Interactive Learning Environments
Vol. 20, No. 5, October 2012, 423–439

Aids to computer-based multimedia learning: a comparison of human tutoring and computer support

H. García Rodicio* and E. Sánchez

Department of Developmental and Educational Psychology, University of Salamanca, Salamanca, Spain

(*Received 5 November 2009; final version received 27 April 2010*)

Learners are usually provided with support devices because they find it difficult to learn from multimedia presentations. A key question, with no clear answer so far, is how best to present these support devices. One possibility is to insert them into the multimedia presentation (canned support), while another is to have a human agent provide them (human tutoring). Human tutoring poses potential advantages: it uses spoken modality, displays non-verbal cues and implies social interaction. However, there is mixed evidence regarding these supposed advantages, and prior research comparing human and computer support presents problems. Our goal was to explore whether the advantages of human tutoring actually exist while avoiding the problems of prior research. In one experiment, participants learned Geology from a multimedia presentation including one of three forms of support: human tutoring, canned support or no support. After viewing the presentation, participants solved retention and transfer tests. Results revealed that participants in the human tutoring condition outperformed those in the other two conditions, who did not differ from each other. This means that human tutoring is advantageous, a fact that has implications in the design of support devices in multimedia learning.

Keywords: multimedia learning; monitoring process; support devices; human tutoring

1. Introduction

As learners find it difficult to learn from multimedia instructional messages (see below), they are often provided with some kind of support devices (e.g. corrective feedback, prompts). These devices assist learners in the execution of the learning processes necessary to build coherent mental models, such as the monitoring process, through which learners check their emerging understanding. A key question is what the best way of providing support devices is: a conventional or a technology-based approach. In the conventional approach, which we will call *human tutoring*, the learner–computer interaction is mediated by a human agent who provides instructional support (as can happen in the classroom). In the technology-based approach, support devices are inserted into the computerised material (henceforth,

*Corresponding author. Email: garciarodicio@usal.es

ISSN 1049-4820 print/ISSN 1744-5191 online

http://dx.doi.org/10.1080/10494820.2010.500090
http://www.tandfonline.com

canned support) so that there is a strict human–computer interaction. Although using a human agent to provide support devices might have some advantages, such as the use of the spoken modality, the display of non-verbal cues and the inclusion of a social component, it is not clear whether these advantages are actually present, as will be shown below. For this reason, it seems warranted to compare human and computerised forms of support. In fact, a prior experiment (Azevedo, Cromley, & Seibert, 2004) explored this question; it had, however, one limitation: not only the form of support but also its quantity and quality differed between the conditions. Therefore, our goal was to carry out a comparison of human and computerised support avoiding the limitations observed in prior research. From a theoretical perspective, the results of such a comparison may elucidate whether or not the supposed advantages of human tutoring are a reality. On the practical side, shedding light on the question may be useful when designing support devices; for instance, the results may provide suggestions about how to support students who are learning from book-based and computer-based multimedia presentations in the classroom.

1.1. The need for support in learning from multimedia presentations

Multimedia learning occurs when cognitive processes are executed to generate a mental model of the topic described in a presentation involving words and pictures (Mayer, 2001; Schnotz, 2005). Several difficulties may arise in doing so. More accurately, learners have to assimilate many elements and interconnections (Pollock, Chandler, & Sweller, 2002) within and between verbal and pictorial information sources (Ainsworth, Bibby, & Wood, 2002) whilst this wealth of information has to be processed in the limited capacity of the working memory (Sweller, van Merriënboer, & Pass, 1998). This circumstance makes it likely for learners to construct poor mental models.

In order to deal with that circumstance, learners are supposed to self-regulate their learning. According to some approaches (Chi, 2000; Hacker, 1998; Moos & Azevedo, 2009; Otero, 2002), this mainly consists of *monitoring* our own emerging understanding. In fact, monitoring is a critical process in achieving deep learning from instructional materials (Azevedo, Guthrie, & Seibert, 2004; Cain, Oakhill, & Bryant, 2004; McNamara, 2004). Nevertheless, such a process is difficult for learners to perform (Commander & Stanwyck, 1997; Otero & Kintsch, 1992).

Given these difficulties, if instructional materials are to be effective, they have to be enriched with *support devices*. Support devices for the process of monitoring can be, for instance, corrective feedback contingent on learners' responses to inserted questions (Campbell & Mayer, 2008), passages making learners' typical misconceptions explicit (Mikkilä-Erdmann, 2001) or prompts asking learners to keep track of their ongoing understanding (Hausmann & Chi, 2002).

1.2. Different ways of providing support devices

When providing learners with support devices, educators can follow at least two approaches, while it is still unclear which is the better. One possibility is to insert support devices into the instructional material itself. For instance, in helping learners to monitor their understanding, multimedia presentations can include questions to learners and provide predetermined feedback on their responses (see Figure 1). Another possibility is to use a human agent, physically present near learners, to

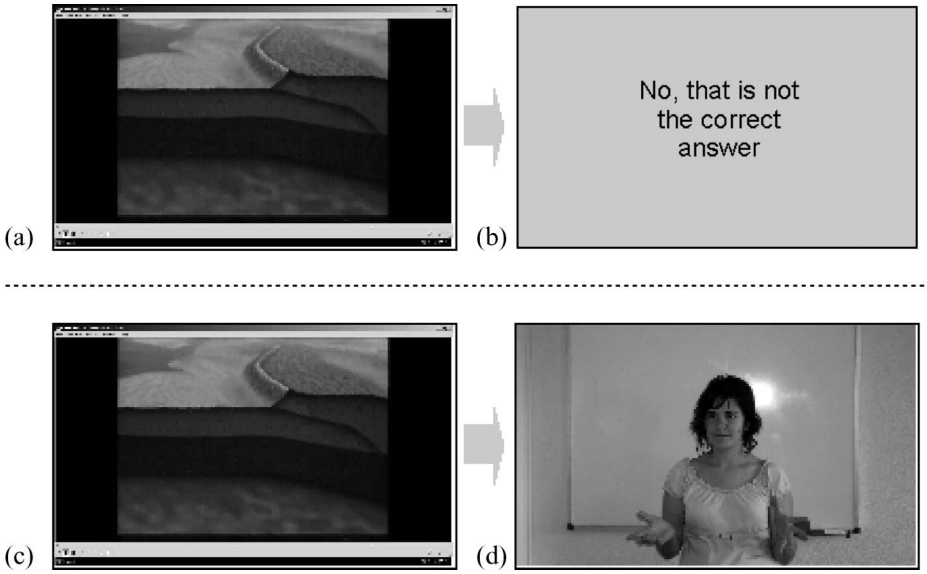

Figure 1. An animation with concurrent narration followed by canned support (b). The same animation with concurrent narration followed by human tutoring (d). The screenshots (a,c) are extracted from the materials used in the present experiment.

provide support. For instance, during the use of a multimedia presentation, the human agent can pose questions to learners providing feedback on their responses as an aid to monitoring (see Figure 1). The first strategy, called *canned support*, responds to a technology-based approach, in which learners interact with computerised instructional resources on their own. The second one, called *human tutoring*, responds to a more conventional approach, in which a human agent mediates learner–computer interaction.

The human tutoring approach presents potential advantages over canned support. First, human agents use the spoken modality to provide support devices while there is ample evidence towards this modality being better than the visual one for the presentation of verbal information in multimedia learning (see Ginns, 2005 for a meta-analysis of the modality principle). Second, human agents convey multiple non-verbal cues, such as facial expressions, eye gaze, hand gestures or body postures that might be helpful when facilitating learning due to their communicative value (Ekman et al., 1987; Kleinke, Staneski, & Berger, 1975; Valenzeno, Alibali, & Klatzky, 2003; Walters & Walk, 1988). Third, the presence of an agent engages learners in a social interaction, which increases their involvement in the task and, thus, enhances learning (Moreno, Mayer, Lester, & Spires, 2001).

Based on these potential advantages, one may predict human tutoring to be better than canned support. Nevertheless, these supposed advantages may not necessarily be present for several reasons. First, although the spoken is better than the written modality when presenting words and pictures which are to be mentally integrated (Ginns, 2005), it is not clear if the same holds true for words presented in isolation, such as those serving as support devices. In fact, there are discrepant results regarding the effect of the modality of support devices originating from the field of multimedia learning. Seufert, Schütze, and Brünken (2009) had participants who learn Chemistry from a multimedia presentation. Half of them received spoken prompts asking them to mentally integrate words and pictures whereas the other half received identical prompts presented in written modality. The former participants outperformed the latter. Sánchez and García Rodicio (2008) asked participants to learn Geology from a multimedia presentation. As well as animations with narration, the presentation comprised elaborations aimed at revising learners' emerging understanding. They were presented in either spoken or written modality. Participants receiving written elaborations performed better on retention and transfer tests, as compared to those receiving spoken elaborations. Furthermore, there is also mixed evidence in this respect originated in other fields, such as example-based or problem-based computerised instruction (Atkinson, 2002; Graesser et al., 2003; Moreno & Mayer, 2002; Moreno et al., 2001). Atkinson (2002), for instance, found that elaborations accompanying worked-out examples in the topic of probability were equally effective in either written or spoken modality, while Moreno et al. (2001), when asking participants to learn Botany from an environment with problems to solve, found that participants receiving spoken feedback outperformed those receiving written feedback. Overall, results are not conclusive regarding which modality is better in the provision of verbal aids.

Second, although non-verbal cues do indeed provide information, it is not clear that such information can facilitate learning. Evidence supporting this doubt comes from research on animated pedagogical agents. Atkinson (2002) had participants learn probability from an environment involving worked-out examples. The environment included support devices in the form of elaborations explaining the rationale behind the solutions to the examples. These elaborations were inserted into the computerised environment either with or without an accompanying animated agent. The agent was a green parrot exhibiting non-verbal cues, such as hand gestures. In one of two experiments, participants viewing the agent outperformed those in the without-agent condition. Graesser et al. (2003) asked participants to learn computer literacy by solving problems in a computer-based environment. They received feedback from either a visible agent, displaying multiple cues (e.g. eye gaze, facial expressions), or an invisible one. This variable made no significant difference, although there was a trend in the expected direction. Baylor and Ryu (2003) had participants learn psychology by using an environment including a set of problems to solve. Participants received feedback on their steps towards solutions from either a visible or an invisible agent. There were no differences between conditions in a transfer test. The same goes for the experiments of Moreno et al. (2001). In instructing participants on Botany they presented them with problems to solve and feedback on their steps towards solutions. For some participants feedback was provided by Herman, an alien bug displaying non-verbal cues; others did not see Herman. Seeing Herman (or even a human-like agent) had no impact. Taken in conjunction, the experiments exploring the effect of non-verbal cues on learning provide a mixed pattern of results.

Third, people often experience a human–human interaction when using computers (Nass, Moon, Fogg, Reeves, & Dryer, 1995). Nass et al. (1995) recruited a sample of participants, categorised as dominant or submissive. Participants interacted with dominant or submissive computer programmes. Participants interacting with a programme similar to them reported a preference with respect to programmes not similar to them. This resembles the law of attraction in human–human relationships. Therefore, even when using computers, one can experience a human–human interaction. This contrasts with the rationale behind the persona effect (Moreno et al., 2001). This notion assumes that the more social presence (e.g. audible voice, visible face) an agent has, the more involvement learners have in the task, which in turn benefits learning. In keeping with the results in Nass et al. (1995), one should expect involvement even when rough computer programmes (with almost no social presence) are used.

In conclusion, although human tutoring poses potential advantages over canned support, it is not clear if these are actually present. This doubt was confirmed by examining research on the impact on learning of the modality of support devices, non-verbal cues and human presence.

1.3. Prior work comparing human and computer support

Given that evidence with regard to the potential advantages of human tutoring is not conclusive, it seems reasonable to compare human tutoring and canned support. A prior experiment by Azevedo et al. (2004) attempted to make this comparison but with limited success, as the reader will appreciate.

Azevedo et al. (2004) had participants learn about the human circulatory system from a hypermedia presentation. Part of the participants used the presentation and received a list of hints helping them to monitor their emerging understanding. Other participants used an identical presentation having access to a human tutor, who provided them with monitoring aids. The rest of the participants received only the presentation, with no support devices. Results revealed that participants in the human agent condition learned more deeply, as compared to those in the other two conditions, which did not differ from each other. This means that the potential advantages of human tutoring proved to be actual advantages. A shortcoming of this experiment, however, is that the support devices in the list were not strictly the same as those provided by the human tutor. They were not so because the tutor (a) decided when learners needed support and (b) was free to provide learners with the most appropriate aid. This made support devices in the human condition more frequent and custom-built, as compared to those in the list condition. So, the advantage of the human tutoring condition might be due to either the way in which support devices were delivered (i.e. via human or via computer) or the quantity and quality of support devices used by the tutor. Therefore, it remains to be explored what would happen if support devices in the canned and human conditions were strictly the same with only one exception: the way in which they are presented. A comparison meeting such constraints would be a more stringent test of the supposed advantages of human tutoring.

1.4. Overview of the experiment

The goal of the present experiment was to compare canned and human forms of support in learning from a computer-based multimedia presentation. Special care was taken in keeping support devices strictly equal in both conditions, which represents an improvement in comparison with prior research.

With that goal in mind we carried out the following experiment. High-school students learned Geology (plate tectonics) from a multimedia presentation including animations with concurrent narration and different support configurations. Support devices were constructed as follows. During the course of the presentation, multiple-choice questions were posed for which participants had to find an answer, receiving corrective and elaborative feedback on their responses. Corrective feedback detected limitations in learners' responses helping learners to monitor their emerging understanding; elaborative feedback revised learners' responses helping them to construct coherent mental representations. Participants in the canned support condition received the questions and the corrective and elaborative feedback on the computer screen, presented in written modality. Those in the human tutoring condition received the questions and the corrective feedback from a human agent, while the elaborative feedback was displayed on the computer screen. Those in the control condition received no questions and no feedback but only the statements comprised in the elaborative feedback, presented on the computer screen. After using the multimedia materials, participants were asked to solve retention and transfer tests.

2. Method

2.1. Participants and design

Eighty-four ninth-graders studying high-school in Las Palmas (Spain) participated in this experiment. Participants, who volunteered to take part in the experiment (with parents' permission), were randomly assigned to one of three conditions. Twenty-seven participants served in the human tutoring condition, 29 served in the canned support condition and 28 served in the control condition. The mean age of the sample was 14. The sample was $\sim$60% females. The ethnicity of all participants was Spaniard, Spanish being their first language. They all came from medium-income families. All participants reported that they used computers frequently (for gaming and Internet browsing). They all had comparable levels in their reading comprehension skill, as measured by a standardised test, namely, PROLEC-SE (Ramos & Cuetos, 1999): $F(2, 81) = 1.48$, MSE $= 3.72$, $p > 0.10$. This means that there were no significant differences in their perceptual skills (i.e. vision) or their ability to build coherent mental representations from instructional texts.

The experiment had a one-factor design with the form of support (human/canned/no support) as the between-subjects factor. Retention and transfer test scores were used as the dependent variables. Prior domain knowledge was used as a control variable.

2.2. Materials

The *prior knowledge test* consisted of a set of seven open-ended questions. These questions tested basic Geology notions and issues directly addressed in the material to-be-learned. These questions were the following: "What is a tectonic plate?"; "Can continents move horizontally? Explain why"; "Is it possible for the Earth's surface to be recycled? Explain why"; "How are mountains formed?"; "How are volcanoes formed?"; "Why do some mountains have volcanoes while others do not?"; "Place the following elements in the illustration: ridge, subduction, tectonic plates, volcanoes, magma" (in this question an illustration was shown; participants had to match names from a list with parts of the illustration).

The *multimedia presentation* included both (a) animations with concurrent narration and (b) support devices (see Figure 1). The animations with concurrent narration described several events concerning plate tectonics, namely, (1) the three layers of the internal structure of the Earth and their relations, (2) convection currents, (3) ridges and the process through which new crust is created, (4) convection currents as the origin of plate movements and collisions, (5) the collision between a continental and an oceanic plate and its consequences on the Earth's surface, the Andes range as an example of this type of collision, (6) the collision between two continental plates and its consequences on the Earth's surface, the Himalaya range as an example of this type of collision, (7) the destruction of crust in the trenches. The modules were presented in a sequential fashion and, in conjunction, lasted 583 s (10 min approx.).

The mental model we wanted participants to build from the presentation was the following: "The Earth's core is a very warm sphere. It makes the magma (i.e. a doughy substance made of molten rock) in the mantle heat up and, thus, approaches the crust. When the magma cools again, it moves away from the crust. As a result, convection currents are formed: magma is continuously moving up and down. The magma in the mantle surfaces through the ridges (i.e. rifts in the crust) because of the movement the currents provoke. Once it is on the surface, it gets colder and solidifies creating new crust. Convection currents also push the plates, as these are floating on the magma, making them move away from each other. When moving, plates can collide with other plates. There are different kinds of collisions depending on the plates engaged in the crashes, namely, one continental plate and one oceanic plate or two continental plates. Thus, plates can collide and one can sink (in the trenches) inside the Earth originating mountains with volcanoes (e.g. the Andes range); or they can collide and move vertically originating mountains without volcanoes (e.g. the Himalaya range). When a plate sinks in the trenches it is destroyed and becomes magma again whereas the magma surfacing through the ridges solidifies creating new crust. This means there is a continuous recycling loop between trenches and ridges."

The material also included *support devices* helping participants to detect and repair possible understanding problems. Based on our prior studies (Sánchez, García Rodicio, & Acuña, 2009), the most common misunderstandings were identified. There were three aspects typically misunderstood by participants in prior studies[1]. First, the notion of ridge was distorted. A number of participants thought of ridges as isolated and small cracks instead of large cracks all over the Earth's surface. Second, the specific features of the Andes and the Himalaya plates' collisions were mixed-up. Many participants underestimated the particular characteristics of each type of collision, building a mixed model in which features of both types of collision were included. Finally, most participants could not grasp the idea of the recycling loop linking the activity in the ridges and that in subduction. They grasped either the former or the latter part of the relationship but not the link between them. Accordingly, three supporting devices were created, each involving two parts. All the devices had a corrective feedback part, which helped participants to monitor their learning process to detect misunderstandings, and an elaborative feedback part, which assisted participants in revising and repairing their flawed mental representations. An example of these aids is shown in Figure 2 (translated into English from the original). As shown in this Figure, during the presentation a question was posed to the participant (i); then he/she received corrective feedback on his/her answer (iii); finally, elaborative feedback revised and repaired possible flaws in his/her understanding (iv).

Participants who chose the correct answer ((c) in the example) were told that a more accurate response was, nevertheless, needed. Therefore, they were redirected to the elaborative part. This means that all participants saw the elaborative part.

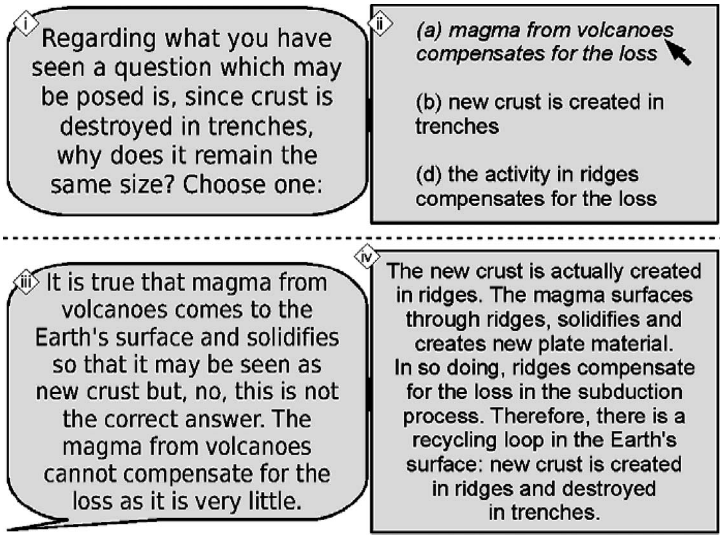

Figure 2. Extract from the support devices in the material to be learned. i: A critical question that participants answered. ii: A list of possible answers (here the participant chose (a), which is not correct). iii: Corrective feedback telling the participant what aspects are wrong in his/her understanding. iv: Elaborative feedback revising his/her flawed understanding.

Depending on the experimental condition, participants received either both corrective and elaborative feedback parts or only elaborative. Participants in the human and canned support conditions were provided with both corrective and elaborative parts; those in the control condition received elaborative parts but no corrective parts at all. In these cases, explanations in the elaborative parts were not presented as a response to the participant's answer but as additional information in the materials (e.g. "A question you must consider is that new crust is created in the ridges ..."). These explanations were included in the control condition in order to ensure that all conditions strictly had access to the same contents to-be-learned.

Moreover, depending on the condition, participants received corrective feedback either from a human agent (human tutoring condition) or from the presentation itself (canned support condition). The human agent was present at the experimental session. She was an experienced lecturer (Nadezhna Castellano) who was instructed to use her expressiveness. Expressiveness included here: intonational and temporal variations in speech, facial expressions, eye gaze, hand gestures and body postures. A convincing corrective feedback was that clarifying that the communicative intention was to warn about a misunderstanding and displaying an attitude of importance and urgency. Special attention was paid to doing this in a natural way. In using variations in speech, stress (i.e. relative prominence of syllables), high tones (i.e. in relation to the baseline frequency) and slow rhythm (i.e. syllables uttered by time unit) were applied to relevant passages. In using non-verbal cues, the agent was instructed to frown, look at participants' eyes, use beat gestures and lean toward the participants when uttering relevant passages. The experimental session was video-recorded and two judges checked the recordings to ensure that (a) the agent was both natural and convincing when providing corrective feedback and that (b) the words uttered by the agent matched those included in canned support devices.

Therefore, the canned support condition displayed the same words uttered by the agent but on the computer screen, presented in written modality.

The *retention test* consisted of a set of five open-ended questions. These questions required participants to recall key information that was presented in the multimedia material. The test included these questions: "Why are tectonic plates permanently moving and crashing?"; "What are convection currents?"; "What is a ridge?"; "Explain how the Earth's surface can be recycled"; "What are the differences between the plate collisions in the Andes and those in the Himalaya?"

The *transfer test* consisted of a set of nine open-ended questions. These questions presented some hypothetical conditions to participants, who had to predict some results based on the knowledge they had acquired. The test involved the following questions: "Imagine that convection currents stop working, what would happen?"; "Imagine tectonic plates stop moving, how would you explain that?"; "Imagine that convection currents start moving at half their speed, how would you explain that?"; "What would you expect if ridges were small, isolated holes in the Earth's surface?"; "What would you expect if the sub-duction process destroyed more crust than that created in ridges?"; "In the future, will there be more/less crust than now?"; "Could the Himalaya range have volcanoes?"; "Why are there both mountains and volcanoes in Italy but only mountains in Spain?"; "Imagine eruptions in the Andes stop, how would you explain that?"

2.3. Procedure

A given participant was randomly assigned to either the human, canned or no support condition. The participant was seated in front of his/her individual computer and headphones. Nearby, an experimenter was present in the session. It was the same experimenter for all participants. She was the person in charge of providing the instructions for the experiment, delivering and collecting the tests, and ensuring that the computers worked properly. She also worked as the tutor in the human tutoring condition.

First, the participant received some basic instructions from the experimenter. These instructions were as follows (identical in all experimental conditions): "Thank you for participating in this experiment. We are interested in how people learn from multimedia instructional materials. You will be asked to use a computer-based multimedia presentation on plate tectonics. Please, pay attention to this presentation as after watching it you will have to solve some questions. Before using the computer material, we want you to fill in a prior knowledge test on Geology. Please, try to remember all the things you know about the subject." The participant was also told not to interrupt or disturb the experimenter (video-recordings of the sessions made it possible to ensure this). This was particularly important for the human tutoring condition: although participants in this condition received support devices from the tutor, maintaining open interaction with her was not allowed.

After receiving the instructions, the participant was handed the prior knowledge test. Solving this test took no more than 15 min.

When the participant had finished filling in this test, he/she started using the material. The presentation lasted about 900 s (15 min) including the animations with narration (583 s, 10 min approx.) and the support episodes.

The participant, then, viewed the animations with concurrent narration describing the events listed before about the plate tectonics theory. In addition, he/she received three support devices (each including corrective and elaborative feedback parts) in one of the forms described above. Specifically, participants in the human tutoring condition received corrective feedback from the experimenter. She was instructed to present support devices in a natural and expressive (via prosody and non-verbal cues) fashion, as explained before. After each corrective feedback, these participants viewed the elaborative feedback on the computer screen, presented in written modality. The same was true for participants in the canned support condition, except for the fact that they also viewed the corrective feedback on the screen, presented in written modality. The time that these participants spent on corrective feedback frames was recorded, showing that they did process them, instead of just skipping them[2]. Participants in the control condition viewed the elaborative parts on the screen but received no corrective feedback.

After the presentation, the participant was given the retention test. Solving this test took no more than 15 min. Then he/she was given the transfer test, having no more than 15 min to solve it. After this final test was collected, the participant was seen off. Each session lasted about 75 min.

2.4. *Scoring*

A rater (i.e. the one in charge of rating the tests) scored all the tests unaware of the condition of each participant. A second rater also scored $\sim$30% of the tests. Inter-rater agreement was 0.93 on the prior knowledge test, 0.81 on the retention test and 0.91 on the transfer test. Disagreements were solved by consensus.

A template with possible answers was developed for all the tests. It included accurate, correct but incomplete and incorrect answers for each question (see Table 1). These answers yielded 2, 1 or 0 points, respectively. Total scores ranged from 0 to 14 in the prior knowledge test. Total scores ranged from 0 to 10 points in the retention test. Total scores ranged from 0 to 18 in the transfer test.

2.5. *Results*

Prior knowledge was analysed first in order to ensure that all conditions had a comparable level in this variable. To this end, a one-way analysis of variance with condition (human tutoring/canned support/control) as the between-subjects factor was used. Then, in order to compare the impact on learning of the different conditions, the retention and transfer test scores were analysed using a one-way analysis of variance with condition as the between-subjects factor. *Post-hoc* pair-wise comparisons, based on a Scheffé's F test, were conducted afterwards (when an analysis of variance reveals a significant effect, *post-hoc* comparisons make it possible to explore which means are significantly different from each other). Cohen's d was calculated as a measure of effect size whenever there was a significant effect (or a trend). According to Cohen (1988), values from 0.00 to 0.30 were interpreted as small effects, values from 0.40 to 0.60 as medium effects and values from 0.70 to 2.00 as large effects. An α of 0.05 was used throughout this article. All scores are shown in Table 2.

Regarding prior knowledge, there were no significant differences between conditions, as revealed by an analysis of variance, $F(2, 81) = 1.60$, $MSE = 6.70$, $p > .10$. This result indicates that participants in the three experimental conditions exhibited similar levels of prior knowledge on Geology.

Table 1. Questions in the prior knowledge, retention and transfer tests and accurate answers.

Test	Question	Accurate answer
Prior knowledge	"What is a tectonic plate?"	"They are blocks dividing the Earth's surface into pieces, which move permanently causing different phenomena"
	"Can continents move horizontally? Explain why"	"Continents move because they are plates or part of a plate, which in turn are moving permanently"
	"Is it possible for the Earth's surface to be recycled? Explain why"	"It is recycled by the creation of new crust in ridges and the destruction of old crust in trenches"
	"How are mountains formed?"	"For mountains to be created, two tectonic plates must collide and push each other"
	"How are volcanoes formed?"	"For volcanoes to be created, two plates must collide and push each other in such a way that cracks are formed in one of them, through which magma surfaces"
	"Why do some mountains have volcanoes while others do not?"	"They differ in the plates involved in their crashes: continental–continental to produce mountains, oceanic–continental to produce volcanoes"
	"Place these elements in the illustration"	One mark on a crack over the Earth's surface ("ridge"); one mark on the region where an oceanic plate sinks ("subduction"); some marks on the pieces into which the Earth's surface is divided ("plates"); one mark on mountains with eruptions ("volcanoes"); one mark on the substance under plates ("magma")
Retention	"Why are tectonic plates moving and crashing permanently?"	"Plates move because they are floating on magma, which in turn is moving permanently due to convection currents, and plates collide because sometimes two moving plates converge"
	"What are convection currents?"	"Convection currents are created by cold and warm magma going up and down"
	"What is a ridge?"	"Ridges are cracks all over the Earth's surface dividing it into plates. Through ridges magma surfaces and solidifies creating crust"
	"Explain how the Earth's surface can be recycled"	"Old crust is destroyed in the subduction process whereas new crust is created in ridges"
	"What are the differences between plate collisions in the Andes and those in the Himalaya?"	"The kind of plates involved in the collision: continental-continental *vs*. oceanic-continental. The processes in each collision: plates pushing each other *vs*. one plate

(*continued*)

Table 1. (*Continued*).

Test	Question	Accurate answer
		forming cracks in the other. The results of each collision: mountains *vs*. volcanoes"
Transfer	"Imagine that convection currents stop working, what would happen?"	"Plates would not move and, hence, would not crash into each other. As a consequence, there would be no mountains and volcanoes"
	"Imagine tectonic plates stop moving, how would you explain that?"	"For plates to stop moving, convection currents would have to stop working, which would be due to a reduction in the core's temperature"
	"Imagine that convection currents start moving at half their speed, how would you explain that?"	"It could be explained by stating that the core's temperature had decreased to half its original level"
	"What would you expect if ridges were small, isolated cracks in the Earth's surface?"	"Then magma would not surface through ridges, which would prevent the Earth's surface from being recycled"
	"What would you expect if the subduction process destroyed more crust than that created in ridges?"	"Crust would disappear, so the mantle would be exposed to the environment"
	"In the future, will there be more/less crust than now?"	"If crust is always being created and destroyed, there will be the same amount now as in the future"
	"Could the Himalaya range have volcanoes?"	"The Himalaya range could not have volcanoes because none of its plates have cracks, which is due to the kind of plates involved in this collision (continental–continental)"
	"Why are there both mountains and volcanoes in Italy but only mountains in Spain?"	"In Italy an oceanic plate and a continental plate are crashing into each other. Conversely, in Spain there are two continental plates"
	"Imagine eruptions in the Andes stop, how would you explain that?"	"One might explain this by saying that convection currents are not working, so plates are not pushing each other and, hence, the cracks through which magma surfaces are not formed"

Table 2. Means and standard deviations (in brackets) of all conditions in all variables.

	Human tutoring	Canned support	Control (no support)
Prior knowledge	4.03 (3.39)	2.87 (2.21)	3.07 (2.00)
Retention	2.30 (2.13)	0.83 (1.47)	1.21 (1.57)
Transfer	4.37 (3.24)	1.86 (1.88)	2.43 (2.52)

With regard to the retention test, there were significant differences between the conditions. This difference was reliable, as indicated by an analysis of variance, $F(2, 81) = 5.32$, $MSE = 3.02$, $p < .01$. Post-Hoc pair-wise comparisons revealed that the human tutoring condition was better than the canned support condition ($p < .01$) and marginally better than the control condition ($p < .08$). The sizes of these effects were large ($d = 0.80$) and medium ($d = 0.58$), respectively. The canned support and control conditions did not differ from each other ($p > .50$). Overall, this means that participants in the human tutoring condition were more able to recall the key concepts of the plate tectonics theory, as compared with those in the canned support and control conditions.

Regarding the transfer test, once again, there were significant differences between the conditions. An analysis of variance confirmed this, $F(2, 81) = 5.74$, $MSE = 8.33$, $p < .01$. *Post-hoc* pair-wise comparisons revealed that the human tutoring condition was better than both the canned support condition ($p < 0.01$) and the control condition ($p < 0.05$), which did not differ from each other ($p > 0.50$). The sizes of the effects were large ($d = 0.82$) and medium ($d = 0.67$), respectively. The result indicates that participants in the human tutoring condition were more able to apply the knowledge they had acquired to solve novel tasks, as compared with their counterparts.

3. Discussion

Learning from multimedia presentations requires processing a number of verbal and pictorial elements and interconnections within the limited capacity of the working memory. This circumstance makes it difficult to build coherent mental models and calls for the execution of the monitoring process, through which learners keep track of their emerging understanding. However, monitoring is difficult to execute. Because of this difficulty, learners are usually provided with support devices helping them in monitoring their understanding to achieve deep learning. A key question is how these devices should be presented. One approach, canned support, consists of inserting support devices into the multimedia presentation itself. Another one is to have human agents provide learners with support devices. In the light of its potential advantages (namely, the use of the spoken modality, the display of non-verbal cues and the inclusion of a social component), one might expect human tutoring to be better than canned support. However, these potential advantages might not be actual advantages, as indicated by the mixed results concerning their effects. In the experiment reported here, participants learned Geology from a multimedia presentation and received support in one of three forms: human tutoring, canned support, no support.

The results showed that participants in the human tutoring condition outperformed those in the canned support condition in both retention and transfer. This indicates that those participants receiving support devices from a human agent were more able to recall and use the key information presented in the multimedia material. According to this result, it may be interpreted that the advantages of human tutoring are actually realised. In other words, using the spoken modality, displaying non-verbal cues and incorporating a social component are beneficial when providing support devices.

The results in the present experiment extend those found by Azevedo et al. (2004). As was argued before, these researchers carried out an experiment very similar to that reported here but exhibiting one limitation. They provided learners with aids to multimedia learning in human and canned forms finding that the former was better than the latter. However, the support devices used in each condition were not strictly equal. Here, this variable was under control, making the interpretation of our results more accurate: it was the form of presentation, not the quantity and quality of the support devices that made the difference.

A question stimulated by our research is which of the advantages of human tutoring or what combination of them are the most significant? In other words, although the combination of spoken modality, non-verbal cues and human presence was effective, it is not clear what would happen if one or two of these elements were to be removed. Tentatively, it is the combination of the three elements that has a critical impact since the impact of either modality, non-verbal cues or human presence in isolation was unclear (see *1.2. Different ways of providing support devices*).

Another interesting finding in the present experiment was that the human tutoring condition outperformed the control condition. This can be interpreted as follows. On the one hand, as long as participants in the control condition exhibit lower levels of performance, it represents additional support for the idea that learning from multimedia can be difficult because of the problems learners have to overcome when doing so. On the other hand, it also means that, with the appropriate support, learners are able to achieve learning. Such appropriate support can be feedback to critical questions provided by a human agent, such as that used here.

There were no differences between the performances in the canned support and control conditions. This is important to the extent that learners are perhaps not able to learn deeply from multimedia presentations on their own, even when these include support devices. In order to explain why supporting devices inserted into the presentation were not helpful, the concept of *structural knowledge* (Goldman & Rakestraw, 2000; Sánchez & García, 2008) might be of help. In the field of text comprehension, this term refers to the ability of readers to recognise and use discourse markers (e.g. "in sum", "on the other hand") as cues guiding the construction of coherent representations from text. There are two assumptions behind the concept. First, learners acquire structural knowledge through their experience with instructional materials. Second, without such ability, readers cannot recognise and use the markers, being ineffective in their inclusion into texts. According to this framework, it is likely that participants in our experiment had low prior experience with the support devices used here, which prevented them from taking advantage of the devices, unless a human agent provided these.

On the practical side, based on the results reported here, it seems recommendable to provide learners with support via human tutoring (at least for agents, learners and materials similar to those used here). Accordingly, teachers asking students to learn from book-based or computer-based multimedia presentations (e.g. illustrated text, narrated animations in CD-Rom) might provide human assistance rather than rely on the support devices of the material itself. This would allow students to learn more deeply. The problem is that using humans requires more resources than using computers. Alternatively, computerised support devices might embody the advantages of human tutoring. However, if not only the spoken modality but human presence is needed to make support devices effective, it is not possible to do that. Future experiments exploring which of the elements or combination of elements is the critical one would clarify this question.

One limitation of the present experiment is that only one kind of support device was used. Corrective feedback on learners' responses to critical questions is one kind of aid but there is a wide range of support devices at our disposal. For instance, in aiding learners to revise their emerging mental representations, texts (Lorch, Lorch,

& Inman, 1993) or example-based instructional environments (Renkl, 2002) usually include elaborations clarifying critical concepts. As these elaborations can be complex (i.e. they comprise many ideas and interconnections between them), it is possible that a computerised version is better than human tutoring for their presentation, since the written modality allows readers to control the input (Sánchez & García Rodicio, 2008). Future research should explore whether the advantage of human tutoring holds for different kinds of aids. Other limitations are that we used a particular student population, a particular topic and a particular human agent. Specifically, we used high-school rather than older students, we used geology instead of biology or physics and we used a young woman tutor rather than an old/man tutor. So, the question remains open: would we find the same effects had we used a different population, topic or agent?

4. Conclusion

The supposed advantages of human support over other forms were not clear, so we conducted an experiment comparing human and canned forms of support. Human tutoring was better, indicating that its potential advantages had been realised. This makes progress on prior research, as the effect of the supposed advantages of human tutoring was mixed. Moreover, in practical terms, the finding means that if we want learners to profit from support devices, it is recommendable to have a human agent to provide them. Such a guideline may be of help for practitioners involved in the design of support mechanisms. An issue to explore in future research is which of the advantages of human tutoring or what combination of them is the critical one.

Acknowledgements

This research was partly supported by funding from Junta de Castilla y León y Fondo Social Europeo and Ministerio de Ciencia e Innovación de España (EDU2009-13077) awarded to the first and second author, respectively. The authors would like to thank Nadezhna Castellano for her valuable assistance in developing the materials and conducting the experiment.

Notes

1. In these studies, different instructional materials were used. This means that the cause of the misunderstandings was not their design but the intrinsic complexity of the topic.
2. Specifically, a rule was followed according to which, if the time spent on corrective feedback was less than 15 s, it was assumed that the participant had skipped it and it was left out of the experiment.

Notes on contributors

Héctor García Rodicio got his PhD in Educational Psychology from the University of Salamanca, Spain (2009, February). He is now teaching and pursuing his research at the University of Salamanca. His research interests include multimedia learning, text comprehension, self-regulated learning, the use of modality in computer-based environments and the design of instructional support.

Emilio Sánchez is Professor of Developmental and Educational Psychology at the University of Salamanca, Spain, where he has served since 1983. He is the author of several books and more than 50 articles in Spanish and international peer-reviewed journals. He is mainly interested in reading literacy.

References

Ainsworth, S., Bibby, P., & Wood, D. (2002). Examining the effects of different multiple representational systems in learning primary mathematics. *Journal of the Learning Sciences*, *11*, 25–61.

Atkinson, R.K. (2002). Optimizing learning from examples using animated pedagogical agents. *Journal of Educational Psychology*, *94*, 416–427.

Azevedo, R., Cromley, J.G., & Seibert, D. (2004). Does adaptive scaffolding facilitate students' ability to regulate their learning with hypermedia? *Contemporary Educational Psychology*, *29*, 344–370.

Azevedo, R., Guthrie, J.T., & Seibert, D. (2004). The role of self-regulated learning in fostering students' conceptual understanding of complex systems with hypermedia. *Journal of Educational Computing Research*, *30*, 87–111.

Baylor, A., & Ryu, J. (2003). Does the presence of image and animation enhance pedagogical agent persona? *Journal of Educational Computing Research*, *28*, 373–395.

Cain, K., Oakhill, J., & Bryant, P. (2004). Children's reading comprehension ability: Concurrent prediction by working memory, verbal ability, and component skills. *Journal of Educational Psychology*, *96*, 31–42.

Campbell, J., & Mayer, R.E. (2008). Questioning as an instructional method: Does it affect learning from lectures? *Applied Cognitive Psychology*, *23*, 747–759.

Chi, M.T.H. (2000). Self-explaining expository texts: The dual processes of generating inferences and repairing mental models. In R. Glaser (Ed.), *Advances in instructional psychology* (pp. 161–238). Mahwah, NJ: Lawrence Erbaum Associates.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Erlbaum.

Commander, N.E., & Stanwyck, D.J. (1997). Illusion of knowing in adult readers: Effects of reading skill and passage length. *Contemporary Educational Psychology*, *22*, 39–52.

Ekman, P., Friesen, W.V., O'Sullivan, M., Chan, A., Diacoyanni-Tarlatzis, I., Heider, K., ...Tzavaras, A. (1987). Universal and cultural differences in the judgement of facial expressions of emotion. *Journal of Personality and Social Psychology*, *53*, 712–717.

Ginns, P. (2005). Meta-analysis of the modality effect. *Learning and Instruction*, *15*, 313–331.

Goldman, S.R., & Rakestraw, J.A. (2000). Structural aspects of constructing meaning from text. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (pp. 311–335). Hillsdale, NJ: Lawrence Erlbaum Associates.

Graesser, A.C., Moreno, K., Marineau, J., Adcock, A., Olney, A., & Person, N. (2003). AutoTutor improves learning of computer literacy: Is it the dialog or the talking head? In U. Hoppe, F. Verdejo, & J. Kay (Eds.), *Proceedings of artificial intelligence in education* (pp. 47–54). Amsterdam: IOS Press.

Hacker, D.J. (1998). Self-regulated comprehension during normal reading. In D.J. Hacker, A.C. Graesser, & J. Dunlosky (Eds.), *Metacognition in educational theory and practice* (pp. 165–191). Mahwah, NJ: Lawrence Erlbaum Associates.

Hausmann, R.G.M., & Chi, M.T.H. (2002). Can a computer interface support self-explaining? *Cognitive Technology*, *7*, 4–14.

Kleinke, C., Staneski, R., & Berger, D. (1975). Evaluation of an interviewer as a function of interviewer gaze, reinforcement of subject gaze, and interviewer attractiveness. *Journal of Personality and Social Psychology*, *31*, 115–122.

Lorch, R.F., Lorch, E.P., & Inman, W.E. (1993). Effects of signaling topic structure on text recall. *Journal of Educational Psychology*, *85*, 281–290.

Mayer, R.E. (2001). *Multimedia learning*. New York: Cambridge University Press.

McNamara, D.S. (2004). SERT: Self-explanation reading training. *Discourse Processes*, *38*, 1–30.

Mikkilä-Erdmann, M. (2001). Improving conceptual change concerning photosynthesis through text design. *Learning and Instruction*, *11*, 241–257.

Moos, D.C., & Azevedo, R. (2009). Self-efficacy and prior domain knowledge: To what extent does monitoring mediate their relationship with hypermedia learning? *Metacognition and Learning*, *4*, 197–216.

Moreno, K., Mayer, R.E., Lester, H.A., & Spires, J.C. (2001). The case for social agency in computer-based teaching: Do students learn more deeply when they interact with animated pedagogical agents? *Cognition and Instruction*, *19*, 177–213.

Moreno, R., & Mayer, R.E. (2002). Learning science in virtual reality multimedia environments: Role of methods and media. *Journal of Educational Psychology*, *94*, 598–610.
Nass, C., Moon, Y., Fogg, B.J., Reeves, B., & Dryer, D.C. (1995). Can computer personalities be human personalities? *International Journal of Human–Computer Studies*, *43*, 223–239.
Otero, J. (2002). Noticing and fixing difficulties in understanding science texts. In J. Otero, J.A. León, & A.C. Graesser (Eds.), *The psychology of science text comprehension* (pp. 281–307). Mahwah, NJ: Lawrence Erlbaum Associates.
Otero, J., & Kintsch, W. (1992). Failures to detect contradictions in a text: What readers believe versus what they read. *Psychological Science*, *3*, 229–235.
Pollock, E., Chandler, P., & Sweller, J. (2002). Assimilating complex information. *Learning and Instruction*, *12*, 61–86.
Renkl, A. (2002). Learning from worked-out examples: Instructional explanations supplement self-explanations. *Learning and Instruction*, *12*, 529–556.
Sánchez, E., & García, J.R. (2008). The relation of knowledge of textual integration devices to expository text comprehension under different assessment conditions. *Reading and Writing*, *22*, 1081–1108.
Sánchez, E., & García Rodicio, H. (2008). The use of modality in the design of verbal aids in computer-based learning environments. *Interacting with Computers*, *20*, 545–561.
Sánchez, E., García Rodicio, H., & Acuña, S.R. (2009). Are instructional explanations more effective in the context of an impasse? *Instructional Science*, *37*, 537–563.
Schnotz, W. (2005). An integrated model of text and picture comprehension. In R.E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 49–69). Cambridge: Cambridge University Press.
Seufert, T., Schütze, M., & Brünken, R. (2009). Memory characteristics and modality in multimedia learning: An aptitude treatment interaction study. *Learning and Instruction*, *19*, 28–42.
Sweller, J., van Merriënboer, J.J.G., & Paas, F. (1998). Cognitive architecture and instructional design. *Educational Psychology Review*, *10*, 251–296.
Valenzeno, L., Alibali, M.W., & Klatzkya, R. (2003). Teachers gestures facilitate students learning: A lesson in symmetry. *Contemporary Educational Psychology*, *28*, 187–204.
Walters, K.L., & Walk, R.D. (1988). Perception of emotion from moving body cues in photographs. *Bulletin of the Psychonomic Society*, *26*, 112–114.