Journal of Teaching Language Skills (JTLS)
35(2), Summer 2016, ISSN: 2008-8191
pp. 123-160

On Anthropomorphism in Technology-Enhanced Language Learning: Does Modality Matter in Agent-Based Multimedia Instruction on L2 Idioms?

Rahman Sahragard
Professor
Shiraz University
rsahragard@rose.shirazu.ac.ir

Alireza Ahmadi
Associate Professor
Shiraz University
arahmadi@shirazu.ac.ir

Hamed Babaie Shalmani *
PhD Student
Shiraz University
babaie@iaurasht.ac.ir

Abstract

The present study aimed to satisfy a twofold purpose: On the one hand, it sought to verify the postulation that agent-based instruction could offer a compromise approach to teaching L2 idioms where form and meaning would be equally emphasized during instruction. Given that anthropomorphism has not been much under scrutiny, this research, on the other hand, sought to ascertain whether learning and retention of English idioms would be differentially impacted when two different modalities of virtual tutors —anthropomorphic and non-anthropomorphic— were present in the tutorial. To this aim, the participants of the study received instruction on 128 English idioms from human teachers, a multimedia application featuring a humanoid virtual teacher, or a piece of multimedia courseware with a non-anthropomorphic virtual tutor. Analysis of the post-intervention measures of L2 idiom knowledge revealed that agent-based instruction had proved more effective in improving both learning and retention of the target idioms among the participants. A further finding was that despite the greater motivational benefits of the humanoid virtual tutor, it had not privileged the participants, performance-wise.

Keywords: anthropomorphism, cognitive load, modality effect, pedagogical agents, redundancy effect

Received: 02/04/2016 Accepted: 11/07/2016

* Corresponding author

The non-compositionality of a great many idioms in English has always presented a serious challenge to L2 learners as they strive desperately to come to grip with the problem of mapping forms to meanings (Grant, 2003). The lack of appropriate teaching materials and the paucity of a clear methodology to help remove this stumbling block are among the major factors that have further aggravated the difficulty associated with the learning and retention of L2 idioms (Cameron & Low, 1999). Despite the fact that myriad strategies for efficient teaching of English idioms have been employed by ELT practitioners worldwide, none has fared better in ensuring optimal learning on the part of language learners largely owing to their excessive obsession with meaning-making and contextualization (Chen & Lai, 2013).

As one famous pedagogic strategy that places a strong emphasis on the role of context in L2 idiom acquisition, *associative fluency* (Johnson & Rosano, 1993) requires that students should find as many interpretations as possible for a single metaphor or an idiom by analyzing possible relationships or associations between a particular concept and the other concepts in the target language. As one example, in the idiomatic expression *he looks at the world through rose-tinted spectacles*, one hypothesis is that by analyzing the relationship between the color of a rose, a pair of spectacles, and the personality of the individual to whom the idiom applies, one can help make the idiomatic meaning transparent and hence fix it in the learner's memory; however, the central issue is that the mere transparency of meaning as revealed through the analysis of relationships would not necessary guarantee internalization of form as well (Abel, 2003).

Image formation (Paivio & Walsh, 1993) as yet another common strategy requires that learners should form mental images of the situations in which L2 idioms can be appropriately used. By allowing learners to generate novel images of situations, the efficiency of the search for relevant information in memory is enhanced, and the likelihood of efficient coding of idiomatic meanings is increased accordingly. Yet the strategy attracts the same criticisms as associative fluency: Mental representations can greatly aid comprehension of metaphoric expressions, in particular, of transparent idioms; nevertheless, form still receives scant attention and is hence prone to forgetting over time (Nippold & Duthie, 2003).

In a similar vein, *etymological elaboration* has been cited in the literature as one panacea for the problem of teaching L2 idioms. The rationale behind the use of this strategy comes from the idea that the meanings of idiomatic expressions are usually not arbitrary but motivated (Boers, Demecheleer, & Eyckmans, 2004). Students are presented with the history of origin or source domains of figurative expressions that are assumed to promote comprehension of idioms via the provision of elaborate verbal descriptions. Notwithstanding, while verbal elaborations provide an extra pathway for expediting the process of comprehension and enhancing recall, the strategy is not without its shortcomings and has often been reported not to have liven up to its claim to help decipher the meanings of culture-specific idioms and also to aid encode the L2 forms efficiently in the learner's memory (Vasiljevic, 2015).

Given the preoccupation of the aforementioned strategies with meaning, contextualization, and semantic analysis, and the familiar complaint ELT teachers receive from their students about their finding it challenging to make persistent form-meaning connections when it comes to the learning of figurative expressions, a concerted attempt needs to be made to develop a compromise approach that targets both profound comprehension and prolonged retention of L2 idioms on the part of students.

Over the past decade, advancements in computer technology have gained tremendous momentum such that they have inspired ELT enthusiasts to deliberate on what affordances technology could offer language pedagogy and hence to lay the foundation for the development of more innovative approaches such as Technology-Enhanced Language Learning (TELL), Multimedia Computer-Assisted Language Learning (MCALL), Mobile-Assisted Language Learning (MALL), and Podcasting (Babaie, 2008; Laborda, 2010).

From all technological breakthroughs, however, the design and implementation of animated pedagogical agents in multimedia learning environments represent a major landmark where the presence of virtual tutors in tutorials is assumed to enhance the quality of computer-human interaction, thereby augmenting learners' engagement with the learning task (Dunsworth & Atkinson, 2007). The claim gains support from the *social agency theory*

(Reeves & Nass, 1996) suggesting that emulation of human-to-human interaction in virtual learning environments, which is made possible via the incorporation of pedagogical agents into multimedia courseware, provides the necessary conditions for enhanced learning, as the learner finds himself immersed in a naturalistic environment that gives him the impression that the communication is actually taking place between two real humans (Heidig & Clarebout, 2011).

In addition to improving the naturalness of computer-human interaction through simulation of human daily communication, unique characteristics of virtual tutors likewise correlate with the effect they produce on student learning. It is contented that *lifelikeness* (i.e., looking like human beings and/or behaving like them) as one of these distinct qualities empowers pedagogical agents to coordinate speech with non-verbal forms of communication such that they appear as realistic, believable, and hence engaging teachers, which in turn further improves their effectiveness as attention grabbers in tutorial situations (Cassell & Thorisson, 1999). AutoTutor, Baldi, COSMO, Herman the Bug, and STEVE (Graesser, Person, & Harter, 2001; Massaro & Cohen, 1994) are but a few examples of pedagogical agents shown to have served learners by arousing their curiosity, garnering and directing their attention to the learning task being introduced, and generally, by generating strong motivation in them through performing a blend of roles such as “supplanting”, “scaffolding”, “demonstrating”, “modelling”, “coaching”, and “testing” at the time of instruction (Clarebout, Elen, Johnson, & Shaw, 2002, pp. 270-271).

More recent experiments have also lent support to the view that the attention-getting caliber of pedagogical agents is one major determinant of learners’ success with cognitive tasks. Yung and Pass (2015), for example, sought to examine how the presence of an animated pedagogical agent would affect student learning from an instructional animation on cardiovascular system. The tutor used in the study was capable of gesticulating and pointing at the time of instruction, thus engaging students’ attention and focusing their concentration on important concepts introduced. The study revealed that cuing by the pedagogical agent had had a positive effect on learning efficiency

thanks to the participants' increased ability to differentiate between relevant and redundant information presented at the time of instruction.

In yet another experiment by Dai, Raine, Roscoe, Cai, and McNamara (2011), the efficacy of an intelligent tutoring system called Writing-Pal was assessed concerning the development of essay writing skills among the participants of the study. The tutoring system comprised a number of strategy training modules that would provide witting strategy instruction on the three distinct phases of writing ability development, namely prewriting, drafting, and revising. Instruction was delivered in a virtual classroom where a virtual teacher called Dr. Julie established dialogues with other cartoon characters as they asked about the principles and mechanics of writing in each phase of development. Analysis of the students' essays as well as examination of their attitudes clearly suggested that the participants had found the pedagogical agent effective enough in engaging their attention throughout instruction and hence the whole learning experience enjoyable and memorable.

Carlotto and Jaques (2016) likewise explored how varying degrees of agent embodiment in tutorial settings would affect Brazilian students' learning of English in a CALL environment with embedded pedagogical gents. Attributes such as the image effect (i.e., the visual presence of a tutor can lead to more efficient learning); the modality effect (i.e., auditory information coupled with visual texts can facilitate information processing); and the embodiment effect (i.e., a fully embodied agent capable of gesturing and showing emotions and delivering the content aurally) were the object of investigation. The participants received treatment on the present perfect verb tense under a no-agent, a voice-only, a static-agent, or an embodied-agent condition. Analysis of the posttest scores as well as the students' responses to the questionnaire items revealed that the presence of a fully embodied agent had greatly privileged the participants, albeit the agent's voice was believed to have better contributed to the students' learning than its gestures and movements.

Whereas coordination of speech with facial expressions and body movements is one hallmark of pedagogical agents, not all virtual tutors feature the caliber to combine a wide range of human paralinguistic behaviors with verbal information, and accordingly, two distinct agent modalities, namely

anthropomorphic and *non-anthropomorphic*, can be identified. Anthropomorphic agents are those that resemble human beings and are able to gesticulate, wear various facial expressions, and generally display a vast array of human non-verbal forms of communication that are in synchrony with speech (Gulz & Haake, 2006). By comparison, non-anthropomorphic tutors are not humanoid in appearance and are able to show only a small subset of human non-verbal behaviors. An animated pointing finger, an arrow symbol with transition effects, a blinking rectangle, or flashing texts coupled with stunning visual effects are all characteristic examples of non-anthropomorphic agents (Murano, Gee, & O'Brian Holt, 2011).

Notwithstanding the fact that research exploring the contributions of animated pedagogical agents to student learning has been burgeoning since the advent of *intelligent tutoring systems* (Sklar & Richards, 2010), very few experiments have sought to compare the effects of different agent modalities on learning from agent-based multimedia instruction. Accordingly, whether *anthropomorphism* matters and the extent to which different modalities of virtual tutors produce comparable or differential effects on learning in multimedia environments has not been much under scrutiny. The studies conducted so far (e.g., Hongpaisanwiwat & Lewis, 2003; Yılmaz & Kılıç-Çakmak, 2012) have presented cogent arguments to support their preference for either the anthropomorphic or the non-anthropomorphic agent, or to show no preference for either of the two modalities; however, no plausible explanation has been provided as to why these studies have yielded contradictory findings. A dearth of comparative research to help settle the controversies surrounding the issue of anthropomorphism in agent-based instruction and hence the lack of convincing evidence suggesting the superiority of one agent modality over the other, then, should instigate studies examining the subject matter through the conduct of a closer inspection and a more detailed exploration.

Objectives of the Study

On the one hand, the present study aims to compare the effects of agent-based instruction with those of a conventional method (i.e., teacher-fronted instruction) on the learning and retention of English idioms among Iranian EFL learners. Given that studies (Allen, Crossley, Snow, & McNamara, 2014; Corbeil, 2007; Duffy & Azevedo, 2015; Ergül & Koç, 2013; Johnson, Rickel, & Lester, 2000; Noma & Badler, 1997) have shown that virtual tutors offer the potential to serve as powerful attention-getting devices that garner and focus the learners' attention on salient aspects of the learning task, it can be hypothesized that agent-based instruction may represent a compromise approach to language pedagogy such that while meaning is emphasized through the provision of contextual cues, definitions, and examples of use in the tutorial, equal importance is accorded to form by the virtual tutors increasing the saliency of L2 structures at the time of instruction. This can possibly be accomplished through increased noticing of L2 forms via promoting students' engagement with the subject matter being introduced and combining this with visual effects such as keeping an eye gaze on L2 forms, pointing to L2 structures, and so on. The present study, then, seeks to measure the efficiency of this balanced approach in comparison to that of a mainstream method of teaching L2 idioms as well as to test the tenability of the assumptions upon which it is premised.

On the other hand, given that little is known about whether anthropomorphism can be an issue in agent-based instruction, the overriding objective of the present study is to compare tutorial situations featuring two distinct modalities of pedagogical agents in an attempt to ascertain whether the learning outcome can be differentially impacted when an anthropomorphic and a non-anthropomorphic virtual tutor embark on teaching English idioms to different constellations of language learners. Considering that experiments investigating the role played by different pedagogical agents in furthering student learning have produced inconclusive results, the present study seeks to shed light on current conceptualizations and theorizing in the area of agent-based instruction by drawing on a mixed methods research design that combines quantitative and qualitative data to provide grounds for more informed judgments to be made on the overall

utility of pedagogical agents for language learning as well as the issue of agent modality in MCALL settings.

Research Questions

Given the above-mentioned objectives, the present study aims to address the following questions:

Q_1: Is there a statistically significant difference between the performance level of EFL learners who receive treatment through teacher-fronted instruction and that of those who receive treatment via agent-based instruction in regard to their learning and retention of English idioms?

Q_2: Is there a statistically significant difference between the performance level of EFL learners who receive instruction from an anthropomorphic pedagogical agent and that of those who receive instruction from a non-anthropomorphic virtual tutor in regard to their learning and retention of English idioms?

Q_3: In the views of the participants, what are seen as the pros and cons of learning from virtual tutors? Specifically, what is the promise, if any, of agent-based instruction and how it might be different from conventional methods as told by the participants?

Method

Participants

The participants were 150 students who were studying EFL at the Islamic Azad University of Rasht in the north of Iran. To homogenize the participants, care was exercised to recruit BA students who had already passed a course on English idioms. Student participation was voluntary and the participants would receive remuneration as their reward.

An IELTS proficiency test of receptive skills was administered to an initial pool of 287 EFL learners from whom students who had obtained an overall band score of 4.5 or 5 on the test were identified as intermediate-level learners and were hence entitled to take part in the study.

A digital randomizer called SuperCool Random Number Generator[1] was then employed to randomly select only 150 participants from the population

[1] http://www.supercoolbookmark.com/random/Download.aspx

of qualified candidates. The program came with an option to generate a set of random numbers from within a specified range. The prospective participants were first assigned a number from unity to a number representing the last individual identified as a qualified candidate. Next, the application randomly selected 150 numbers in such a way that all the qualified participants whose corresponding numbers were in the list of chosen numbers could be included in the final pool. The same randomizer was also utilized to randomly assign the final pool of the candidates to five equivalent groups of participants, with two serving as the pilot groups, two as the experimental groups, and one as the control group.

Materials

Out of 18 textbooks reviewed, five idiomatic texts were ultimately chosen as the base materials of the study. These involved (a) *All Clear! Idioms in Context Book 2* (Fragiadakis, 1992), (b) *Basic Idioms in American English Book 2* (Setzler, 1981), (c) *Practicing Idioms* (Watson, 1991), (d) *Speak English Like an American* (Gillett, 2004), and (e) *Street Talk 3* (Burke, 1995).

Authors' strong assertions about their books featuring high frequency English idioms and finding the selected idioms in the category of common idioms in idiom dictionaries such *American Heritage Dictionary of Idioms* (1997), *Cambridge International Dictionary of Idioms* (1998), *Collins COBUILD Dictionary of Idioms* (1995), and so on were two main standards used for the selection of the idiom texts as well as high frequency idioms.

Units with a limited number of exercises as well as units with recurring themes were excluded from selection. The final sample contained 128 idioms in 12 units of instruction chosen based on the suitability of the topic, neutrality of the content, as well as the diversity of the range of exercises offered.

Instruments

The proficiency test of receptive skills. For the learners to reap much benefit from instruction, it was essential that they should be able to read and listen fairly well in English. The rationale was that the participants would receive explanations of idiomatic meanings aurally, via listening to the virtual tutors, and visually, through their following on-screen texts. To this end, a

sample copy of an IELTS test taken from *Cambridge IELTS 9* (2013) examination papers was administered to the learners at the beginning of the experiment. The test comprised 80 items, half of which measured the students' listening ability, and the other half, their reading ability. The raw scores obtained on individual modules were calculated by computing the aggregates of the points the participants received on individual test items. These were then converted to band scores using an online IELTS band score calculator and logged by the researchers for later access.

The test of English idioms. In order to ascertain the participants' prior familiarity with the idiomatic expressions introduced over the course of the experiment and also to gauge their degree of learning from the course, two parallel versions of an English idiom test were developed, with one serving as a pretest and the other as immediate and delayed posttests of L2 idiom knowledge. Each version of the test also would measure receptive and productive knowledge of L2 idioms through recognition and production modules that initially contained 128 items in fill-in-the-blank, matching, and cloze formats.

The focus group interviews. At the end of the study, a total of 15 participants were randomly chosen from each experimental group and were interviewed in an attempt to share their feelings and also their hands-on experience with the researchers. Specifically, the probes aimed to poll the participants' opinions as to their overall impression of working with the multimedia applications, the efficacy of agent-based instruction, what they liked most about the virtual tutors, the amount of control they could exercise over the multimedia units, the adequacy of explanations and examples presented, the utility of feedback given on their responses by the pedagogical agents, and so on.

The multimedia applications. Two pieces of multimedia courseware, each drawing on one agent modality (i.e., anthropomorphic or non-anthropomorphic), were authored by one of the researchers to serve as instructional instruments. The multimedia application with an embedded anthropomorphic agent capitalized on a humanoid virtual tutor that was capable of coordinating speech with a vast range of human paralinguistic behaviors such as gesticulating, using various hand gestures (e.g., pointing,

waving, etc.), making eye contact with students, giving a nod, fixing a steady gaze on L2 forms, and so on. It also featured the caliber to wear various facial expressions and communicate emotions. By comparison, the other type of courseware featured a non-anthropomorphic virtual tutor in the form of an animated hand with a pointing finger. The non-humanoid tutor, however, was capable of demonstrating only a small subset of human non-verbal behaviors that is, pointing to the idiomatic forms at the time of instruction. Both types of applications, however, comprised 12 units of instruction that were computerized versions of the units chosen based on the aforementioned selection criteria. They also introduced 128 English idioms to the experimental groups through several transitional modules.

Procedures

At the beginning of the experiment, each pilot group received one of the two versions of the test of L2 idioms that underwent standardization through several rounds of analysis. The participants took the production test first and sat the test of recognition after a short break. Their responses to individual items were scored and then entered as binary data into a digital item statistics analyzer called Test Analysis Program[2] featuring the capability to mark defective items with the number (hash) sign.

It was decided that items with item facility values close to zero be kept and easy items be removed from the final drafts of the tests, as easy items would indicate that the participants already knew the idioms. The exclusion of easy items, however, made the scores non-comparable across the tests, as some modules were left with more items and some with fewer ones. Likewise, some items were of different types, measuring the students' knowledge of different idioms. To remedy the problem, then, a digital file and directory comparison tool called ExamDiff Pro[3] was employed. The comparator allowed for the cross-comparison of the tests drafts by highlighting extra items and those not shared by all four modules in different colors.

Following the exclusion of extra and mismatched items, too, the participants' exam papers were rescored and the reliability coefficients of the

[2] http://www.ohio.edu/people/brooksg/downloads/tap.exe
[3] http://www.prestosoft.com/edp_examdiffpro.asp

tests were computed through the use of a Cronbach's alpha. With 84 items (out of 128) left on the tests modules, the indices of reliability reported for the pretest of production was 0.91; for the pretest of recognition, it was reported as 0.8; for the posttest of production, it turned out to be 0.86; and for the posttest of recognition, it was reported as 0.84.

As for the content validity of the tests, the remaining test items were checked by the researchers to ensure that they would still measure the knowledge of a representative sample of the idioms that appeared in the selected units. Next, a Monte Carlo simulation was conducted to demonstrate the tests construct validity. The simulation could help provide a dependable estimate of the probability of recovering a specified number of factors when the number of iterations was set to 1000. It worked by comparing the eigenvalues derived from the real dataset (i.e., by analyzing the participants' test performance scores) with those obtained for a randomly generated set of data (Pallant, 2013).

The simulation results revealed that three factors could be recovered for all four modules, suggesting that the three sections of the tests (i.e., fill-in-the-blank, matching, and cloze) might have measured three logical aspects of the L2 idiom knowledge and hence the high chance of recovery for three factors. The small number of participants in the pilot groups could be another explanation for the possibility of test items correlating with more than one single factor (Bryant & Yarnold, 1995).

Following the standardization of the idiom tests, one test was administered to the experimental and control groups as a pretest so as to demonstrate the participants' prior familiarity, if any, with the target idioms to be taught as well as the homogeneity of the study groups at the beginning of the experiment.

The learners' responses to the test items on both recognition and production modules were binary scored, that is, they would receive one mark for correct answers and the total score possible would be 84. As for the production tests, the participants would receive points only when they supplied the blanks in the stems with the target idioms in full, or when they produced a matching target idiom for an underlined phrase in the stems. Incomplete productions would receive no mark.

The analysis of the pretest results revealed that the participants demonstrated a lackluster performance on both pretests of receptive and productive knowledge ($M_{R\ anthropomorphic}$ = 16.16, SD = 4.14; $M_{P\ anthropomorphic}$ = 7.26, SD = 3.54; $M_{R\ non\text{-}anthropomorphic}$ = 14.90, SD = 3.65; $M_{P\ non\text{-}anthropomorphic}$ = 6.10, SD = 3.25; $M_{R\ control}$ = 15.03, SD = 4.53; $M_{P\ control}$ = 7.46, SD = 2.95), suggesting that they would need to learn a great number of the target idioms. A further finding was that the three groups belonged to the same population prior to the delivery of instruction ($p > 0.05$).

Having taken the pretest, the participants in the experimental groups began receiving treatment on 128 English idioms through 12 agent-based multimedia units. They sat at computer terminals, wore headsets, and launched the applications at the researchers' signal. Next, the participants watched an intro movie and were redirected to the Login Page a while later to enter their personal information. Entering all details was essential for successful logging into the multimedia courseware.

Once the participants specified their names, e-mail addresses, and other required data, they were taken to the Orientation Module where a humanoid agent or a background voice zeroed in on the importance of learning idioms, the themes under which the idioms of the unit would be presented, what different buttons would do when interacted by the participants, and so on.

Next, both experimental groups watched an animation of two or more virtual characters establishing dialogues about a subject matter and presenting the target idioms under a relevant theme. Figures 1 and 2 below show the images of virtual characters under the two experimental conditions where they set the scene for contextualized presentation of the target idioms.

Figure 1. The virtual characters under the anthropomorphic condition. Dialogues were established and the target idioms were introduced under a relevant theme to allow for the contextualized learning of L2 idioms.

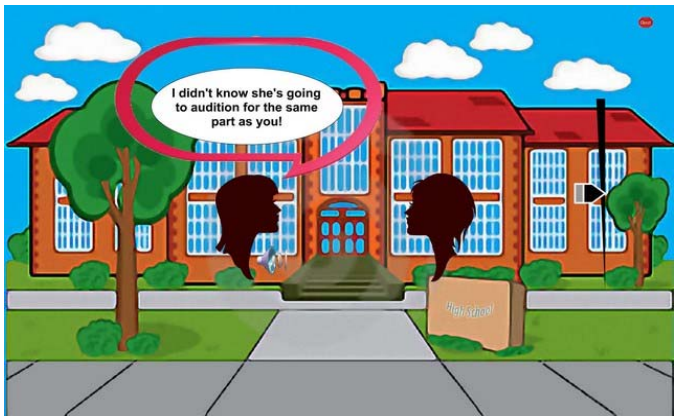


Figure 2. The virtual characters under the non-anthropomorphic condition. While the anthropomorphic group watched animated virtual characters conversing, the non-anthropomorphic group viewed static images of talking silhouettes that were used to signal turn-taking only.

Having watched the animations of virtual characters and ventured a wild guess at the idiomatic meanings, the participants moved on to have their conjectures checked by the applications. Here the anthropomorphic tutor or the background voice started posing questions to the participants and responding in an appropriate manner depending on their choices. The questions were of the multiple-choice type and appeared on a virtual blackboard in the background. The options were selectable from a list of choices that came with radio buttons.

The humanoid teacher would praise the participants by giving them a round of rapturous applause in case they responded correctly to the comprehension checks, or express his regret when the selected option was wrong. As for the non-anthropomorphic group, only verbal feedback was given to inform the participants of the correctness of their choices.

Next, the pedagogical agents embarked on furnishing the participants with pertinent definitions, explanations, and examples of use in the Instruction Module. For the anthropomorphic group, the humanoid tutor behaved in much

the same way as a human teacher: He coordinated body movements with speech and demonstrated a wide range of non-verbal behaviors at the time of instruction. He spoke slowly and eloquently, allowing the participants to deliberate on the content delivered.

As for the non-anthropomorphic condition, an animated hand with a pointing finger would pop up at the time of instruction. The pointing finger appeared at different positions (sometimes blinking and sometimes moving in different directions) when pointing to and highlighting the idiomatic forms displayed by on-screen texts; however, it was not capable of wearing facial expressions or communicating emotions. Figures 3 and 4 below show the difference between the two multimedia conditions under which instruction was delivered by the two agent modalities.

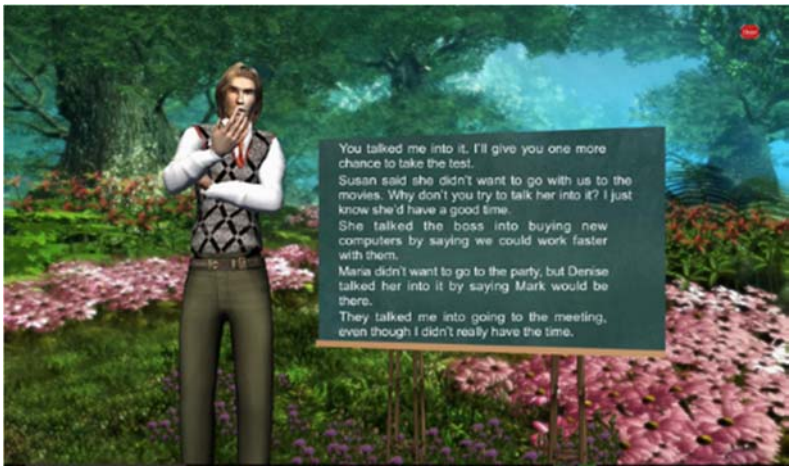


Figure 3. Humanoid tutor in the Instruction Module. Jack is gesticulating as he explains the idiomatic meanings and presents examples showing the use of the target idioms in appropriate contexts of use.

Figure 4. Pointing finger in the Instruction Module. To highlight the idiomatic expressions being introduced, the pointing finger was animated with visual effects such as blinking and moving effects. Only a small subset of human paralinguistic behaviors (i.e., pointing) could be displayed, however.

At the end of instruction, the participants were given three options to choose from: They could opt for rewinding the whole movie clip; they could listen to individual segments of instruction in the Individual-Parts Scene by interacting with the corresponding multimedia buttons; or they could move on to work with the exercises in the Exercise Module.

The Exercise Module aimed to provide the learners with further practice in manipulating the newly learnt target idioms in different contexts of use. One characteristic feature of the module was that it would supply the participants with feedback by item response, that is, they would receive feedback as soon as they answered every single question.

From the Exercise Module, the participants were taken to the Round-Up Module where they quickly reviewed the idioms they had already learnt jointly with the tutor; and from there, they were taken to the Evaluation Module to show their overall learning from agent-based instruction. There were a total of 10 multiple-choice questions that the participants would have to answer within the time limit. Each item could be attempted only once, however, and the provision of feedback would be delayed until the end of the quiz.

In antithesis to the experimental groups who worked with multimedia courseware with embedded pedagogical agents, the control group received treatment on the target idioms the conventional way: At the beginning, the dialogues included in the print copies of the selected units were read aloud to the participants to allow them to formulate conjectures about idiomatic meanings. Next, they were presented with 10 to 12 comprehension check questions to confirm their hunches on a voluntary basis. Having their surmises confirmed, the participants listened to the same dialogues again as they were being read by the researchers, this time paying heed to the definitions, explanations, and examples provided.

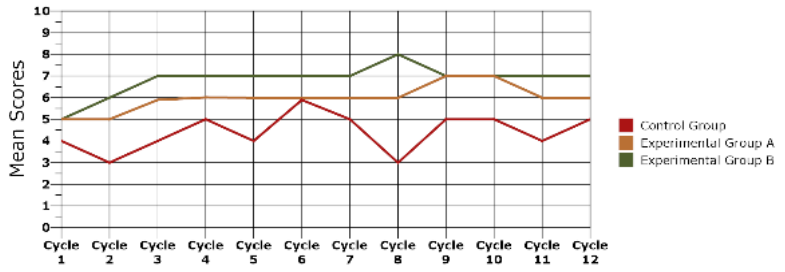
The participants then practiced using the recently introduced idioms in new contexts of use by trying the exercises and listening to the researchers' feedback. The round-up session was much akin to its counterpart under the experimental conditions: The idiomatic expressions were reintroduced through new example sentences that were followed by a brief explanation of idiomatic meanings. Finally, the participants took a paper quiz to demonstrate their overall learning from instruction.

The experiment lasted for six weeks during which treatment was offered for one session per week and for 90 to 100 minutes per session. Each session was divided into two cycles, and in each cycle, one unit of instruction available in either a print or a digital copy was taught. At the end of the experiment, the participants took a different version of the standardized test of idioms as an immediate posttest and again as a delayed posttest two weeks after the study. Focus group interviews were also conducted with half of the participants in the experimental groups.

Results

To gauge the amount of learning that occurred as a result of the treatment effect over the course of the experiment and to identify potential developmental patterns in the participants' performance data, mean scores obtained by the three study groups on the quizzes given at the end of each instructional cycle as well as those obtained on the immediate posttests of recognition and production were retrieved and analyzed. Figure 5 below

shows the plot of mean scores recorded for individual cycles in each treatment session.



	Cycle 1	Cycle 2	Cycle 3	Cycle 4	Cycle 5	Cycle 6	Cycle 7	Cycle 8	Cycle 9	Cycle 10	Cycle 11	Cycle 12
Control Group	4	3	4	5	4	6	5	3	5	5	4	5
Experimental Group A	5	5	6	6	6	6	6	6	7	7	6	6
Experimental Group B	5	6	7	7	7	7	7	8	7	7	7	7

Figure 5. Patterns of development. The experimental groups' gains showed no long-term, unsystematic fluctuation in their level of performance over the course of the study, implying that learning had proceeded more smoothly for these groups of participants than for the control group students.

The total possible score that could have been obtained by the participants in each cycle was seven on average, given that the command of 84 idioms in total should have been measured over 12 cycles of continuous instruction. As can be seen in the figure, the mean score logged for the experimental groups in each cycle was higher than that obtained by the control group learners. This clearly suggests that instruction as delivered by the pedagogical agents had proved comparatively more efficient in expanding the participants' repertoire of the target idioms. Further evidence supporting this claim comes from the fact that the mean scores of the experimental groups displayed small fluctuations and almost remained fixed across the cycles, whereas long-term variation in the scores was observed for the participants of the control group across the treatment sessions.

While the scores obtained by the participants in individual cycles were indicative of the amount of learning that occurred as a result of the treatment they had received on a limited number of idioms taught, the mean scores obtained on the immediate posttests could help show their overall learning from instruction as delivered over the course of the experiment. Tables 1

through 4 below summarize the results of the immediate posttests of receptive and productive knowledge.

Table 1

Means on the Immediate Posttest of Recognition

Descriptive Statistics	Target Groups	Mean	Std. Deviation	N
Mean Scores	Experimental Group A	74.2000	8.71938	30
	Experimental Group B	80.6000	8.37319	30
	Control Group	64.0333	11.03125	30
	Total	72.9444	11.58977	90

Table 2

Results of Tests of Within- and Between-subjects Contrasts

Tests of Within-Subjects Effects		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Treatment	Pillai's Trace	.966	2489.914	1.000	87.000	.000	.966
	Wilks' Lambda	.034	2489.914	1.000	87.000	.000	.966
	Hotelling's Trace	28.620	2489.914	1.000	87.000	.000	.966
	Roy's Largest Root	28.620	2489.914	1.000	87.000	.000	.966
Treatment * Target Groups	Pillai's Trace	.287	17.494	2.000	87.000	.000	.287
	Wilks' Lambda	.713	17.494	2.000	87.000	.000	.287
	Hotelling's Trace	.402	17.494	2.000	87.000	.000	.287
	Roy's Largest Root	.402	17.494	2.000	87.000	.000	.287

Tests of Between-Subjects Effects	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	350948.356	1	350948.356	7564.729	.000	.989
Target Groups	2120.478	2	1060.239	22.854	.000	.344
Error	4036.167	87	46.393			

Scheffé's Test	(J) Target		Std. Error	Sig.	95% Confidence Interval

(I) Target Groups	Groups	Mean Difference (I-J)			Lower Bound	Upper Bound
A	B	-6.4000*	2.43961	.036	-12.4759	-.3241
	C	10.1667*	2.43961	.000	4.0908	16.2425
B	A	6.4000*	2.43961	.036	.3241	12.4759
	C	16.5667*	2.43961	.000	10.4908	22.6425
C	A	-10.1667*	2.43961	.000	-16.2425	-4.0908
	B	-16.5667*	2.43961	.000	-22.6425	-10.4908

Note. Experimental Group A = Anthropomorphic Group, Experimental Group B = Non-Anthropomorphic Group, & Control Group = Teacher-Fronted Instruction. The within- and between-subjects contrasts are both statistically significant ($p < 0.05$) and meaningful as suggested by large Eta Squared values. In addition, the claim that the treatment produced differential impacts on student learning under all the three conditions is substantiated by the Scheffé's test result ($p < 0.05$).

Table 3

Means on the Immediate Posttest of Production

Descriptive Statistics	Target Groups	Mean	Std. Deviation	N
Mean Scores	Total	6.9444	3.27867	90
	Experimental Group A	72.1000	7.97561	30
	Experimental Group B	77.2000	5.40370	30
	Control Group	60.5667	8.83443	30
	Total	69.9556	10.22874	90

Table 4

Results of Tests of Within- and Between-Subjects Contrasts

Tests of Within-Subjects Effects		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Treatment	Pillai's Trace	.983	4990.433	1.000	87.000	.000	.983
	Wilks' Lambda	.017	4990.433	1.000	87.000	.000	.983
	Hotelling's Trace	57.361	4990.433	1.000	87.000	.000	.983
	Roy's Largest Root	57.361	4990.433	1.000	87.000	.000	.983
	Pillai's Trace	.446	34.980	2.000	87.000	.000	.446

Treatment * Target Groups	Wilks' Lambda	.554	34.980	2.000	87.000	.000	.446
	Hotelling's Trace	.804	34.980	2.000	87.000	.000	.446
	Roy's Largest Root	.804	34.980	2.000	87.000	.000	.446

Tests of Between-Subjects Effects	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	266112.450	1	266112.450	8375.834	.000	.990
Target Groups	1884.933	2	942.467	29.664	.000	.405
Error	2764.117	87	31.771			

Scheffé's Test (I) Target Groups	(J) Target Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
A	B	-5.1000*	1.94855	.037	-9.9529	-.2471
	C	11.5333*	1.94855	.000	6.6805	16.3862
B	A	5.1000*	1.94855	.037	.2471	9.9529
	C	16.6333*	1.94855	.000	11.7805	21.4862
C	A	-11.5333*	1.94855	.000	-16.3862	-6.6805
	B	-16.6333*	1.94855	.000	-21.4862	-1.7805

Note. Both within- and between-subjects contrasts are statistically significant at the preset alpha level ($p < 0.05$) and are also meaningful as indicated by large Eta Squared values. In addition, the treatment effects produced under all the three conditions are not equivalent in size as suggested by the Scheffé's test result ($p < 0.05$).

As can be seen in the tables, the mean scores reflect a marked improvement in the participants' receptive and productive knowledge of the target idioms over the course of the study in comparison to their lackluster performance prior to their receiving treatment. This clearly suggests that the three modalities of instruction had proved effective in improving the learners' test performance scores across the pre- and post-intervention measures of L2 idiom knowledge. Such a claim gains support from the results of the tests of within-subjects effects, in particular, from that of Wilk's Lambda (Pallant, 20103), the most robust and reliable measure of all, suggesting that the within–group difference is statistically significant at the preset alpha level (p

< 0.05). Large Eta Squared values ($\eta > 0.14$) provide further corroborative evidence for this assumption. A further finding is that the interaction effect (Treatment * Target Groups) suggests that there was no similar change in the participants' test scores across the pre- and posttests. In other words, the treatment effects were not similar for the learners of different groups.

A glimpse at the statistics for the between-subjects contrasts as well as the Scheffé's test result reveals that the group effect is also statistically significant ($p < 0.05$). This suggests that knowledge of the target idioms did not improve equally well for the participants in the three study groups. In other words, it can be contended that the treatments had produced differential effects on student learning, resulting in varying degrees of learning success. Furthermore, as can be seen, the mean difference is significant in favor of both the experimental groups and the experimental group who had received treatment under the non-anthropomorphic condition. This strongly suggests the greater efficiency of agent-based instruction relative to that of teacher-fronted instruction as well as the superiority of the non-anthropomorphic condition when it comes to the learning of L2 idioms.

Likewise, in an attempt to measure the effect, if any, of the treatment on the participants' retention of L2 idioms, the mean scores obtained by the three study groups on the delayed posttests of recognition and production were retrieved and analyzed. Tables 5 through 8 below summarize the results of the delayed posttests of receptive and productive knowledge.

Table 5

Means on the Delayed Posttest of Recognition

Descriptive Statistics	Target Groups	Mean	Std. Deviation	N
Mean Scores	Experimental Group A	72.0333	8.51969	30
	Experimental Group B	78.5333	8.62927	30
	Control Group	61.3667	10.29390	30
	Total	70.6444	11.53628	90

Table 6

Results of Tests of Within- and Between-Subjects Effects

Tests of Within-Subjects Effects		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.295	36.375	1.000	87.000	.000	.295
	Wilks' Lambda	.705	36.375	1.000	87.000	.000	.295
	Hotelling's Trace	.418	36.375	1.000	87.000	.000	.295
	Roy's Largest Root	.418	36.375	1.000	87.000	.000	.295
Time * Target Groups	Pillai's Trace	.005	.237	2.000	87.000	.790	.005
	Wilks' Lambda	.995	.237	2.000	87.000	.790	.005
	Hotelling's Trace	.005	.237	2.000	87.000	.790	.005
	Roy's Largest Root	.005	.237	2.000	87.000	.790	.005

Tests of Between-Subjects Effects	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	927799.606	1	927799.606	5553.387	.000	.985
Target Groups	8691.878	2	4345.939	26.013	.000	.374
Error	14535.017	87	167.069			

Scheffé's Test (I) Target Groups	(J) Target Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
A	B	-6.5000*	2.37119	.027	-12.4054	-.5946
	C	10.6667*	2.37119	.000	4.7612	16.5721
B	A	6.5000*	2.37119	.027	.5946	12.4054
	C	17.1667*	2.37119	.000	11.2612	23.0721
C	A	-10.6667*	2.37119	.000	-16.5721	-4.7612
	B	-17.1667*	2.37119	.000	-23.0721	-11.2612

Note. Whereas there was a considerable change in test performance scores over time as indicated by the main effect for "Time" ($p < 0.05$), the decline in performance was similar across the study groups as suggested by an insignificant interaction effect (Time * Target Groups, $p > 0.05$). The between-subjects effects are also sizable, suggesting that the differential effects of the treatment under the three conditions were kept throughout time ($p < 0.05$).

Table 7

Means on the Delayed Posttest of Production

Descriptive Statistics	Target Groups	Mean	Std. Deviation	N
Mean Scores	Experimental Group A	70.4000	6.34415	30
	Experimental Group B	75.2000	5.99080	30
	Control Group	57.4333	6.61077	30
	Total	67.6778	9.79776	90

Table 8

Results of Tests of Within- and Between-Subjects Effects

Tests of Within-Subjects Effects		Value	F	Hypothesis df	Error df	Sig.	Partial Eta Squared
Time	Pillai's Trace	.229	25.846	1.000	87.000	.000	.229
	Wilks' Lambda	.771	25.846	1.000	87.000	.000	.229
	Hotelling's Trace	.297	25.846	1.000	87.000	.000	.229
	Roy's Largest Root	.297	25.846	1.000	87.000	.000	.229
Time * Target Groups	Pillai's Trace	.021	.949	2.000	87.000	.391	.021
	Wilks' Lambda	.979	.949	2.000	87.000	.391	.021
	Hotelling's Trace	.022	.949	2.000	87.000	.391	.021
	Roy's Largest Root	.022	.949	2.000	87.000	.391	.021

Tests of Between-Subjects Effects	Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	852432.050	1	852432.050	9701.490	.000	.991
Target Groups	9408.100	2	4704.050	53.537	.000	.552
Error	7644.350	87	87.866			

Scheffé's Test (I) Target Groups	(J) Target Groups	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
A	B	-4.8000*	1.63191	.016	-8.8643	-.7357
	C	12.9667*	1.63191	.000	8.9024	17.0309
B	A	4.8000*	1.63191	.016	.7357	8.8643
	C	17.7667*	1.63191	.000	13.7024	21.8309
C	A	-12.9667*	1.63191	.000	-17.0309	-8.9024
	B	-17.7667*	1.63191	.000	-21.8309	-13.7024

Note. The main effect for "Time" suggests that the change in test performance scores over time was noticeable for the participants in each study group ($p < 0.05$). Yet the interaction effect (Time * Target Groups) is not significant ($p > 0.05$), suggesting that the effect of time was similar across the study groups. In addition, the tests of between-subjects contrasts clearly show that the treatments had produced differential effects on the participants' retention ($p < 0.05$).

As can be seen, the statistics reported for the tests of within-subjects effects clearly shows that the main effect for "Time" is statistically significant ($p < 0.05$) and meaningful at the same time as indicated by large Eta Squared values ($\eta > 0.14$). The implication is that there was a noticeable change in the test scores over time for the participants in individual groups. Notwithstanding, the interaction effect (Time * Target Groups) is not significant ($p > 0.05$) and at the same time is too small to be meaningful ($\eta < 0.06$). This suggests that the decline in performance and hence the drop in test scores was similar for the three groups, or there was the same change in test performance scores over time for all participants.

Analysis of the between-subjects effects likewise suggests that the group effect is statistically significant ($p < 0.05$). This shows that the amount of retention varied for the participants of different groups—an assumption that receives further support from the Scheffé's test result ($p < 0.05$). In other words, it can be argued that the treatments had kept their effects over time and hence the groups' differential performance on the delayed posttests, too. A further glimpse at the tables also reveals that here again the mean difference is significant in favor of the experimental treatments and that the non-anthropomorphic condition is still superior to the anthropomorphic condition in terms of instructional efficiency.

Discussion

In this section, the main findings of the study are discussed with reference to the three research questions addressed. The overarching qualitative question is divided into ancillary probes that are answered along with the first two quantitative questions. When analyzing transcribed interviews to poll the participants' opinions about the key aspects of agent-based instruction, the researchers drew on a model of inductive data analysis

by Strauss and Corbin (1998) suggesting that qualitative data analysis involves codification of data fragments in three distinct phases, namely "open", "axial", and "selective" coding (pp. 101-103). A specialized qualitative data analysis tool called MAXQDA[4] was also employed to expedite the arduous process of data analysis.

A brief review of the results presented in the previous section suggests that agent-based instruction had proved comparatively more effective in developing both receptive and productive knowledge of the learners, which lends support to the assumption that this modality of instruction represents a compromise approach to teaching English idioms such that both form and meaning are emphasized equally well during instruction. This finding provides an empirically justified answer to the first research question asking about the comparability of the effects of agent-based instruction with those of a conventional method of teaching English idioms on student learning and retention. There are two explanations for the present finding:

One justification is that learning from pedagogical agents offered the participants a new, enjoyable experience where "persona effect" (Ryu & Baylor, 2005, p. 291), produced as a result of virtual tutors mimicking human behaviors, increased their lifelikeness and hence their believability such that it created the impression that the participants were actually communicating with real human teachers (van der Meij, van der Meij, & Harmsen, 2015; Veletsianos & Miller, 2008). Accordingly, enhanced naturalness of computer-human interaction can be a possible explanation for the students' sustained motivation throughout instruction and hence their augmented learning from virtual tutors.

The persona effect, lifelikeness, and believability of pedagogical agents, which make them powerful attention gabbers such that student motivation is likely to sustain longer at the time of instruction, in turn seem to stem from their caliber to draw on social cues to establish rapport and maintain interpersonal relationships with learners through, for example, making eye contact, flashing a smile, giving a nod, or offering praise, and visual cues where, for instance, pointing to or keeping a steady gaze on L2 structures, using a special hand gesture, or displaying a flashing effect is used to capture and direct learners' attention to the learning task (Johnson, Ozogul, & Reisslein, 2014). The pedagogical agents in the present study were capable of

[4] http://www.maxqda.com/

using one or both types of cues and hence were able to directly (via visual signaling) or indirectly (via social cues) capture and sustain the participants' attention throughout instruction.

When asked about the overall utility of agent-based instruction, what they liked most about the applications, and which features they found the most useful, a large percentage of the participants in both experimental groups (92% in the anthropomorphic and 86% in the non-anthropomorphic group) reported that they had greatly enjoyed working with the multimedia units and that they had found the whole experience engaging and fruitful in improving their knowledge of L2 idioms. One participant, for example, described his hands-on experience in the following way:

Vignette 1: What I found truly interesting in the applications was the way the agent greeted me and introduced himself like a human. He then explained everything in detail: He talked about the goals of the unit, the functions of the buttons… His teaching was perfect, the clarifications were great, and there were a lot of exercises to do…

In addition to improved quality of computer-human interaction adding to the overall novelty of experience, increased levels of motivation, as well as enhanced noticing of L2 forms being a possible explanation for the greater sufficiency of agent-based instruction, another justification is that instruction as delivered via pedagogical agents is more or less learner-centered where learners feel to be in control of aspects of the environment and hence their learning. Such a claim receives support from a number of studies (e.g., Collentine, 2011; Heift, 2007; Toyoda, 2001) that have explored issues of *learner autonomy*, *learner-controlled interaction*, and *individualized instruction* or *adaptive CALL* in CALL environments. It has been argued that some degree of control is implicit in most tutorial situations such that parameters like the amount of tutor's talk, the number of clarification requests, the provision of feedback on responses, the amount of time to be spent on doing various learning tasks, what to do next and how to do it, and so on can be controlled, allowing learners to explore the environment and exploit the available resources within a heuristic-learning framework (Chapelle & Mizuno, 1989).

In the present study, the participants were able to be in control of part of their learning by, for example, enjoying complete freedom to skip over the Rewind/Individual Parts Scene to do the exercises, or to rewind parts of instruction they believed they had not completely or carefully listened to, or

to receive feedback based on their unique responses to questions. This contrasts with the control condition where the learning environment was more under the rein of the human teachers, and the clarifications and feedback were typically given based on group performance or on voluntary requests by some participants. When asked about the extent to which they felt to be in control of their learning, again a sizable percentage of the participants (72% in the anthropomorphic and 74% in the non-anthropomorphic group) reported that they could partially control the pace of learning and that they had found the whole experience to be partly learner-directed and individualized. The following vignette represents one participant's view on her being in control of aspects of instruction:

Vignette 2: I believe I could somehow proceed at my own speed when working with the units. For instance, I could watch the virtual teacher several times to listen to the explanations if I felt I forgot the meanings of some idioms; or I could choose to get feedback…

A further review of the findings under the results section reveals that the non-anthropomorphic group delivered a superior performance relative to even the anthropomorphic group, suggesting that the two modalities of pedagogical agents had produced differential effects on the students' learning and retention of the target idioms. This finding provides an empirically justified answer to the second research question exploring whether anthropomorphism would be a covariate, moderating the effects of virtual tutors. Mayer and Moreno's (1998) Generative/Cognitive Theory of Multimedia Learning and Sweller's (2005) Cognitive Load Theory are time-honored, viable theories of multimedia learning that bear strong relevance for the interpretation of this second finding.

Out of 10 instructional design principles proposed by Moreno (2006) and her associates as underpinnings of the Generative Theory of Multimedia Learning, two principles—the *modality* and *redundancy* principles—are key to an understanding of whether and how anthropomorphism affects learning from virtual tutors. The modality principle suggests that when learners are presented with relevant materials that are available in multiple modalities (e.g., text, picture, audio, etc.), processing is facilitated, as different memory channels are engaged. A strong modality effect, then, is produced that in turn maximizes the learning efficiency.

Notwithstanding, when redundant information is available in one modality, the promising modality effect should vie with the harmful

redundancy effect to compensate for the limited capacity of the working memory when engaged in information processing. This hinges on the amount of extra information that is available in one modality such that the greater the amount of redundant information, the weaker the modality effect appears in the tutorial setting. For example, it has been suggested that learners can acquire knowledge better from graphics and narration (spoken texts) than from graphics and visual texts (Mayer, 2009). The rationale is that graphics and visual texts are both processed in the visual memory, resulting in a high cognitive load.

The redundancy principle lends support to Sweller's (2005) Cognitive Load Theory suggesting that instruction should be delivered in such a way as to reduce the *extrinsic cognitive load* that results from an inefficient instructional design where redundant materials overload the working memory capacity at the time of information processing. Convictions are strong that a heightened level of extrinsic load produces what is technically referred to as the *split-attention effect* (Ayres & Sweller, 2005) where the overloaded memory module (i.e., verbal or visual memory) compensates for its limited capacity by forcing learners to divide their attention between different bits of information that are to be processed simultaneously than sequentially for active construction of knowledge to take place. In other words, different visual or verbal stimuli jostle for students' attention, resulting in them devoting their attention to one bit of visual/verbal information at one time and to another bit at a later time. Student learning, then, is affected owing to the inability to seamlessly integrate information in dual modalities or different bits of information in a single modality.

In the present study, modality effect might have been produced as a result of presenting information via pedagogical agents, visual, and spoken texts. Yet redundancy effect could have come about in the learning environment as well, as the agents' non-verbal behaviors were processed in the participants' visual memory in addition to visual texts. Consequently, split-attention effect could have been produced as a result of the learners' attention being divided between the agents showing these behaviors and the on-screen and spoken texts. Seamless integration of information, then, is affected due to visual memory overload.

Given that agents' paralinguistic behaviors can be deemed to be redundant or extra bits of information that increase the size of the redundancy effect in tutorial situations, one explanation that can be given for the

differential effects of the two modalities of virtual tutors on student learning and retention is that as for the anthropomorphic condition, greater redundant information might have been present in the visual modality than in the auditory modality, as the anthropomorphic agent was able to display a wide range of human non-verbal behaviors. Accordingly, a stronger redundancy effect might have been produced leading the participants to struggle for processing both visual texts and the agent's non-verbal behaviors concurrently in their visual memory. The learners, then, should have divided their attention between the on-screen texts and the tutor's behaviors to allow for the visual memory to free up space for information processing. Contiguous or parallel processing of information could have been affected, however.

As for the non-anthropomorphic pointing finger, the tutor was capable of mimicking only a small subset of human non-verbal behaviors. Accordingly, little redundant information could have existed in the visual modality and hence there was a strong possibility for greater modality effect to be present in the tutorial. Thanks to a smaller extrinsic cognitive load, the visual memory might not have been occupied to the extent that the participants' attention would have to become split, allowing for seamless, uninterruptable integration of information. Further evidence corroborating such a view comes from the participants' responses to qualitative probes that could help shed light on potential cognitive and affective benefits of the two agent modalities.

When asked to share their opinions about the emotional or affective impact, if any, of the virtual tutors on their curiosity, motivation, and engagement with the learning task, a vast majority of the participants in the anthropomorphic group (around 97%) reported that they had found the agent's behaviors motivating and that they had been greatly influenced by the charisma, persona, and appeal of the anthropomorphic virtual tutor. One participant, for example, shared his impression in this way:

Vignette 3: I really enjoyed the appeal of the virtual teacher: He was attractive and looked quite charming; the tone of voice was strong but also calm and friendly; the facial expressions, body movements, and reactions looked very real… Overall, seeing the teacher behaving like a human and coordinating movements was highly interesting to me…

Whereas a great many participants in the anthropomorphic group found the tutor's behaviors motivating and charming, only a tiny percentage of the participants in the non-anthropomorphic group (around 32%) found the

agent's behavior highly attractive and exciting. Even those who liked the agent's behavior complained that they had found it a bit difficult to identify themselves with the tutor as a real teacher, largely owing to the fact that they could see no face or person on the stage posing or reacting to their responses like a human being. Below is one participant's view on the degree to which she found the agent's behavior motivating and lovable:

Vignette 4: There's nothing special about the virtual teacher. Though the effects and directional movements looked interesting, I could only hear a voice in the background talking about the idioms and their uses in different situations. I could see no one on the stage I could identify with…

Yet when asked about their opinions as to the visual or cognitive impact, if any, of the virtual tutors on their attention, learning, and memory, only a small number of the participants in the anthropomorphic group (around 43%) reported that the humanoid tutor could efficiently sustain their attention throughout instruction, while a fairly large percentage (around 84%) of the learners in the non-anthropomorphic group asserted that they had found the pointing finger effective enough in directing and holding their attention for an extended period of time. The following vignettes represent the participants' ideas about this characteristic feature (i.e., the ability to effectively sustain students' attention throughout instruction) of the anthropomorphic and non-anthropomorphic agents, respectively:

Vignette 5: I believe both the teacher's behavior and the texts were useful in helping me concentrate on the expressions being taught in that I felt they highly persisted in my memory; but, watching the teacher gesturing somehow kept me from holding my attention all the time…

Vignette 6: I never thought highlighting effects like pointing in e-learning environments could've had such a positive effect on memory. Also, I don't think following the visual effects and graphic texts was that distracting; to the contrary, I found it rather useful in helping me concentrate all the time…

What can be gleaned from the descriptions given above is that it is likely that a great number of the participants in the anthropomorphic group found the agent slightly distracting as they struggled to process aural and visual texts, on the one hand and the agent's behaviors, on the other hand. The implication is that active construction of knowledge resulting from contiguous presentation and simultaneous processing of information could have been repeatedly interrupted, reducing the learning efficiency. Conversely, in all likelihood, the non-anthropomorphic group might have had little difficulty

processing the agent's behavior, as the visual texts and the non-verbal behavior of pointing could have been processed as essentially one unified visual element along with the spoken texts in the verbal memory. Accordingly, the bits of information available in dual modalities could have been more or less balanced, resulting in a strong modality effect to come about in the tutorial setting.

Conclusions, Implications, and Future Studies

The present study has found that agent-based instruction can hold great promise for the teaching of L2 idioms by emphasizing a balanced approach where both form and meaning are equally emphasized during instruction. It has also revealed that anthropomorphism is one dimension of agent modality that may correlate with the effects of virtual tutors on students' learning and retention in tutorial situations. Among the suggested explanations was the idea that the epitome of a non-anthropomorphic tutor would produce a stronger modality effect thanks to the presence of little redundant visual information, fewer occurrences of split-attention effect, facilitated processing, and hence seamless integration of information in dual modalities.

Considering that bundles of multimedia courseware and tutorials with embedded pedagogical agents have not been used extensively in language pedagogy, it is advisable that ELT teachers, especially in EFL contexts, show wider recognition for the delivery of agent-based instruction in educational settings, appreciating the fact that this type of instruction can benefit language learners in much the same way as it privileges and continues to privilege the learners of other disciplines.

Language learners can likewise gain great advantage from agent-based instruction in the following ways: (a) learning from pedagogical agents can be potentially motivating and engaging thanks to their being lifelike or their lifelike human attributes; (b) the learning experience itself presents an exciting opportunity for learners to engage in learning tasks in a new, convivial atmosphere; (c) there is a strong possibility for learners devoting sustained attention to diverse elements (sub-schemas) of the knowledge or skill being learnt by following the agent drawing on social and visual signaling; and (d) there are great opportunities for individualized, learner-controlled instruction where learners can more or less control the pace of learning by personalizing their learning experience.

Finally, as for courseware designers, the findings of the present study suggest that it is essential that multimedia applications should deliver instruction in such a way as to reduce the extrinsic cognitive load to allow for the working memory to allocate more processing power to *germane load* (Sweller, 2010) or that portion of mental load required for the formation of schemata. In the present study, efficient learning of L2 idioms required that the participants should integrate three interrelated sub-schemas or aspects of idiom knowledge (i.e., knowledge of form, meaning, and appropriate use) into higher-order schemata. For this to happen, a heightened level of germane load should serve as a precondition for effective knowledge construction. One way to accomplish this is by incorporating into multimedia courseware pedagogical agents that are assumed to deliver the course content in optimal ways. Non-anthropomorphic virtual tutors, as the study suggests, can serve as one good candidate vying for incorporation into tutorials, thanks to the likelihood of their delivering instruction in ways that maximize the germane load by reducing the extrinsic load. The possibility for a stronger modality effect produced as a result of the balanced processing of information in dual modalities provides further support for their inclusion in educational applications.

The present study employed a relatively small sample size (30 participants in each study group) who might not have been representative of the population of EFL learners. It is recommended, therefore, that future studies should recruit larger groups of learners to examine whether the claims made in the present study are substantiated and actually hold for a bigger constellation of participants. In the present study, only a small sample of English idioms were taught through the multimedia units. Despite the fact that a concerted attempt was made to choose high frequency idioms, they might not have typified all the L2 idioms used in everyday communication. Future samples can, then, include a wider range of English idioms. The present study explored only one dimension of agent modality along which different categories of pedagogical agents could be identified. It might be intriguing to examine other dimensions of agent modality and their implications for language learning. For example, virtual tutors can differ with respect to degrees of realism such that they can be identified as “unrealistic”, “moderately realistic”, and “highly realistic” agents (Sahimi et al., 2010, p.175), or with respect to the stereotypic role they play in the tutorial and hence can be identified as “peer-like” or “expert-like” agents (Liew, Tan, &

Jayothisa, 2013, p. 275), or even with respect to their gender so that they can be attracted by different genders (Ozogul, Johnson, Atkinson , & Reisslein, 2013). Finally, future studies may examine what implications agent-based instruction carries for the teaching of other components or skills of the L2 such as its vocabulary and grammar, and whether or not level of proficiency can also interact or correlate with other variables to affect learning from pedagogical agents.

References

Abel, B. (2003). English idioms in the first language and second language lexicon: A dual representation approach. *Second Language Research*, *19(*4), 329-358.

Allen, L. K., Crossley, S. A., Snow, E. L., & McNamara, D. S. (2014). L2 writing practice: Game enjoyment as a key to engagement. *Language Learning & Technology*, *14*(2), 124-150. Retrieved from *http://llt.msu.edu/issues/june2014/allenetal.pdf*

Ayres, P., & Sweller, J. (2005). The split-attention principle in multimedia learning. In R. E. Mayer (Ed.), *Cambridge handbook of multimedia learning* (pp. 135–147). UK: Cambridge University Press.

Babaie, H. (2008). On the effects of help options in MCALL programs on the listening comprehension of EFL learners. *TELL Journal*, *2*(6), 27-47.

Boers, F., Demecheleer, M., & Eyckmans, J. (2004). Cross-cultural variation as a variable in comprehending and remembering figurative idioms. *European Journal of English Studies*, *8*(3), 375-388.

Bryant, F. B., & Yarnold, P. R. (1995). Principal components analysis and exploratory and confirmatory factor analysis. In L. G. Grimm & R. R. Yarnold (Eds.), *Reading and understanding multivariate statistics* (pp. 99-136). Washington, DC: American Psychological Association.

Cambridge University Press. (2013). *Cambridge IELTS 9: Authentic examination papers from Cambridge ESOL*. New York: Cambridge.

Cameron, L., & Low, G. (1999). Metaphor. *Language Teaching*, *32*(1), 77-96.

Carlotto, T., & Jaques, P. A. (2016). The effects of animated pedagogical agents in an English-as-a-Foreign-Language learning environment. *Journal of Human Computer Studies*, *90*, 1-32.

Cassell, J., & Thorisson, K. (1999). The power of a nod and a glance: envelope vs. emotional feedback in animated conversational agents. *Applied Artificial Intelligence*, *13*(4), 519- 538.

Chapelle, C., & Mizuno, S. (1989). Student's strategies with learner-controlled CALL. *CALICO Journal*, *7*(2), 25-47.

Chen, Y., C., & Lai, H., L. (2013). Teaching English idioms as metaphors through cognitive-oriented methods: A case in an EFL writing class. *English Language Teaching*, *6*(6), 13-20.

Clarebout, G., Elen, J., Johnson, W. L., & Shaw, E. (2002). Animated pedagogical agents: An opportunity to be grasped? *Journal of Educational Multimedia and Hypermedia*, *77*(3), 267-286.

Collentine, K. (2001). Learner autonomy in a task-based 3d world and production. *Language Learning & Technology*, *15*(3), 50-67.

Corbeil, G. (2007). Using the French tutor multimedia package or a textbook to teach two French past tense verbs: Which approach is more effective? *CALICO Journal*, *24*(2), 313-330. Retrieved from https://calico.org/html/article_647.pdf

Dai, J., Raine, B. R., Roscoe, R., Cai, Z., & McNamara, D. S. (2011). The Writing-Pal tutoring system: Development and design. *Journal of Engineering and Computer Innovations*, *2*(1), 1-11.

Duffy, M. C., & Azevedo, R. (2015). Motivation matters: Interactions between achievement goals and agent scaffolding for self-regulated learning within an intelligent tutoring system. *Computers in Human Behavior*, *52*, 338-348. http://dx.doi.org/10.1016/j.chb.2015.05.041

Dunsworth, Q., & Atkinson, R. K. (2007). Fostering multimedia learning of science: Exploring the role of an animated agent's image. *Computers and Education*, *49*(3), 677-690.

Ergül, E., & Koç, M. (2013). The role of animated agents in web-based distance education. *Procedia—Social and Behavioral Sciences*, *83*, 1016-1022. http://dx.doi.org/10.1016/j.sbspro.2013.06.189

Graesser, A. C., Person, N., Harter, D. (2001). Teaching tactics and dialog in AutoTutor. *International Journal of Artificial Intelligence in Education, 12,* 257-279.

Grant, L. E. (2003). *A corpus-based investigation of idiomatic multiword units*. (Doctoral dissertation). Available from ProQuest Dissertations and Theses database.

Gulz, A., & Haake, M. (2006). Design of animated pedagogical agents - A look at their look. *International Journal of Human-Computer Studies, 64*(4), 322-339.

Heidig, S., & Clarebout, G. (2011). Do pedagogical agents make a difference to student motivation and learning? A Review of Empirical Research. *Educational Research Review, 6,* 27-54.

Heift, T. (2007). Learner personas in CALL. *CALICO Journal, 25*(1), 1-10. http://dx.doi.org/10.1558/cj.v25i1.1-10

Hongpaisanwiwat, C., & Lewis M. (2003) Attentional effect of animated character. In M. Rauterberg, M. Menozzi, & J. Wesson (Eds.), *Human-computer interaction-INTERACT '03* (pp. 423-430). Nieuwe Hemweg, Amsterdam: IOS Press.

Johnson, A. M., Ozogul, G., & Reisslein, M. (2014). Supporting multimedia learning with visual signalling and animated pedagogical agent: moderating effects of prior knowledge. *Journal of Computer Assisted Learning, 31*(2), 97-115.

Johnson, J., & Rosano, T. (1993). Relation of cognitive style to metaphor interpretation and second language proficiency. *Applied Psycholinguistics 14,* 159-175.

Johnson, W. L., Rickel, J. W., & Lester, J. C. (2000). Animated pedagogical agents: Face-to-face interaction in interactive learning environments. *International Journal of Artificial Intelligence in Education, 41,* 41-78.

Laborda, J. G. (2010). Contextual clues in Semi-direct interviews for computer assisted language testing. *Procedia Social and Behavioral Sciences, 2,* 3591-3595.

Liew, T.-W., Tan, S.-M., & Jayothisa, C. (2013). The effects of peer-like and expert-like pedagogical agents on learners' agent perceptions,

task-related attitudes, and learning achievement. *Educational Technology & Society, 16*(4), 275–286.

Massaro, D. W., & Cohen, M. M. (1994). Visual, orthographic, phonological, and lexical influences in reading. *Journal of Experimental Psychology: Human Perception and Performance, 20*, 1107-1128.

Mayer, R. E. (2009). *Multimedia learning* (2nd Ed.). NY: Cambridge University Press.

Mayer, R. E., & Moreno, R. (1998). A split-attention effect in multimedia learning: Evidence for dual processing systems in working memory. *Journal of Educational Psychology, 90*, 312-320.

Moreno, R. (2006). Learning in high-tech and multimedia environments. *Current Directions in Psychological Science, 15*, 63-67.

Nippold, M. A., & Duthie, J. K. (2003). Mental imagery and idiom comprehension: A comparison of school-age children and adults. *Journal of Speech, Language, and Hearing Research, 46*, 788-799.

Noma, T., & Badler, N. I. (1997). A virtual human presenter. *Proceedings of the IJCAI Workshop on Animated Interface Agents: Making Them Intelligent*. San Francisco: Morgan Kaufmann.

Ozogul, G., Johnson, A. M., Atkinson, R. K., & Reisslein, M. (2013). Investigating the impact of pedagogical agent gender matching and learner choice on learning outcomes and perceptions. *Computers & Education, 67*, 36-50.

Paivio, A., & Walsh, M. (1993). Psychological processes in metaphor comprehension and memory. In A. Ortony. (Ed.), *Metaphor and thought* (pp. 307-328). Cambridge: Cambridge University Press.

Pallant, J. (2013). *SPSS survival manual* (5th Ed.). NY: Open University Press.

Reeves, B., & Nass, C. (1996). *The media equation: How people treat computers, televisions, and new media like real people and places*. University Press, Stanford, California.

Ryu, J., & Baylor, A. (2005). The psychometric structure of pedagogical agent persona. *Tech., Inst., Cognition and Learning, 2*, 291-314.

Sahimi, S. M., Zain, F. M., Kamar, N. A. N., Samar, N., Rhman, Z. A., Majid, O., Luan, W. S. (2010). The pedagogical agent in online

learning: Effects of the degree of realism on achievement in terms of gender. *Contemporary Educational Technology*, *1*(2), 175-185.

Sklar, E., & Richards, D. (2010). Agent-based systems for human learners. *The Knowledge Engineering Review*, *25*(2), 1–25.

Strauss, A., & Corbin, J. (1998). *Basics of qualitative research techniques and procedures for developing grounded theory*. Thousand Oaks: Sage Publications.

Sweller, J. (2005). Implications of cognitive load theory from multimedia learning. In R. E. Mayer (Ed.), *The Cambridge handbook of multimedia learning* (pp. 19-30). New York: Cambridge University Press.

Sweller, J. (2010). Element interactivity and intrinsic, extraneous, and germane cognitive load. *Educational Psychology Review*, *22*(2), 123-138.

Toyoda, E. (2001). Exercise of learner autonomy in project-oriented CALL. *CALL-EJ Online*, *2*(2), Retrieved from *http://callej.org/journal/2-2/toyoda.html*

Van der Meij, H., van der Meij, J., & Harmsen, R. (2015). Animated pedagogical agents' effects on enhancing student motivation and learning in a science inquiry learning environment. *Education Tech Research Dev*, *63*, 381-403.

Vasiljevic, Z. (2015). Effects of etymology and pictorial support on the retention and recall of l2 idioms. *Electronic Journal of Foreign Language Teaching*, *12*(1), 35-55.

Veletsianos, G., & Miller, C. (2008). Conversing with pedagogical agents: A phenomenological exploration of interacting with digital entities. *British Journal of Educational Technology***,** *39*(6), 969-986.

Yılmaz, R., & Kılıç-Cakmak, E. (2012). Educational interface agents as social models to influence learner achievement, attitude and retention of learning. *Computers & Education*, *59*, 828-838.

Yung, H. I., & Pass, F. (2015). Effects of cueing by a pedagogical agent in an instructional animation: A cognitive load approach. *Educational Technology & Society*, *18*(3), 153-160.