. And if we analyze *B* we notice that is says: c_bMeaning cannot be learned in isolation, which could be expressed in the next equivalent way: f one is isolated, then one cannot have meaning comprehension, which equivalent to p_2.Thus, there is a relation between p_2 y *B*, namely, an equivalence relation and, hence, *B* is the right answer.
C	If *C* is the right answer, then the garden metaphor must express the value of independent learning, but that contradicts p_2, and thus, *C* cannot be the right answer.
D	Suppose *D* is the right answer. Option *D* asserts that "the relation with people is fundamental for every person". If *D* was the right answer, it would be possible to find a proof of *D* from a combination of p_1 and p_2, but there is no such relevant combination.

Source: Personal elaboration

Received: 13-08-2015 Accepted: 13-11-2015

After the test, the students where interrogated with a semi-structured survey in order to know the heuristics employed in the justification of their answers.

Procedure

The test was applied in regular circumstances and under optimal environmental conditions. For the statistical analysis we used GraphPad Prism, 5.01. (GraphPad Software Inc., San Diego, USA). The obtained data distribution normality was evaluated with a Shapiro-Wilk test. We picked Fisher and Xi-squared tests because they allow us to analyze the contingency tables in order to determine whether the number of (in)correct answers is independent from the language or the type of reasoning. The Kruskal-Wallis test, a non-parametric alternative to ANOVA, compares the means obtained in each group, which allowed us to analyze the answers obtained by academic area. Finally, we used Spearman's correlation to evaluate the possible association between two variables (the length of the text and the number of (in)correct answers).

Comparison with respect to language

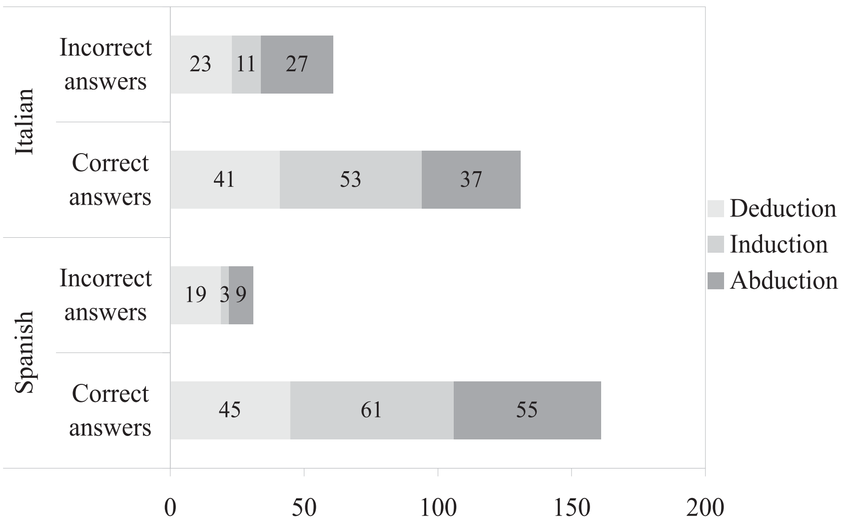
The number of correct answers obtained in Spanish (161) is greater than the obtained in Italian (129). Using an exact Fisher test we determined that there were no statistical differences with respect to language (p=0,0002).

Comparison with respect to type of reasoning

The number of correct answers in Italian with respect to the types of reasoning presented a decreasing order: induction (53), deduction (41), and abduction (37); while in Spanish it presented the next order: induction (61), abduction (55), and deduction (45). Using a Xi-squared test we determined that there were statistical differences between correct and incorrect answers with respect to the types of reasoning in both Italian (χ^2=9,995, p=0,0068) and Spanish (χ^2=15.08, p=0,0005) (Graph I).

Received: 13-08-2015 Accepted: 13-11-2015

GRAPH I. Comparison between correct and incorrect answers with respect to the types of reasoning in both Italian and Spanish



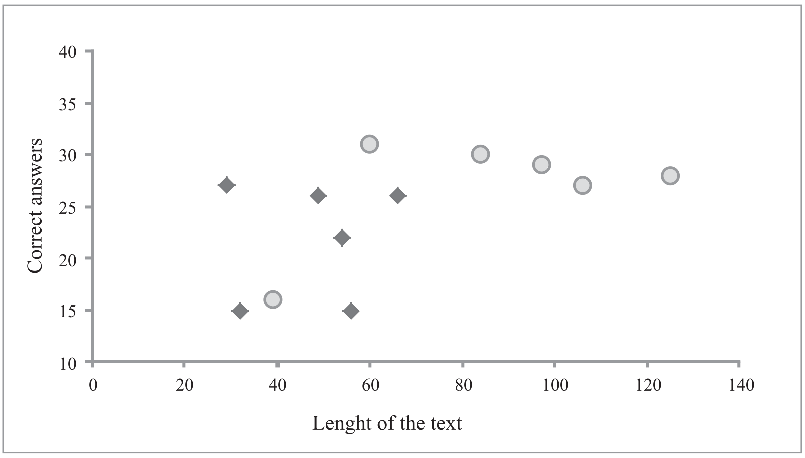
Source: Personal elaboration

Comparison with respect to the length of the text

The Spearman test did not show any statistical correlation between the length of the text and the number of (in)correct answers in Italian (r=0,2648, p=0,6121) nor in Spanish (r=0,08671, p=0,9194) (Graph II).

Revista de Educación, 371. January-March 2016, pp. 150-179
Received: 13-08-2015 Accepted: 13-11-2015

GRAPH II. Correlation between the length of the text and the number of (in)correct answers in Italian (diamonds) and Spanish (dots)



Source: Personal elaboration

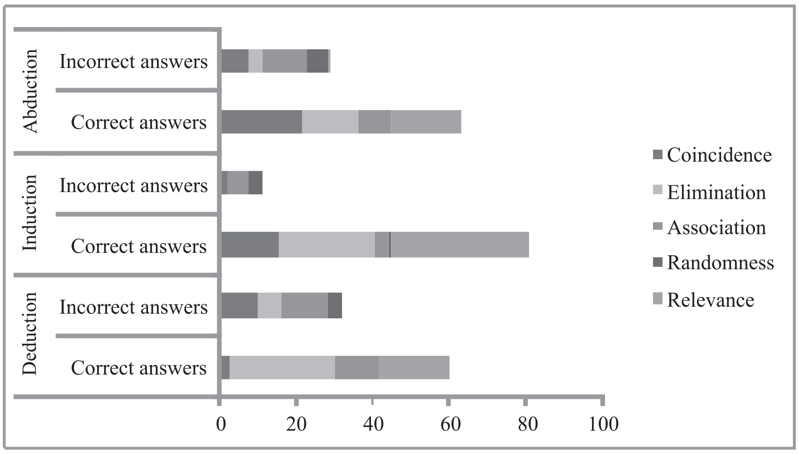
Comparison with respect to academic area

The Kruskal-Wallis test used to analyze the number of correct answers by type of reasoning with respect to academic areas (Social Sciences and Humanities, Administration, and Science) did not show any statistical difference in deduction (k=0,9632, p=0,6178), induction (k=3,541, p=0,1702), or abduction (k=0,1211, p=0,9412).

Heuristics and types of reasoning

As we can see in Graph III, in deductive reasoning the most used heuristic was elimination. The Xi-squared test results showed, however, that the heuristic of relevance was the most efficient with respect to this type of reasoning (χ^2=343.0, p=0,0001). Within inductive reasoning the most used heuristic was relevance and it was also the most efficient with respect to this kind of reasoning (χ^2=470,5, p=0,0001). Within abductive reasoning the most employed heuristic was coincidence; however, the most efficient heuristic for this type of reasoning was relevance (χ^2=260,3, p=0,0001).

GRAPH III. Frequency of use of heuristics w.r.t. types of reasoning



Source: Personal elaboration

Discussion

It is possible to find a homomorphism between the inferential model and the abstract structure of multiple-choice tests that require inferential comprehension. This homomorphism can be analyzed from the points of view of design (*top-down*) and resolution (*bottom-up*). From the former, it is possible to design multiple-choice tests using a taxonomy of types of reasoning; from the latter, it is possible to register the heuristics employed in the resolution of these tests. Since these points of view can be studied experimentally, we have performed a series of measurements that we are going to discuss.

When we started this study we expected the students to perform better in Spanish than in Italian. The reason behind this expectation is evident: being Spanish their mother tongue, they should have greater domain of this language requiring less effort in reading (González Gutiérrez, 2000). Our results confirm this expectation.

We also expected the students to have more correct answers in deductive reasoning with respect to the other types of reasoning, since

Received: 13-08-2015 Accepted: 13-11-2015

deduction is simpler in that the body of the text provides complete information to find an answer (i.e., to infer the conclusion). However, the results showed that there were more correct answers in inductive reasoning, both in Spanish and Italian. This can be the case because, although induction is not the simplest form of reasoning, it is possibly the one we are most used to in order to solve everyday problems.

We had two extra questions: the first one related to the length of the body of the text (for the assumption that there could be a directly proportional relation between the length of the text and its difficulty); the second one, related to the student's academic area (for the assumption that they would be more familiar with certain types of reasonings or contents according to their academical education) (Velásquez *et al.*, 2008, p.134). These questions are not for free: in some occasions the students reported that when they confronted longer texts or unfamiliar texts, they had a difficult time to find the answers.

However, the data we obtained does not show any correlation between the length of the text and the number of correct answers, nor statistical differences when comparing such number with the academic area, both in Spanish and Italian. Roselli, Matute & Ardila (2004) mention that the attention required for reading some text depends on the familiarity with the content and certain skills (p.31), thus, probably the difficulty the students report is an issue of attention, not necessarily of reasoning. Hence, in principle, any student could solve reading comprehension tests disregarding her academical formation, which is consistent with the assumptions of an inferential reading comprehension test.

From the point of view of resolution we noticed that within the questions that required deductive reasoning the most used heuristic was elimination, while the most efficient, i.e., the one that preserved maximum correctness with minimum incorrectness, was relevance. We can provide an explanation for this phenomenon: in the process of elimination the consequences of each possible answer are considered as hypothesis and these get discarded by an inferential process, similar to *modus tollens*, in which by suppressing consequences, hypothesis get canceled (Flores & Fautsch, 1981, p.45). However, the heuristic of relevance was more efficient because deductive reasoning supposes that all the information is given in the body of the text.

In the inductive questions the most used heuristic, and also the most effective, was relevance. This may be explained because inductive reasoning requires a causality standard similar to relevant logic's *entailment*, which supposes that the truth of the premises provides a guarantee to accept the truth of the conclusion (Hawthorne, 2014).

Finally, in the abductive questions the most used heuristics was coincidence, but the most efficient was relevance. This could be the case because abduction requires a set of previous beliefs and coincidence is the heuristic that promotes the use of the body of the text together with one's base of beliefs, which allows to derive the correct answer and, in this case, obtain the most explanatory.

Conclusions

As Hughes asserts (2003), the teacher may contribute to the improvement of evaluation by elaborating better tests or by supporting the people involved in the elaboration of tests (p.5). In this sense, our study suggests that the identification of types of reasoning (in the design of the tests) and heuristics (in the resolution of the tests) could contribute to the evaluation of inferential reading comprehension in a foreign language, since this identification provides a structure of design and measurement of such comprehension, which could get us nearer to a solution to the problem of inferential reading comprehension.

Some limitations of this study were the size of the sample, the number of questions used in the tests, and the costs our proposal has in terms of time and effort, for the teacher would have to dedicate more resources to design the tests and measure the results; however, it seems that this could be a practice that could have favorable consequences in the academical life of the students, consequences that, surely, would require a longer period of time to be observed, which could also be part of a further study.

Currently we are working in the elaboration of a manual to aid the design of inferential reading comprehension tests, with Italian as a foreign language, that implement the criteria we have described in this study (types of reasoning and heuristics) with the goal of providing didactic tools for the teaching-learning process.

Received: 13-08-2015 Accepted: 13-11-2015

References

Alderson, J. C., Clapham, C. & Wall, D. (1995). *Language Test Construction and Evaluation*. Cambridge: Cambridge University Press.

Aliseda, A. (1997). *Seeking Explanations: Abduction in Logic, Philosophy of Science and Artificial Intelligence.* Amsterdam: Institute for Logic, Language and Computation, University of Amsterdam.

Aliseda, A. (1998). La abducción como cambio epistémico: C.S. Peirce y las teorías epistémicas en inteligencia artificial. In *Analogía* (12) UNAM, México, pp.125-144.

Anderson, A. R. & Belnap, N. (1975). *Entailment: the logic of relevance and necessity*, Vol. I. Princeton: University Press

Bachman, L. F. (1990). *Fundamental Considerations in Language Testing*. Oxford: Oxford University Press.

Bratman, M. (1999). *Intention, plans, and practical reason*. Stanford: Center for the Study of Language and Information.

Cohen, A.D. (1991). Testing linguistic and communicative proficiency: The case of reading comprehension. In M.E. McGroarty & C.J. Faltis (Eds.), *Languages in school and society: Policy and pedagogy*, 383-408. Berlin: Mouton de Gruyter.

Cohen, M.R. & Nagel, E. (1934). *An introduction to logic and scientific method*. New York: Harcourt, Brace and company.

Cook, R.T. (2009). *A Dictionary of Philosophical Logic*. Edinburgh: Edinburgh University Press.

Copi, I.M. (1979). *Symbolic Logic* (5a ed.). New Jersey: Prentice Hall. Tr. Cast. Copi, I. (1998). *Lógica simbólica*. México: Compañía Editorial Continental.

Douven, I. (2011). Abduction. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring ed.). Retrieved from http://plato.stanford.edu/archives/spr2011/entries/abduction/

Elosúa, M. R. & García, E. (1993). *Estrategias para enseñar y aprender a pensar*. Madrid: Ediciones Narcea.

Enderton, H. (2001). *A Mathematical Introduction to Logic* (2a ed.). Orlando, Florida: Harcourt/Academic Press. Tr. Cast. Enderton, H. (2004). *Una introducción matemática a la lógica* (2a ed.). Trad. José Alfredo Amor Montaño, Instituto de Investigaciones Filosóficas, UNAM.

Farr, R., Pritchard, R. & Smitten, B. (1990). A description of what happens when an examinee takes a multiple-choice reading comprehension test. In *Journal of Educational Measurement*, 27(3), 209-226.

Flores, M.A. & Fautsch, E. (1981). *Temas selectos de matemáticas*. México: Editorial Progreso.

González Gutiérrez, M.C. (2000). La habilidad de la lectura: Sus implicaciones en la enseñanza del inglés como lengua extranjera o como segunda lengua. In *Revista de Ciencias Humanas* [on line] n.19.

Hawthorne, J. (2014). Inductive Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of philosophy* (Winter 2014 ed.). Retrieved from http://plato.stanford.edu/archives/win2014/entries/logic-inductive

Hughes, A. (2003). *Testing for Language Teachers* (2a ed.). Cambridge: Cambridge University Press.

Klein, G.A. (1998). Recognition-primed decision making. In *Sources of power: How people make decisions*. Cambridge, MA: MIT Press, p.15-30.

Mares, D.E. (2004). *Relevant Logic. A Philosophical Interpretation*. Estados Unidos: Cambridge University Press. Cap. 1

Mares, D.E. (2014). Relevance Logic. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of philosophy* (Spring 2014 ed.). Retrieved from http://plato.stanford.edu/archives/spr2014/entries/logic-relevance/

Mates, B. (1972). *Elementary Logic*. New York: Oxford University Press. Tr. Cast. Mates, B. (1979). *Lógica Matemática Elemental*. Madrid: Tecnos.

Méndez, J.M. (1995). Lógica de la relevancia. In *Lógica: Enciclopedia Iberoamericana de Filosofía*. Madrid: Trotta, pp.237-270.

Morado, R. & Savion, L. (2002). Rationality, Logic, and Heuristics. In *International Conference on Artificial Intelligence*, Vol. 2, CSREA Press, pp.791-797.

Okasha, S. (2002). *Philosophy of Science: A very Short Introduction*. Oxford: Oxford University Press, cap. 2.

Palencia del Burgo, R. (1990). La evaluación como diagnóstico y control. In Bello, P., Feria, A., Ferrán, J.M., García Hernández, T., Gómez, P., Guerrini, M.C., López Hernández, J., Martín Catalán, D., Martín Viaño, M.M., Martos, M.R., Mata Barreiro, C., Navarro, A., Palencia, R., Ravera, M., Salaberri, S., Sierra, J.M., & Verdú, M. *Didáctica de las segundas lenguas. Estrategias y recursos básicos.* Madrid: Aula XXI Santillana.

Pearl, J. (2000). *Causality: Models, Reasoning, and Inference*. Cambridge: Cambridge University Press.

Pérez Zorrilla, M.J. (2005). Evaluación de la comprensión lectora: dificultades y limitaciones. In *Revista de Educación*, num.

Received: 13-08-2015 Accepted: 13-11-2015

Extraordinario, pp.121-138. Retrieved from http://www.revistaeducacion.mec.es/re2005_08.htm
Popper, K.R. (1972). *Conjectures and refutations*: the growth of scientific knowledge. London: Routledge. Tr. Cast. Popper, K. (1983). *Conjeturas y refutaciones*, Madrid: Paidós.
Priest, G. (2008). *An Introduction to Non-Classical Logic: From If to Is*. Cambridge: Cambridge University Press.
Ribeyro, J.R. (1993). The insignia. In *Marginal Voices: Selected Stories*. (D. Douglas, Trans.) Texas: University of Texas Press.
Rodríguez Rodríguez, R. (2005). Abducción en el contexto del descubrimiento científico. In *Revista de Filosofía de la Universidad de Costa Rica***.** 43 (109/110), 87-97.
Roselli, M., Matute, E. & Ardila, A. (2004). Características neuropsicológicas y aprendizaje de la lectura en niños hispanohablantes. In Matute, E. (coord.) *Aprendizaje de la lectura. Bases biológicas y estimulación ambiental.* México: Universidad de Guadalajara.
Savater, F. (1997). *El valor de educar***.** Barcelona: Ariel.
Tversky, A. & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. In *Science*, Vol. 185, 1124–1131. doi:10.2307/1738360
Van Dalen, D. (2004). *Logic and Structure* (2a ed.). Berlín: Springer-Verlag.
Velásquez, M., Cornejo, C. & Roco, A. (2008). Evaluación de la competencia lectora en estudiantes de primer año de carreras del área humanista y carreras del área de la salud en tres universidades del consejo de rectores. In *Revista de Estudios pedagógicos* [online]. Universidad Austral de Chile. 34(1), pp.123-138.

Contact address: Alejandra Platas-García, Benemérita Universidad Autónoma de Puebla, Facultad de Lenguas, Centro de Lenguas Extranjeras. Avenida San Claudio y 22 Sur, sin número. Edificio CAALE. Colonia Jardines de San Manuel, Puebla, C.P. 72570. México. E-mail: aplatasg@gmail.com

Received: 13-08-2015 Accepted: 13-11-2015

La *Revista de Educación* es una publicación científica del Ministerio de Educación, Cultura y Deporte español. Fundada en 1940, y manteniendo el título de *Revista de Educación* desde 1952, es un testigo privilegiado de la evolución de la educación en las últimas décadas, así como un reconocido medio de difusión de los avances en la investigación y la innovación en este campo, tanto desde una perspectiva nacional como internacional. La revista es editada por la Subdirección General de Documentación y Publicaciones, y actualmente está adscrita al Instituto Nacional de Evaluación Educativa de la Dirección General de Evaluación y Cooperación Territorial.

GOBIERNO
DE ESPAÑA

MINISTERIO
DE EDUCACIÓN, CULTURA
Y DEPORTE

NIPO línea: 030-15-016-X
NIPO ibd: 030-15-017-5
ISSN línea: 1988-592X 0034-8082
ISSN papel: 0034-8082

www.mecd.gob.es/revista-de-educacion