Running Head: DESIGNING PEDAGOGICAL AGENT STUDIES

# Five Design Principles for Experiments on the Effects of Animated Pedagogical Agents

Richard E. Clark and Sunhee Choi

Rossier School of Education, University of Southern California

University of Southern California

Waite Phillips Hall 601C

Los Angeles, California 90089-0031

Phone (310) 377 7220

Fax (310) 909 8606

Email: clark@usc.edu

First draft submitted: September 19, 2005

Revised draft submitted: February 19, 2005

Published in the *Journal of Educational Computing Research*, 32(3), 209-223.

## Abstract

Research on animated pedagogical agents (agents) is viewed as a very positive attempt to introduce more pedagogical support and motivational elements into multi-media instruction. Yet, existing empirical studies that examine the learning benefits of agents have had very mixed results, largely due to the way that they are designed. This paper will suggest five design principles for future research on the impact of agents on learning and motivation including: (a) The Balanced Separation Principle describes need for adequate controls that tease out the specific type of learning and/or motivational support the agent is providing; (b) The Variety of Outcomes Principle suggests different measures to test a variety of learning and motivation outcomes that may be influenced by agents; (c) The Robust Measurement Principle advises researchers to pay special attention to the reliability and construct validity of experimenter designed measures; (d) The Cost-Effectiveness Principle recommends the collection of data on the relative cost of producing agent and non-agent treatments; and (e) The Cognitive Load Principle alerts those who plan treatments for experiments to exercise caution when developing and testing agents that are visually and aurally "noisy" or complex.


# Five Suggestions for the Design of Experiments on the Effects of Animated Pedagogical Agents

## Introduction

Animated pedagogical agents (agents) are defined by Craig, Gholson and Driscoll (2002) as "a computerized character (either humanlike or otherwise) designed to facilitate learning" (p. 428). Atkinson (2002) suggests that these animated computer-based instructional agents " ... reside in the learning environment by appearing as animated "humanlike" characters, which allows them to

exploit ... communication typically reserved for human-human interaction ... [and] can focus a learner's attention by moving around the screen, using gaze and gesture, providing ... feedback and conveying emotions" (p. 416-417). Agents are a product of recent technological advances in computer animation and user interface design. Advocates suggest that they have great potential for human learning (e.g., Sampson, Karagiannidis, Kinshuk, 2002). Since the design of agents is a recent, welcome and visible attempt to provide pedagogical support during multimedia instruction, a review of early research studies seems warranted[1].

Examples of Agents

Agents with names such as Herman, Steve, Adele, and Auto Tutor have been developed recently to serve a variety of instructional goals in computer-based instruction. Different agents have been assigned distinct pedagogical roles and functions during instruction including, for example to: (a) provide a learner with study suggestions and information based on their progress in a lesson (Baylor, 2002b; Moreno, Mayer, Spires, & Lester, 2001); (b) demonstrate interactively how to perform tasks (Johnson, Rickel, & Lester, 2000; Samson et al., 2002); (c) direct learners' attention on certain elements or aspects of instructional presentations with familiar and natural methods, such as gestures, locomotion, or gaze (Atkinson, 2002); and (d) provide nonverbal as well as verbal feedback on learners' actions (André, Rist, & Müller, 1999; Cassell & Thórisson, 1999; Craig, Driscoll, Gholson, 2004). Many agents provide nonverbal feedback through behaviors such as head nodding for approval, head-shaking for disapproval, jumping up and down for congratulating students' success, and a look of puzzlement for disagreement or negative feedback (Johnson et al., 2000).

[1] We have examined all of the animated pedagogical agent studies we could locate in peer reviewed journals and a number of papers presented in conferences. We would be interested in receiving copies of unpublished technical reports, and unpublished or published studies we have inadvertedly missed. Please send any references or documents to the first author, clark@usc.edu.

Presumed Benefits from Agents

Discussions about the uses of agents in computer based instruction suggest that they might have at least three primary types of learning benefits: (a) Agents may have a positive impact on learners' motivation and how positively they value computer-based learning programs; (b) They might help learners to focus on important elements of learning material crucial for successful learning; and (c) They may also provide learners with context-specific learning strategies and advice (Choi & Clark, 2004).

Despite the claimed effects, very little empirical research has been conducted on the pedagogical effectiveness of agent-based learning environments. According to Baylor (2002a), the majority of agent studies have so far tended to focus on system architecture, development, functionalities, and implementation (e.g., Alpert, Singley & Fairweather, 1999; André, Rist, & Muller, 1998; Bull, Greer, & McCalla, 2003; Lester, Towns, & Fitzgerald, 1999; Ogata, Liu, Ochi, & Yano, 2001; Person, Graesser, Kreuz, & Pomeroy, 2001, Song, Hu, Olney, & Graesser, 2004), rather than agents' learning or motivation impact.

Moreover, the results of the few pedagogically-relevant, published empirical studies are varied (Clark, 2003; Dehn & van Mulken, 2000). In some studies, agent-based instruction results in higher learning scores and/or more positive attitudes towards lessons (e.g. Bossler & Massaro, 2003; Mitrovic & Suraweera, 2000; Moundridou & Virvou, 2002; Ryokai, Vaucelle, & Cassell, 2003), whereas in others, agents produce no learning or motivational benefits (e.g. Andre et. al, 1999; Baylor, 2002b; Craig, Driscoll, & Gholson, 2004; Mayer, Dow, & Mayer, 2002). Yet in many experiments, results are mixed and somewhat confusing (e.g. Atkinson, 2002; Moreno et. al., 2001). The view suggested in this discussion is that these early mixed results are largely due to the way that studies are designed. The purpose of this article is to suggest five design principles for future research in this area and, where possible, to give positive examples (and non-examples) of

each principle from existing agent studies. Most of the principles described here could be applied to many other research areas. We are not suggesting that the principles should be applied exclusively to agent studies. However, all new and evolving research questions tend to be susceptible to certain kinds of design errors more than other. In our discussion, therefore, we have attempted to isolate those we feel are most pressing for agent studies. The goal of these suggestions is to increase the utility of studies that examine the presumed benefits of agents and other pedagogical devices that are invented to aid instruction in new technology-driven instructional environments.

## Five Principles for the Design of Agent Studies

### 1. The Balanced Separation Principle: Separate Pedagogical Agents from Pedagogical Methods

When designing agents studies, control for the type of hypothesized learning and/or motivational support the agent is providing in a balanced, alternative condition where the same type of learning support is provided by a lower technology, non-agent condition. In other words, if the agent is providing a specific type of instructional support, study designs should include a "low technology" alternative method of providing the same type of support to a comparison or control group. Any pedagogical support provided by an agent can also be provided in a 'lean' format. Dehn & van Mulken (2000) explain that without this type of design control, "... differences between the two conditions cannot be attributed exclusively to ... the agent" (italicized words added, p. 18). An adequate test requires that the non-agent or control condition provide all of the learning and motivational support available from the agent condition, otherwise a comparison will be potentially confounded by the uncontrolled effects of the instructional methods the agent provides and the agent itself.

Confusion about the source of measured benefits. For example, Atkinson (2002) compared a "voice plus agent" group with "voice only" and "text only" groups (Experiment 2). In the voice plus agent group, participants listened to the agent's verbal explanations and saw the agent

highlighting relevant information on the screen simultaneously by using pointing gestures. Alternatively, participants in the voice only and text only conditions only received explanations delivered either in voice or text, respectively. In other words, participants in the voice only and text only groups did not have the benefit of a visual, highlighting indicator for important information, which might have forced the participants to use their scarce cognitive resources to connect verbal explanation with related visual information on the screen. Therefore, although the voice plus agent group outperformed the other two groups in far-transfer performance, it is problematic to attribute the obtained learning benefit exclusively to the presence of the agent. The critical learning support provided by the agent--directing learner's attention to the key information in the screen display, was not available to the two comparison groups. A leaner version of the agent's pointing gesture would be to simply use an animated arrow and/or to underline the same information selected by the agent in the comparison conditions. Other studies which also failed to control the types of instructional and motivation supports provided in agent and alternative conditions include Moundridou & Virvou (2002) and Ryokai et al. (2002).

Good examples of agent study design. André and colleagues (1999) conducted a well-controlled study that avoided this design pitfall. To find empirical support for the affective and cognitive benefits of their "PPP Persona" agent, they exposed participants to two different memory tasks--a technical description (the operation of pulley systems) and an informational presentation that included the names, pictures, and office locations of fictitious employees. Both experiment and control versions provided the same treatments except that the control groups did not have the PPP Persona agent. The control group heard a voice conveying the same explanations as the agent provided to the experimental group. The agent's pointing gesture was replaced with an arrow that pointed to important information in the control condition. Following the presentations, participants' affective reaction to the agent and control condition was measured through a questionnaire whereas

the cognitive impact was measured by comprehension and recall questions. The results showed significant differences only in the affective measures. Participants interacting with the PPP Persona agent for the technical description found the presentation less difficult and more entertaining. The positive effects, however, disappeared for the informational presentation about the fictitious employees. Participants reported that the PPP Persona agent was less appropriate for employee information and less helpful as an attention direction aid. No significant achievement differences were found between the experimental and control groups for either the technical description or information presentation tasks on comprehension or recall measures. Thus, in this well-designed study, the agent did not provide learning or motivational benefits that translated to greater learning. Yet, because of the adequate design, there is the serendipitous finding that learners may believe that agents are more appropriate and likeable in some learning tasks but not in others.

Craig et al. (2002) also employed an adequate design where participants learned the process by which lightning occurs presented through an agent and through alternative multimedia (i.e., picture, narration, or animation). An animated agent that pointed to important instructional elements on a computer screen was contrasted with a sudden onset highlighting (i.e., color singleton or electronic flashing) and animation of the same information (without the agent) for comparison groups. The narrative information was synchronized simultaneously with the agent's pointing gestures, separated and provided prior to the agent's pointing, or in a third condition, with a sudden onset of highlighting and animation of relevant parts of an instructional picture. Craig et al.'s results indicated that the agent made no difference in learners' performance both in cognitive load assessment and performance tests (i.e., retention, matching, and transfer). Rather, they reported a significant benefit from both a sudden onset and animation of parts of the pictures for focusing learners' attention. This may be an example of an effect that van Merrienboer (1997) calls "just in time" learning support.

The advice to provide lower technology comparisons should also be applied to experiments that do not compare an agent condition with a non-agent condition, but simply examine agent effects using a pre- and post-experiment, within subject design. In this case, it is necessary to isolate the agent factor from other instructional and motivational supports. For instance, Bosseler and Massaro (2003) tested the potential of Baldi, a three-dimensional computer-animated talking head, on vocabulary and language learning in children with autism using a within subject design. Even though the results showed significant vocabulary learning both in immediate and delayed retention tests, it is not clear whether the measured benefits were due to the talking head or other instructional supports because the vocabulary lesson consisted of images of the vocabulary items and icon-based feedback as well as the animated talking head.

Provide a low technology comparison. Moreno et al. (2001) demonstrated that when the voice of an agent was replaced by a human voice and the image of the agent was deleted, students were still able to learn the material better if an instructional method (i.e., interactivity and contingent feedback) was appropriate for the learning task (Experiment 3). Therefore, it is possible to conclude that the interactive environment that can be realized through human voice is enough to promote deeper learning as well as retention. In addition, Moreno et al. found that the presence of the agent's visual image did not have any impact on affective or cognitive aspect of learning, while the modalities of the agent made significant differences in learning outcomes (Experiment 4 and 5). Craig et al. (2002) also found that highlighting important information was as effective for learning as the agent treatment but the highlighting was significantly cheaper to produce. Thus, it can be inferred that there may often be a far simpler and less expensive way (e.g., highlighting and animation of pictures) to achieve the same learning and motivational benefit.

Are agents necessary for learning? These results also imply that the differences made in student learning may not be due to the agent by itself or any increased motivation or attention

caused by the agent, but rather due to the pedagogical method provided by the agent. Thus we should ask a question, 'Is the animated pedagogical agent the only way to deliver these types of instructional methods in a computer-based environment?'. If alternative ways can deliver the same instruction with the same learning and motivation, but with less cost, shouldn't we choose the least expensive option?

Erickson (1997) argued that the adaptive functionality of an instructional system is often enough for learners to perform a task and achieve the same outcome without the guidance of an agent. He further suggested that when including an agent, instructional designers should think about what benefits and costs the agent would bring, and far more research should be conducted on how people experience agents. Furthermore, Nass and Steuer (1993) found that simply using a human voice without the image of an agent was sufficient to induce learners to use social rules when interacting with a computer. Moreno and colleagues (2001) also noted that learners may form a social relationship with a computer itself without the help of an agent and thus, the image of an agent might not be necessary to invoke a social agency metaphor in a computer-based learning environment. Bayer (2002b), Craig et al. (2004) and Mayer et al. (2003) found no effect of agent image on learning outcomes.

2. The Variety of Measured Outcomes Principle: Test for Complex Problem Solving and Transfer

The second recommendation is that pedagogical agent studies test a variety of learning outcomes including memory for facts, conceptual understanding of processes and "how to" problem solving procedures as well as the transfer of new knowledge to solve similar problems (near transfer) and also related but highly novel problems (far transfer). Many existing agent studies tested factual recall and solving of very simple problems to measure students' cognitive learning outcomes (Lester, Converse, Stone, Kahler, & Barlow, 1997; Mitrovic & Suraweera, 2000; Moundridou & Virvou, 2002) as well as learners' subjective reactions to the computer

instruction and learning experience in a questionnaire format. In order to design effective computer-aided instruction, however, it is crucial to know if agents work better than other formats for instruction on a variety of more challenging learning and motivational outcomes including conceptual understanding of process, principles, and problem-solving measures that provide both near and novel transfer problems along with motivational measures that go beyond simple reaction questionnaires.

Two studies with similar findings. In the first two of Moreno et al. (2001) studies, an agent who "personalized" instruction did not aid recall of facts or the transfer of simpler knowledge to solving problems. These treatments also failed to demonstrate a positive impact on student ratings of the "understandability and difficulty" of problems or their motivation to learn the material. However, their personalizing treatment using the agent was found to be superior for the solving of complex problems that required more cognitive effort. Atkinson (2002) also reported superior performance of the voice plus agent group in far-transfer problems, but found no significant difference in near-transfer problems between the voice plus agent group and the voice only group. The fact that two separate research efforts have now found far transfer learning benefits (but not near transfer or recall benefit) for voice plus agents is a serendipitous finding worth further study.

Nevertheless, it should be noted that the two studies reporting far transfer benefits did not follow the first principle of controlling for instructional methods by comparing agent delivered instruction with low technology, non-agent versions of the instructional method. Moreno et al. (2001), in fact, provided different practice conditions and information for the agent group than were provided to the control group in Experiment 1 and 2. Participants in the agent group were asked to design a plant based on environmental conditions (e.g., the amount of sunlight and rainfall) and then, received verbal feedback from the agent on their choices. On the contrary, participants in the control group were given onscreen text explaining about design choices under specific

environmental conditions (without the agent), but they were not allowed to design plants. These different practice conditions might have made a significant contribution to the outcome regardless of the influence of the agent. It is therefore possible that the far (but not near) transfer benefits may be due, in large part, to a personalizing effect that is facilitated by agents.

In addition, the fact that only a small number of knowledge domains have been the focus of agent studies poses another challenge for agent researchers to overcome. Due to the lack of the variety of subject matters tested in the field, it is still not known what types of domain could benefit most from agents with what kinds of functionalities. To date, most pedagogical agent studies have employed discovery-based learning environments and science instruction even though there is compelling evidence that the agent effect is domain specific (Dehn & van Mulken, 2000). For instance, van Mulken, André, and Müller (1998) report that learners perceived an agent more helpful and appropriate when it presents technical information than it does non-technical information. Sproull, Sproull, Subramani, Kiesler, Walker, and Waters (1996) also found that subjective ratings of an agent were lower than text when it was embedded in a career-counselling system. Nevertheless, both findings are still limited to subjective measures of agent effect, so further research into this issue using objective measures is warranted.

3. The Robust Measurement Principle: Insure the reliability and construct validity of all measures

It is very important to insure that all measures in agent studies, including outcome measure, are reliable and have construct validity. Without reliable measures that connect with past studies of learning from instruction, the results obtained from a statistical analysis might not support claims about the effects of agents (Pedhazur & Schmelkin, 1991). Many agent studies appear to use questionable, researcher-designed measuring instruments and assessment procedures without reporting the reliability or construct validity. The only exception to this informal observation that we have found so far is Baylor & Ryu's study (2003) which investigated the effects of different

agent facial properties.

Motivational measures. While a number of studies have claimed to find increased student motivation due to the instructional use of agents, these claims are largely based on the use of a variety of experimenter-designed Likert-type scales that employed very different questions to (presumably) measure the same construct. In most cases it was not obvious that questions or instruments were designed after even a cursory review of current motivation theories or research. Thus, it is difficult to conclude whether these studies in effect measured any aspect of the construct commonly called 'motivation'. Furthermore, many studies adopted an intuitive, general folk psychology view of motivation in that they provided a definition of motivation that in fact refers to some aspects of interest, which is a separate construct (Pintrich & Schunk, 2002). Although there is a certain relationship between interest and motivation, it is necessary to differentiate these two constructs. Because an in-depth discussion of motivation and its measures is beyond the scope of this paper, we recommend the excellent discussion by Pintrich & Schunk (2002) about "motivational indexes". They recommend the measurement of at least three different types of motivational outcomes: a) Active choice--moving from intention to action. This variable is relatively easy to measure since once someone starts doing something, they have exercised active choice. This outcome is often measured when students have more than one option, for example selecting a classroom-based or computer-based version of the same course or when computer-based courses are offered as electives; b) Persistence--focused work on a learning task, over time, in the face of distractions. This, index can be measured by tracking the consistency of interactions over time in media-based programs, and; c) Mental Effort--the amount of mental effort invested in learning (see a discussion and simple self-report measure developed by Paas 1992 and Paas & van Merrienboer,1994). Presumably, all of the motivational strategies built into agent studies should be intended to impact one or more of these three motivational indexes suggested by

Pintrich and Schunk (2002).

Attention measures. In general, attention has been measured either by administering a subjective, self-report questionnaire (e.g., van Mulken et al, 1998; André et al., 1999) or by calculating reaction times and/or error rates in a learner's responses (e.g., Koda & Maes, 1996; Moundridou & Virvou, 2002). There are two major problems with these types of measurements. With regard to asking learners whether they paid attention to the agent or task, it is very unlikely that learners have a capability to access their internal process and report accurately what they have actually experienced without making any inferences (Dehn & van Mulken, 2000). With regard to reaction time and error rates, researchers tend to interpret data from these two types of measures in very different ways. For instance, Sproull and colleagues (1996) interpreted the longer response times produced by the agent group as an indication of a higher degree of attentiveness. On the contrary, Takeuchi and Naito (1995) took the longer interaction times produced by the agent group as the sign of learners being distracted by the agent. Both interpretations may be correct. Agents might require the investment of much more attention and mental effort than non-agent methods because of increased demands on cognitive processing. Yet the increased demand may not be beneficial to learning (Mayer, 2002). Attention may be a very poor substitute for the motivational indexes described by Pintrich & Schunk (2002).

One exception to the standard proposed in this principle can be found in studies that are designed to use various qualitative or mixed-method approaches to generate new hypotheses, find evidence for the robustness of a class of hypotheses or for extended term evaluation designs (Chatterji, 2004). While observational reliability is always an issue when anything is measured, most types of validity tend not to be in the forefront when generating or substantiating hypotheses.

4. The Cost-Effectiveness Principle: If possible, collect information about the relative cost and benefit of producing the agent and non agent treatments being compared.

A number of critics of media studies have challenged the type of outcomes measured in published studies. Since design attempts to extend our current knowledge about the relationship between independent and dependent measures, the choice of dependent measures is very much an issue when experimental plans are being made. Past reviews of media research (for example Clark, 2001, Levie and Dickie, 1973; Levin and McEwan, 2001; Ross and Morrison, 1989; 1996) have urged designers to emphasize economic outcomes in their designs in addition to learning and motivation. Media and media-based instructional strategies tend to influence the cost of instruction but not learning or motivation. While researchers may not be willing to accept the claim that media-based instructional strategies do not influence learning, there is widespread agreement that they might significantly influence the cost of access to instruction and or the speed of learning ( Cobb, 1997). Yet none of the published studies of agent use that we have located report the relative cost and benefit of agent and non-agent treatments. Since the design and use of agents in instruction is presumably more expensive than low-technology alternatives, it seems rational to suggest that the argument for agent use should include an estimate of its financial benefit as well as its learning benefit. Even if agent use in some contexts can be demonstrated to enhance learning more than leaner alternatives, it seems reasonable to ask about the relative cost of the advantage when the expense is amortized over smaller and larger numbers of students. Ellen Langer (1994) suggests that rational, deliberate, conscious decision making (in any domain) may most often be a myth and argues that “the processes that are most generally understood as leading to decisions, such as integrating and weighing information in a cost-benefit analysis, most often are post-decision phenomena, if they occur at all …. Cognitive commitments are frozen on rigidly held beliefs …. Once a cognitive commitment is reached, choice follows mechanically, without calculation.” (p. 34). It may require considerable persistence and mental effort to overcome our automated inclination to advocate innovations such as agents without adequate supporting evidence that

includes cost-benefit analysis.

Ingredients method of determining the cost of agents and alternatives. While there are a number of emerging ways to determine local costs and efficiencies, one of the soundest and most comprehensive is the "ingredients method" developed by Levin, Glass and Meister (1987) and Levin and McEwan (2001) who have described many approaches to assessing the cost-effectiveness and cost-benefit of various technology-based instructional innovations. Their ingredients method "requires identification of all of the ingredients required for the ... [agent-based] intervention, a valuation or costing of those ingredients and a summation of the costs to determine the cost of the intervention" (Levin, 1988, p. 3). In a K-12 setting, cost is defined as the value of what is given up by using resources in one way rather than for its next best alternative use. For example, if teacher or animation artist time is given up then it may not be used for other purposes. Therefore, the cost of teacher and graphic artist time is assessed by using their hourly cost and assigning a value to what is lost when teachers or artists are assigned to design computer based programs containing agents.

The ingredients method is implemented in two stages. In the first stage, all necessary program ingredients are listed. The identification of ingredients requires that we list distance education program necessities associated with five categories: (a) personnel, (b) facilities, (c) equipment, (d) materials and supplies and (e) all other. In the second stage, each of the ingredients listed in each of the five categories is valued. Space limitations preclude a complete description of the ingredients method but a review of by Levin et al. (1987) and Levin & McEwan (2001) will provide most of the information needed to adequately determine ingredient costs. They reason that failure to conduct a complete cost analysis will give an unrealistic picture of the "replication" expense. He also claims that, in the rare instance where one finds a complete costing of technology-based programs, one often finds evidence that the organizational climate greatly

influences cost-benefit ratios. He presents evidence that when the same distance education program is presented in many different sites, the cost of implementation can vary by 400% (Levin et al., 1987). Some organizations seem much more efficient than others in the design and delivery of technology-based instruction.

5. The Cognitive Load Principle: Avoid Testing Agents That Are Visually and Aurally "Noisy" or Complex.

Working memory will, on the average, tolerate only 4 +/- 2 chunks of information before learners become overloaded. This estimate is more current and accurate than the 7 +/- 2 estimate suggested in the 1950's (Cowan, 2001). Mayer (2001) has documented the learning deficits experienced by learners who are overloaded by adding interesting but conceptually irrelevant material to a computer-based lesson. Clark (1999) has described some of the theories that have been advanced to account for the learning deficits that result when learners are overloaded cognitively. Therefore, it seems important to focus an agent on learning goals and provide both the simplest, least distracting and most conceptually relevant support for the achievement of learning objectives.

Is instructional entertainment beneficial or distracting? The underlying premise of including entertaining elements in instructional presentation is that increased learner interest will lead learners to pay more attention, to persist longer in the system, and to exert more effort to process the lesson. Evidence from research on this premise however, seems to indicate that including entertaining elements that lack conceptual relevance to main ideas of the lesson does not necessarily result in increased learning. In a study by Anderson, Shirey, Wilson, and Fielding (1984), fourth graders read a series of sentences rated for interest value, and were tested on recall. Students allocated more attention and recalled better interesting sentences, but a causal relationship was not found between attention and performance. Shirey & Reynolds (1988) also found that attention did not work a causal mediator between interest and learning with adult learners. Rather,

according to Harp and Mayer (1997, 1998), inserting interesting but conceptually irrelevant text or illustrations actually hinder learning process because they may distract learners away from selecting and processing key elements of the lesson-- “Seductive Details Effect”.

Cognitive load theory and agents. According to cognitive load theory (Kalyuga, Chandler, & Sweller, 1999), the mere presence of an animated pedagogical agent can be detrimental to learning by dividing a learner’s limited cognitive resources into different visual segments in a multimedia presentation. More specifically, cognitive load theory predicts that when animation of an agent is presented simultaneously with other visual information such as graphics or on-screen text, learners need to split their attention between these two sources, and as a consequence, the presence of the agent becomes harmful to learning rather than beneficial. This appears to cause a condition where the learning strategies an agent is attempting to convey are never used by learners. In addition, complex and noisy agents could cognitively overload learners and so make it impossible for them to benefit from the agent’s help.

## Conclusion

The phrase “research design” denotes a plan for conducting an experiment. The plans for an experiment determine, in large part, what we might learn (or fail to learn) from a systematic attempt to test hypotheses. Plans for experiments tend to be developed in one area of inquiry and generalized to another without a careful consideration of the special characteristics of the new context and what will, and will not be learned, as a result of different features in a design. To some extent, the five principles offered in this article to improve the utility and interpretability of studies on animated pedagogical agents could well be applied to all attempts to design studies that examine new electronic media based audio-visual formats for instructional methods. The design problems noticed by critics tend to be similar in media and learning or motivation studies.

Historically, interest in new media has been so intense that it has often overwhelmed our

substantive prior learning about the design of adequate experiments. It also seems to overwhelm the better judgment of some reviewers and journal editors. This was true when movies and television were adopted for use in instruction and it continues to be an issue in the design of studies that attempt to benefit from the introduction of personal computers, multi-media and virtual reality (Clark, 2001). Inadequate designs often provide evidence that supports misconceptions about our favorite hypotheses. In this way, otherwise educated, thoughtful and ethical researchers have unintentionally contributed to misperceptions about the impact of media-based instruction on learning, motivation and related types of performance.

Enthusiasm for pedagogical agent studies, on the one hand, and cautious design on the other, are not necessarily antagonistic. However, as instructional researchers interested in educational computing, our primary commitment and enthusiasm must be to tease out the active ingredients in instruction that provide learning and motivation benefits to students or economic benefits to those who provide education and training. Adequate design and a concern for current theories of learning and instruction will help us develop accurate evidence for the benefits of new technologies similar to the tempting insight about the impact of agents on farther transfer of knowledge that was found in the studies by Moreno et al. (2001) and Atkinson (2002). We may also discover counter-intuitive and negative effects imposed by new instructional devices such as those identified by Mayer (2001) and so advise instructional designers about the pitfalls they should avoid when developing instructional programs. The goal of design is to help us identify the “active ingredients” (Clark and Estes, 1999; 2000) that have led to measured outcomes in a way that will permit us to transfer what we learn and implement it in instructional design and development.

## References

Alpert, S. R., Singley, M. K., & Fairweather, P. G. (1999). Deploying intelligent tutors on the web: An architecture and an example. International Journal of Artificial Intelligence in

Education, 10, 183-197.

Anderson, R. C., Shirey, L. L., Wilson, P. T., & Fielding, L. G. (1984). Interestingness of children's reading material (Report No. 323). Urbana-Champaign, IL: Center for the Study of Reading.

André, E., Rist, T., & Muller, J. (1998). Web persona: A life-like presentation agent for the world-wide web. Knowledge-based Systems, 11(11), 25-36.

Atkinson, R. K. (2002). Optimizing learning from examples using animated pedagogical agents. Journal of Educational Psychology, 94(2), 416-427.

André, E., Rist, T., & Müller, J. (1999). Employing AI methods to control the behavior of animated interface agents. Applied Artificial Intelligence, 13, 415-448.

Bosseler, A., & Massaro, D. (2003). Development and evaluation of a computer-animated tutor for vocabulary and language learning in children with autism. Journal of Autism and Developmental Disorders, 33(6), 653-672.

Baylor, A. L. (2002a). Agent-based learning environments as a research tool for investigating teaching and learning. Journal of Educational Computing Research, 26(3), 249-270.

Baylor, A. L. (2002b). Expanding preservice teachers' metacognitive awareness of instructional planning through pedagogical agents. Educational Technology Research & Development, 50(2), 5-22

Baylor, A. L., & Ryu, J. (2003). Does the presence of image and animation enhance pedagogical agent persona? Journal of Educational Computing Research, 28(4), 373-395.

Bull, S., Greer, J., & McCalla, G. (2003). The caring personal agent. International Journal of Artificial Intelligence in education, 13(1), 21-34

Cassell, J., & Thórisson, K. R. (1999). The power of a nod and a glance: Envelope vs. emotional feedback in animated conversational agents. Applied Artificial Intelligence, 13,

519-538.

Chatterji, M. (2004). Evidence on “what works”: An argument for extended-term, mixed-method (ETMM) evaluation designs. Educational Researcher. 33(9). 3-13.

Clark, R. E. (1999). Yin and Yang cognitive motivational processes operating in multimedia learning environments. In J. van Merrienboer (Ed.), Cognition and Multimedia Design. Herleen, Netherlands: Open University Press.

Clark, R. E. (2000). Evaluating distance education: Strategies and cautions. The Quarterly Journal of Distance Education, 1(1), 5-18.

Clark, R. E. (Ed.). (2001). Learning from Media: Arguments, Analysis and Evidence. Greenwich, CT: Information Age Publishers

Clark, R. E. (2003). Research on web-based learning: A half-full glass. In R. Bruning, C. Horn, & L. M. PytlikZillig (Eds.), Web-Based Learning: What do we know? Where do we go? (pp. 1 – 22). Greenwich, CT.: Information Age Publishers.

Clark, R. E. & Estes, F. (April, 2000) A Proposal for the Collaborative Development of Authentic Performance Technology. Performance Improvement, 40(4) 48-53.

Clark, R. E. & Estes, F. (1999) The Development of Authentic Educational Technologies, Educational Technology, 37(2) 5 - 16.

Craig, S. D., Gholson, B., & Driscoll, D. M. (2002). Animated pedagogical agents in multimedia educational environments: Effects of agent properties, picture features and redundancy. Journal of Educational Psychology, 94(2), 428-434.

Craig, S., Driscoll, D. M., & Gholson, B. (2004). Constructing knowledge from dialog in an intelligent tutoring system: Interactive learning, vicarious learning, and pedagogical agents. Journal of Educational Multimedia and Hypermedia, 13(12), 163-183.

Cobb, T. (1997). Cognitive efficiency: Toward a revised theory of Media. *Educational Technology Research and Development*, *45(4)*, 21-35.

Cowan, N. (2001). The magical number 4 in short-term memory: A reconsideration of mental storage capacity. Behavioral and Brain Sciences, 24(1), 87 – 114.

Dehn, D. M. & van Mulken, S. (2000). The impact of animated interface agents: a review of empirical research. International Journal of Human-Computer Studies, 52, 1-22.

Erickson, T. (1997). Designing agents as if people mattered. In J. M. Bradshaw (Ed.), Software Agents (pp. 79-96). Menlo Park, CA: MIT Press.

Harp, S. F., & Mayer, R. E. (1997). The role of interest in learning from scientific text and illustrations, on the distinction between emotional interest and cognitive interest. Journal of Educational Psychology, 89(1), 92-102.

Harp, S. F., & Mayer, R. E. (1998). How seductive details do their damage, A theory of cognitive interest in science learning. Journal of Educational Psychology, 90, 414-434.

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, 11, 47-78.

Kalyuga, S., Chandler, P., & Sweller, J. (1999). Managing split-attention and redundancy in multimedia instruction. Applied Cognitive Psychology, 13, 351-371.

Kalyuga, S. & Sweller, J. (In Press). Measuring Knowledge to Optimize Cognitive Load Factors During Instruction. Journal of Educational Psychology.

Koda, T., & Maes, P. (1996). Agents with faces: the effect of personification. Proceedings of 5th IEEE International Workshop on Robot and Human Communication, Tsukuba, Japan, 189-194.

Langer, E. (1994). The illusion of calculated decisions. In R. C. Schank, & E. Langer (Eds.),

Beliefs, Reasoning, and Decision Making: Psycho-Logic in Honor of Bob Abelson. Hillsdale, NJ: Erlbaum.

Lester, J., Converse, S., Stone, B., Kahler, S. And Barlow, T. (1997). Animated pedagogical agents and problem solving effectiveness: A large scale empirical evaluation. Proceedings of 8th World Conference on Artificial Intelligence in Education, Kobe, Japan, 23-30.

Lester, J. C., Towns, S. G., & Fitzgerald, P. J. (1999). Achieving affective impact: Visual emotive communication in lifelike pedagogical agents. International Journal of Artificial Intelligence in Education, 10, 278-291

Levie, W. H., & Dickie, K. E. (1973). The analysis and application of media. In R. M. W. Travers (Ed.), Second handbook of research on teaching. Chicago: Rand McNally.

Levin, H. M. (1988). Accelerated schools for at-risk students. New Brunswick, NJ: Center for Policy Research in Education.

Levin, H. M, Glass, G., & Meister, G. R. (1987). Cost-effectiveness of computer assisted instruction. Evaluation Review, 11(1), 50-72.

Levin, H. M., & McEwan, P. J. (2001). Cost-effectiveness analysis: Methods and applications (2nd ed.). Thousand Oaks, CA: Sage.

Mayer, R. (2001). Multimedia learning. Cambridge, England: Cambridge University Press.

Mayer, R. E., Dow, G. T., & Mayer, S. (2002). Multimedia learning in an interactive self-explaining environment: What works in the design of agent-based micorworlds?. Journal of Educational Psychology, 95(4), 806-813

Mitrovic, A. & Suraweera, P. (2000). Evaluating an Animated Pedagogical Agent. Lecture notes in computer science, No 1839, 73-82.

Moreno, R., Mayer, R. E., Spires, H. A., & Lester, 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(2), 177-213.

Moundridou, M., & Virvou, M. (2002). Evaluating the persona effect of an interface agent in a tutoring system. Journal of Computer Assisted Learning, 18(3), 253-261

Nass, C., & Steuer, J. (1993). Anthropomorphism, agency, and ethopoeia: Computers as social actors. Human Communication Research, 19(4), 504-527.

Ogata, H., Liu, Y., Ochi, Y., & Yano, Y. (2001). Neclle: Network-based communicative language-learning environment focusing on communicative gaps. Computers & Education, 37, 225-140.

Paas, F.G.W.C. (1992). Training strategies for attaining transfer of problem solving skills in statistics: A cognitive load approach. Journal of Educational Psychology, 84, 429-434.

Paas, F.G.W.C. & Van Merrienboer, J.J.G. (1994). Measurement of cognitive load in instructional research. Perceptual and Motor Skills, 79, 419-430.

Pedhazur, E. J., & Schmelkin, L. P. (1991). Measurement, Design, and Analysis: An Integrated Approach. Hillsdale, NJ: Lawrence Erlbaum Associates, Publishers.

Person, N. K., Graesser, A. C., Kreuz, R. J., & Pomeroy, V. (2001). Simulating human tutor dialog moves in auto tutor. International Journal of Artificial Intelligence in Education, 12, 23-29.

Pintrich, P. R., & Schunk, D. H. (2002). Motivation in education: Theory, research and practice (2nd ed.). Englewood Cliffs, NJ: Prentice Hall.

Ross, S. M., & Morrison, G. R. (1989). In search of a happy medium in instructional technology research: Issues concerning external validity, media replications, and learner control. Educational Technology Research and Development, 37(1), 19-33.

Ross, S. M., & Morrison, G. R. (1996). Experimental Research Methods. In D. J. Jonassen (Ed.), Handbook of Research on Educational Communications and Technology (pp. 1148 - 1170). New York: Simon & Schuster Macmillan.

Ryokai, R., Vaucelle, C., & Cassell, J. (2003). Virtual peers as partners in storytelling and literacy learning. Journal of computer assisted learning, 19, 195-208.

Sampson, D., Karagiannidis, C., & Kinshuk. (2002). Personalised learning: Educational, technological and standardisation perspective. Interactive Educational Multimedia, 4, 24-39.

Shirey L. L., & Reynolds, R. E. (1988). Effect of interest on attention and learning. Journal of Educational Psychology, 80(2), 159-166.

Song, K., Hu, X., Olney, A., & Graesser, A. C. (2004). A framework of synthesizing tutoring conversation capability with web-based distance education courseware. Computers & Education, 42, 375-388.

Sproull, L. Subramani, M., Kiesler, S., Waler, J. H., & Waters, K. (1996). When the interface is a face. Human-Computer Interaction, 11, 97-124.

Sweller, J., & Chandler, P. (1994). Why some material is difficult to learn. Cognition and Instruction, 12, 185-233.

Takeuchi, A., & Naito, T. (1995). Situated facial displays: Towards social interaction. In I. Katz, R. Mack, L. marks, M. B. Rosson & J. Nielsen (Eds.), Human Factors in Computing Systems: CHI'95 Conference Proceedings (pp. 450-455). New York: ACM Press.

Van Merriënboer, J. J. G. (1997). Training complex cognitive skills: A four-component instructional design model for technical training. Englewood Cliffs, NJ: Educational Technology Publications.

van Mulken, S., André, E., & Müller, J. (1998). The persona effect: How substantial is it? In H. Johnson, L. Nigay & C. Roast (Eds.), People and Computers XIII: Proceedings of HCI'98 (pp.

53-66). Berlin: Springer.