Chapter 11
Sustained Learning in 4th and 5th Graders but not 7th Graders: Two Experiments with a Talking Pedagogical Agent

Bruce L. Mann
Memorial University, Canada

Henry Schulz
Memorial University, Canada

Jianping Cui
Bow Valley College, Canada

Shannon Adams
Brother Rice High School, Canada

ABSTRACT
In this chapter, agent movement and temporal speech cueing were designated for empirical study. In Experiment 1 an agent presented students in grades 4 and 5 (n = 133) with instruction, practice, and feedback on the proper usage of the apostrophe to show singular and plural ownership. Analyses of the data in Experiment 1 showed that modality effects favoured speech cueing over text cueing but agent animation had no effect. In Experiment 2, a different agent presented students in grade 7 (n = 91) with examples and practice questions on multiplying and dividing fractions. Experiment 2 data showed no effects for modality or agent animation. The data reflects previous findings of inconsistent effects in modality research.

DOI: 10.4018/978-1-4666-0137-6.ch011
INTRODUCTION

Pedagogical agents are computer-generated characters that can be programmed to move around the screen and provide advice to students in speech or text bubble. These animated characters can be expected to focus student attention on critical information in a computer program (Dehn & van Mulken, 2000) or even provide individualized scaffolding on an educational website (Lester, Towns, Callaway, Voerman, & FitzGerald, 2000; Moreno, Mayer, & Lester, 2000), though the optimal role for an agent in educational multimedia is unclear (Kim & Baylor, 2006).

LITERATURE REVIEW

In the absence of a comprehensive theory of learning from pedagogical agents, researchers must rely on relevant research in human psychology and education. Theory development in educational psychology has added insight to our understanding of learning from educational media including, though not limited to: that students comprehend oral and written discourse differently (Hildyard & Olson, 1982), fuzzy trace theory (Brainerd & Reyna, 1990), the interactive-compensatory hypothesis (Stanovitch, 1980), dual coding theory (Paivio, 1986), the separate streams hypothesis (Penney, 1989), the split-attention theory of multimedia learning (Chandler & Sweller, 1991), the model of working attention (Baddaley, 1992), the structured sound function model of instructional design (Mann, 1992, 1995, 1997a), the dual processing theory of working memory (Mayer & Moreno 1998), the cognitive theory of multimedia learning (Mayer, 1997, 2001), and the attentional control theory of multimedia learning (Mann, 2006, 2008a, 2009).

LISTENING AND READING IN ADULTS

From these advances and others, we know that when adults listen and look at educational multimedia, they integrate spatial and verbal sensations in their working memory for a short time as they generate meaningful relationships between the spatial store and the verbal (language) stores. Most adults can systematically and completely integrate information from listening and reading (Pressley & McCormick, 1995), by self-initiating an executive control of these different mental processes (listening and reading), as suggested in the literature (Halliday, 1987; Higginbotham-Wheat, 1991; Penney, 1989). During listening, adults acquire the gist or meaning from the auditory sensations (Hildyard & Olson, 1982; Penney, 1989; Reyna, 1992; Brainerd, 1993), and from reading text, acquire the details or surface features (Tannen, 1985), sometimes known as verbatim information learning (Martin & Briggs 1986; Penney, 1989). However when our attention is overloaded or distracted, features can be combined inappropriately. We know that students learn better when the instructional material does not require them to split their attention between multiple sources of mutually referring information (Chandler & Sweller, 1992; Mayer & Moreno, 1998; Mousavi, Low, & Sweller, 1995). Meaningful learning occurs when adults select relevant information in each store, organize the information in each store into a coherent representation, and make connections between corresponding representations in each store (Mayer, 1997).

AUTOMATIC AND CONTROLLED PROCESSING

Whenever adults read and listen to easy or familiar content, they use automatic processing (Schneider & Shiffrin, 1977) also known as pre-attentive processing (Treisman, 1986). Pre-attentive processing
Sustained Learning in 4th and 5th Graders but not 7th Graders

of easy or familiar content can occur in parallel; that is we can handle two or more idea elements simultaneously. Adults implement pre-attentive processing on easy tasks or with highly familiar items. On difficult tasks or with unfamiliar items however, adults will apply controlled processing (Schneider & Shiffrin, 1977); also known as attention focusing (Treisman, 1986). Attention focusing on divided-attention tasks is serial; only one task is handled at a time. Adults will use attention focusing on difficult or unfamiliar divided-attention tasks, once described as “the glue that binds the separate features of a stimulus- such as the colour and shape- into a unitary object” (Matlin, 1989, p. 57). When we pay attention to a visual stimulus, cerebral blood-flow studies show increased activity in the visual and parietal cortex (i.e., the top, back region of the brain) (Robinson & Peterson, 1986). When we shift our attention from visual to auditory modality our blood flow is increased in our prefrontal cortex (located at the top, front region of the brain) (Matlin, 1989). In this way, we gain and hold our own attention, whether it’s a response to an emergency message or the beginning stages of a lesson. Figure 1 illustrates two models of attention on a difficult or unfamiliar task, one showing (a) high mental effort associated with high cognitive load from an unbalanced input from both channels; the other (b) normalized mental effort associated with normal cognitive load from a balanced input from both channels.

Adults and older adolescents can usually be expected to examine stimuli systematically and completely, as suggested in Pressley and McCormick (1995). School-age students however, may have a different experience. At 12 years old the executive process that makes a child fully conscious of their problem solving abilities and allows him or her to relate prior knowledge to a current problem in a systematic way, is still ma-

Figure 1. Two models of attention on a difficult or unfamiliar task. Adapted from Mann, Newhouse, Pagram, Campbell, & Schulz (2002).
Sustained Learning in 4th and 5th Graders but not 7th Graders

turing. For this reason, student enjoyment of multimedia is either uncorrelated or negatively correlated with their learning outcome (Clark, 2001; Clark & Feldon, 2005). Frequently students find feedback following an error more interesting than feedback following a correct response (Ragsdale, 1988). “Many learners will not notice the option to read directions or will try to save time by skipping them” (Alessi & Trollip, 1991, p. 22). “While aesthetically pleasing, feedback provided in text will go unnoticed by the student” (Alessi & Trollip, 1991, p. 72). With increasing age and experience, a child’s processing becomes more efficient. By the time the child reaches adolescence, the executive process permits him or her to reason in a systematic and logical fashion (Mussen, Conger & Kagan, 1974).

This chapter addresses this persistent educational problem of ignoring or forgetting to read important instructions and feedback presented in text or other visual displays from a computer screen. The general goal was to determine whether school-aged children would learn from “agent animation” (movement or no movement) and “modality” (auditory or on-screen text). Agent “Genie” tutored 4th and 5th graders on the proper usage of the apostrophe to show singular and plural ownership. Agent “Peedy” tutored 7th grade students on the multiplication or division of fractions. The interfaces were designed so that the agents’ movements and speech prompts would activate associations between relevant prior knowledge and the new information. It was expected that these young learners would retain more knowledge, and find more creative solutions to problems than students using agents that did not move or talk.

At present, recent findings suggest that some school-aged students using educational multimedia are unable to generate sufficient gist to solve problems; that their under-developed biological capacity to extract gist from speech limits their mental ability to generate sufficient referential connections between the speech prompts and the limited text, and the speech prompts and diagrams. For example some researchers (Shilling, 1991; Weiner, 1991) believe that positive findings from using speech in educational multimedia may only be generalizable to particular populations of learners, such as adults and older adolescents. Shilling’s (1991) study with eighty-one kindergartners reported statistically non-significant learning effects between students using conventional writing materials and/or computers with and without available synthesized speech feedback over an eight month period. Similarly, Wiener (1989) investigated the differential effects of presentation conditions; visual only and visuals cued with speech, on sight-word learning with fifty-five handicapped third graders. The results indicated non-significant statistical differences between the presentation conditions. Two years later however, Weiner (1991) reported different results with twenty-four junior-high students. A recent study with 7th graders in Australia (Mann, Newhouse, Pagram, Campbell, & Schulz, 2002) reported no differences between speech cueing and on-screen text cueing. Witteman and Segers (2010) speculated that the experimental materials in their study were not difficult enough to split attention, and may have accounted for what they called “a reversed modality effect,” though it may have more likely been attributable to an expertise reversal effect (Kalyuga, Ayres, Chandler, & Sweller, 2003; Kalyuga, Chandler, & Sweller, 1999). In a within-subjects design by Segers et al (2008) children received the same four lessons in the same order but in a different format that was presented in a mixed research design: (1) written information only; (2) written information accompanied by representational pictures; (3) oral information only or (4) oral information accompanied by representational pictures. Analysis of the data showed that differences between the presentation conditions were statistically non-significant. It is possible that the design of the study itself contaminated the visual-only treatment with memory...
traces from the speech treatment, as suggested previously (Mann, 1997a). Repeated measures in speech research should only be implemented by occasion, and then only after time lapses between occasions.

**DESIGN UNCERTAINTY**

Most of the published research on multimedia learning calls for clearer directions on whether audio should replace or enhance onscreen instruction and feedback (Barron & Kysilka 1993; Koroghlanian & Klein, 2004)—an audio model for computing (Buxton, 1989). Some authors want purposeful advice with a strong cognitive foundation (Mayer, 1997), others a pragmatic emphasis (Reiser & Gagne, 1983; Smaldino & Lowther, 2007), still others straight narration (Mayer, 2001) or something more than narrated screen text (Bishop, Amankwatia, & Cates, 2008), or something beyond word lists (Gyselinck, Jamet, & Dubois, 2008). Most researchers want “something more,” without saying what something more entails or how this might be achieved. We believe that as long as the function of the audio with the onscreen visuals is left unclear, researchers will keep asking for advice for designing multimedia instruction. Aside from the declared need for purposeful design advice, there is the issue of testing. Multimedia materials are usually tested in factorial experiments to determine their immediate impact (e.g., see Mayer, 2001, for a review), instead of repeated measures to determine the more lasting effects of the treatments (Mann, 1994, 1995, 1997a; Segers, Verhoeven, & Hulstein-Hendrikse, 2008).

**DESIGN HEURISTICS**

Researchers typically use one of eight different design heuristics for integrating audio into computer-based animations, graphics and onscreen text for learning (Mann, 2008; 2009a). Each heuristic differs from the others in its scope or depth of advice, and carries a different assumption about how people learn from multimedia. See Mann (2009) for details about each of these design heuristics and corresponding assumptions about multimedia learning. Table 1 outlines the eight different design heuristics for enhancing animations, graphics and onscreen text with audio for learning.

<table>
<thead>
<tr>
<th>Design Heuristic</th>
<th>Underlying Assumption</th>
</tr>
</thead>
<tbody>
<tr>
<td>Structured Sound Function</td>
<td>Learning occurs when student attention is self-controlled, where the gist is heard and the details read</td>
</tr>
<tr>
<td>Cognitive Load</td>
<td>Learning occurs when extrinsic load is reduced and germane load is sufficient</td>
</tr>
<tr>
<td>Maximum Impact</td>
<td>Learning occurs when auditory and visual stimuli are sufficiently strong</td>
</tr>
<tr>
<td>Balanced Input</td>
<td>Learning occurs when the logogens and imagens make referential connections</td>
</tr>
<tr>
<td>Favorite Method</td>
<td>Learning occurs when the right instructional method is used</td>
</tr>
<tr>
<td>Favorite Feature</td>
<td>Learning occurs when the right technology is used</td>
</tr>
<tr>
<td>Design-By-Type</td>
<td>Learning occurs when an instructional software taxonomy is followed</td>
</tr>
<tr>
<td>Whatever Works</td>
<td>Learning occurs when the developer’s intuition is considered</td>
</tr>
</tbody>
</table>

Table 1. A summary of design heuristics and corresponding underlying assumptions about how people learn from multimedia

Adapted from Mann (2009)
The present research is only concerned with the Structured Sound Function (SSF) model of instructional design. The SSF model has five functions and three structures for relating the sound with the visual events. Figure 2 illustrates the structured sound function (SSF) model for designing the modality of instruction with *convergent temporal sound cueing* highlighted.

Convergent temporal sound cues promote thinking by focusing the learner’s attention in a stepwise procedure toward a specific solution (Mann, 2006, 2008, 2009b) that can set the stage or serve as a signal for specific behaviors to take place (Burton, Moore, & Magliaro, 2004). During convergent goal setting for example, the learner is encouraged to use a variety of sources to solve a problem (e.g., answer look-up) to produce an acceptable result. Selecting the goal and constancy for a sound cue for a visual event is the most important of the three structural components in the SSF model (Mann, 2006).

![Figure 2. The structured sound function (SSF) model for designing the modality of instruction. Adapted from Mann (2008).](image)

<table>
<thead>
<tr>
<th>&lt; <em>Giving the sound a function</em> &gt;</th>
<th>&lt; <em>Structuring the sound with a visual event</em> &gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>The Goal:</strong></td>
<td><strong>The Constancy:</strong></td>
</tr>
<tr>
<td>is convergent or divergent</td>
<td>is continuous or discontinuous</td>
</tr>
<tr>
<td><strong>The Density:</strong></td>
<td></td>
</tr>
<tr>
<td>is massed, spaced or summarized</td>
<td></td>
</tr>
<tr>
<td><strong>A Temporal Prompt:</strong></td>
<td></td>
</tr>
<tr>
<td>that cues that counterpoints that dominates that undermines</td>
<td>Continuous Convergent Temporal Sound Cueing during difficult tasks or with unfamiliar items can focus student attention on the critical visual events</td>
</tr>
<tr>
<td><strong>A Point of View:</strong></td>
<td></td>
</tr>
<tr>
<td>objective, subjective, performer, political, socio-cultural</td>
<td></td>
</tr>
<tr>
<td><strong>Locale:</strong></td>
<td></td>
</tr>
<tr>
<td>real, imaginary</td>
<td></td>
</tr>
<tr>
<td><strong>Atmosphere:</strong></td>
<td></td>
</tr>
<tr>
<td>feeling, mood</td>
<td></td>
</tr>
<tr>
<td><strong>Character’s:</strong></td>
<td></td>
</tr>
<tr>
<td>past, future, personality</td>
<td></td>
</tr>
</tbody>
</table>
The theoretical foundation for the SSF model is the Attentional Control Theory of Multimedia Learning (ACTML) (Mann, 2008a). The ACTML is based on two well-established theories: Baddeley’s (2002) revised model of working attention and Brainerd and Reyna’s (2001) Fuzzy-Trace Theory of gist and verbatim memory. Figure 3 illustrates the structure and process of working attention of students in the Sound groups, according to the Attentional Control Theory of Multimedia Learning.

Figure 3. The attentional control theory of multimedia learning. Adapted from Mann (2008a).
The ACTML has two uses for the instructional designer and interface developer. First, the ACTML illustrates a structure of the learner’s working attention to multimedia, by all accounts a more apt description and accurate label than ‘working memory’ (Baddeley, 2002; Neath, 2000). Baddeley’s framework is useful for conducting experiments that identify the processes involved during the integration of information coming from multiple sources (Gyselinck, Jamet, & Dubois, 2008).

Second, the ACTML describes how learners process information in different modalities differentially (Halliday, 1987; Penney, 1989; Higginbotham-Wheat, 1991), consistent with fuzzy-trace theory (Brainerd & Reyna, 1993, 2001, 2004; Reyna & Brainerd, 1990, 1994; Wolfe, Reyna & Brainerd, 2005). Whereas listeners will attend to the gist or meaning (Hildyard & Olson, 1982; Penney, 1989; Brainerd & Reyna, 1990; Reyna, 1992), readers will attend to details or surface features of the text (Tannen, 1985), also called verbatim information (Martin & Briggs, 1986; Penney, 1989). Hildyard & Olson (1982) reported that performances were better for listening to highlights (gist) and reading the details (substance, verbatim). Memory for the gist of verbal information does not include the exact words that were presented, but instead refers to the meaning those words convey. Similarly, memory for the gist of numerical information does not include memory for the exact numbers that were presented, but instead refers to the patterns or relationships that those numbers convey, their “meaning.” In everyday language, when we say a person remembers the gist of an event, we mean that they have remembered the substance of the event, namely its semantic content. In this paper such terms as “sense” or “meaning” are used to refer to this content. When verbal information is involved, this aspect of the term “gist” is straightforward, and, again, conforms to traditional psycholinguistic usage.

**EXPERIMENT 1**

Most of the previous research with a talking pedagogical agent was conducted with adult or adolescent learners, mostly undergraduate psychology majors or pre-service teachers. New agent research was needed with school-age students. The purpose of Experiment 1 therefore, was to determine whether or not 4th and 5th grade students would learn more about the proper usage of apostrophe from a moving and talking pedagogical agent than students using agents that did not move or talk. Neither the presence/absence of an agent, nor the number of agents present were studied in this research, as “agent presence” was held constant across all learning conditions. Our hypotheses were as follows:

1. First, it was expected that fourth and fifth graders who heard the voice of an agent speaking convergent temporal speech cues would be more likely to remember the principles of singular and plural ownership than those who saw the agent and read the on-screen convergent temporal text cues,

2. Second, it was expected that fourth and fifth graders who heard the voice of an agent speaking convergent temporal speech cues would be more likely to solve problems of singular and plural ownership than those who saw the agent and read the on-screen convergent temporal text cues,

3. Third, it was expected that fourth and fifth graders who saw the agent fly-in and gesture with its outstretched arm, scratch its head, do a magic trick, and fly to a different position on the screen would be more likely to remember the principles of singular and plural ownership than those who saw the Agent fly-in and remain motionless, with no gesturing, no changing position on the screen, and no magic tricks.
Participants in Experiment 1

The participants \( n = 133 \) in this study were 4\textsuperscript{th} and 5\textsuperscript{th} grade students (aged 9-12 years) enrolled at Elementary school in a small city in Eastern Canada, and had consent forms signed by a parent or guardian. Students from the 4\textsuperscript{th} and 5\textsuperscript{th} grades were selected for the study because:

1. None of the participants had been introduced to singular and plural ownership, proper usage of apostrophe prior to, nor during, the experiment.
2. Although they had been exposed to educational multimedia in the school setting, none of these children had seen an Agent.
3. This was a comparatively large sample of children from the two grades.

“Children at this level are eager to cooperate in experiments and are able to persist at them for long periods of time” (Case, 1985, p. 182). The teachers and school Principal were enthusiastic about the research and supportive in every way because literacy levels of the students at this school were below the national average (Maguire, 2001). The participants with proper consent were randomly assigned to one of the four treatment groups: Movement with Speech (M, S), No Movement with Speech (NM, S), Movement with Text (M, T), and No Movement with Text (M, T).

Materials in Experiment 1

Commercial software was not used in the study to reduce the risk of students having had prior access to the content in the treatments. Rather, an educational multimedia website was developed by the researchers specifically designed for grade 4 and 5 students. The multimedia treatments were then pilot tested with a few of the students. The students who were recruited for the pilot testing were subsequently excluded from the study. The treatments were then administered to participating students using the school’s course management system, and accessed from separate computer labs (on different floors of the building), unlike previous studies (Adams et al., 2006; Tabbers et al., 2004) where the treatments were administered consecutively in modified classrooms. The multimedia treatments in this study were administered by the random order generator in the Presentations Tool, and the tests by the random order generator in the Quiz Tool within the course management system. This helped to prevent participating students working together in the same lab from copying from one another. Furthermore, the sound-cued groups and text-cued groups accessed their respective treatments from separate computer labs within the school. This was a fairly normal situation for students in that school where classes of fourth and fifth grade students frequently shared the labs in this way to complete their assignments.

A Microsoft agent called “Genie the Magician” cued the visual events on a customized educational website, the files were uploaded to an independent server, to prevent accidental prior access to the content by the participating students. The agent character was implemented as an ‘advisor’ (Kim & Baylor, 2006) flying-in from the top left hand corner of the screen to an initial position on the top left of the screen in physical proximity to the onscreen graphics and text boxes, in keeping with the spatial contiguity principle (Moreno & Mayer, 1999). Four distinct guided-tutorials were accessible from four separate Internet addresses.

In the on-screen text cues with no agent movement (NMT) condition, Genie flew in and remained motionless. Genie spoke as silent text bubble appearing above its head. But there was no gesturing to the text, no changing position on the screen, and no magic tricks.

In the on-screen text cues with agent movement (MT) condition (shown in Figure 4), Genie flew in and gestured to the text on the screen with its outstretched arm. The prompts appeared as text bubble in a text box, with identical information to the speech condition. The Genie’s movements
in the MT condition were identical to those in the MS condition.

In the speech cues with agent movement (MS) condition the Genie flew in, immediately gestured to the text on the screen with its outstretched arm, and spoke speech cues in a synthesized voice about a future event (direction, instruction hint) or a past event (feedback or reminder). Genie would scratch his head, do a magic trick, and fly to a different position on the screen. The agent was programmed to speak only the gist (not explain or narrate), leaving the details in text and illustrations.

In the speech cues with no agent movement (NMS) condition, the Genie flew in and thereafter remained motionless. The Genie in the NMS condition spoke the same temporal speech prompts as the Genie in the MS condition. But there was no gesturing to the text, no changing position on the screen, and no magic tricks. Scripting the agent’s temporal speech cues was developed with the primary aim of directing attention to the details in the onscreen text, graphic or animation on the screen, as suggested in the Structured Sound Function (SSF) Model (Mann, 2008a).

Participating students were asked to discriminate the correct format of the possessive expression from the incorrect ones. Further, students were asked to transfer the knowledge of the correct usage of apostrophe to make up sentences according to different graphics and pictures. The structure and content of the program were held constant across the treatment conditions as the animation and modality were varied. The language on the

Figure 4. A non-example showing the talking agent cueing the student with a verbal caution about singular ownership

![Image](image-url)
website was directed at the fourth graders and designed to assist their learning, as suggested by Tomasello and Brooks (1999). A literacy expert examined the content and judged it to be appropriate and interesting for fourth and fifth grade students. Following a quality review, the website was revised to eliminate clutter thereby reducing extraneous cognitive load, compatible with the coherence principle (Moreno & Mayer, 2000) and cognitive load theory (Pollock, Chandler, & Sweller, 2002). Two of the school’s best-equipped computer labs were made available for the experiment. One lab was set-up for students in the MT and NMT conditions to watch a pedagogical agent called “Genie,” and read the on-screen instruction and feedback involving the proper usage of apostrophe to show singular and plural ownership. The other computer lab was configured for students in the MS and NMS conditions to watch agent “Genie” and to listen to equivalent instruction and feedback through headphones. In both labs, and prior to administration of the treatment, the researcher used a big screen television connected to a computer to demonstrate how students should navigate the program.

**Instruments in Experiment 1**

A single 12-item test on correct apostrophe usage was developed and pilot-tested in the WebCT quiz tool with 4 students from the target grades levels. Some items were revised, based on their feedback. The items were administered in each testing occasion (i.e., pre-test, immediate post-test, delayed post-test) by the random order generator in the WebCT quiz tool. The test was graded out of total of 30 possible points and checked by researchers. There were 10 multiple choice questions each worth 2 points that asked the participant to choose which sentence was correct, and 2 short-answer questions each worth 5 points that asked the participant to type a sentence showing plural ownership based on a picture. The four experimental conditions in the study were deemed equivalent. The probability that this *read-then-type requirement* might disadvantage speech-cued participants, as suggested recently (Segers, Verhoeven, & Hulstein-Hendrikse, 2008) was very low, for two reasons. First, speech-cued participants had their cognitive load shared between auditory and visual memory systems. Second, speech-cued participants were required to process only the gist portion by listening to speech-cues, and the details portion only by reading the text.

**Design of Experiment 1**

The study used a 3-factor repeated measures (pre-test, immediate post-test, delayed post-test) design with three independent between-subjects factors: agent modality (speech cues vs. text cues), agent animation (movement or no movement), and student grade level (grade 4 or grade 5). We included the delayed post-test in this study to test the content of long-term memory over time, consistent with the Attentional Control Theory of Multimedia Learning (ACTML), which links focusing attention in the executive controller with long-term memory, compatible with the view (Sweller, 2004) that the main purpose of instruction is to build knowledge in long-term memory.

**Procedure in Experiment 1**

The experiment was conducted in three scheduled sessions over a 10-week period. From the initial group of 162 fourth and fifth grade students that were given consent forms to have signed by a parent or guardian, 133 were returned with consent. For each experimental session, the teachers brought the participating students into the labs. Their names were coded and then anonymously and randomly assigned to one of the treatment groups, either: speech cues with agent movement (M, S), speech cues with no agent movement (NM, S), on-screen text cues with agent movement (M, T), or on-screen text cues with no agent movement (NM, T). In a separate session,
133 students with signed consent forms took the pre-test of apostrophe usage and interpretation of rules. Administration of the pre-test took about 20 minutes. Four weeks later, these participating students took the treatment and immediate post-test. Administration of treatment and immediate post-test took about 50 minutes. Six weeks later, the participating students took the delayed post-test, which lasted about 20 minutes.

**Data Analysis in Experiment 1**

An analysis of variance was conducted on the pre-test means of the two grade levels crossed with the four treatment groups. It indicated that the groups did not differ significantly in their prior knowledge, \( F(7,129) = 1.19 \) \( (p = .312) \). This indicated that the eight groups were similar with low prior knowledge of apostrophe usage, and the means further suggested that knowledge was uniformly low (see Table 2 for means). Low prior knowledge was a necessary pre-condition in this study, as explained earlier. Statistical imputations were then conducted by multiple regression with a stochastic component on cases that showed a participating student had taken the treatment but was absent for a test, in accordance with recommended procedures (Baraldi & Enders, 2010; Little & Rubin, 1990; Smits, Mellenbergh, & Vorst, 2002). This was done so that there would be no loss of data and to maintain the initial group similarity. There were 11 missing cases on the pre-test and 4 on the delayed post-test, and these were evenly distributed across the treatment groups (from 2 to 4 imputations per group on the pre-test, and from 0 to 2 on the delayed post-test). These procedures permitted analysis of a complete data set that reflected the expected characteristics of the scores (Enders, 2010; Little, & Rubin, 2002). This resulted in complete data for 133 participating students (see Table 2 for numbers in each group).

The design therefore, consisted of a within-subjects factor (the three repeated measurements: pre-, post-, and delayed test) crossed with the three between-subjects factors (agent animation, modality, and grade level). Treatment effects of animation and modality would be evidenced by interactions with the repeated measures. There was only one statistically significant interaction for the within-subjects analysis: the interaction of the repeated measures with modality, \( F(2, 250) = 5.02 \) \( (p = .007) \). While significant, it should be noted that this interaction effect accounted for approximately 4% of the variance attributed to the within-subjects part of the analysis, and should therefore be described as a small to medium effect. There was a significant within-subjects main effect due to the repeated measures, \( F(2, 250) = 22.29 \) \( (p = .000) \), and a significant between-subjects effect due to modality \( F(1, 125) = 21.54 \) \( (p = .000) \). While these main effects were significant, they were not pursued directly as the interaction noted above was significant, and it included both of these factors (see Table 3).

Simple main effects were tested for the three testing times within the Speech group and within the Text group to determine which of these differences were significant using procedures outlined for repeated-measures analyses. There was a significant change over time for the Speech group accounting for over 27% of the variance associated with this group, \( F(2, 130) = 24.51 \) \( (p = .000) \). Further post hoc tests using the Bonferroni critical value of.017 for the three pair wise tests indicated that all three pairs of means differed significantly. The pre-test and posttest means \( (p = .000) \) had a medium to large effect size of.76, the pre-test and delayed posttest means \( (p = .001) \) had a medium effect size of.43, and the posttest and delayed posttest means \( (p = .001) \) had a small effect size of.31.

The simple main effect results were also significant for the Text group but not as strong, accounting for only approximately 5% of the variance associated with this group on the three tests, \( F(2, 132) = 3.27 \) \( (p = .041) \). Post hoc tests using the critical value of.017 could not detect where these differences occurred, pre-test and
Sustained Learning in 4th and 5th Graders but not 7th Graders

Analysis of the data from Experiment 1 showed that 4th and 5th grade students learned the proper usage of the apostrophe better from animated pedagogical agents when critical information was presented as speech rather than on-screen text. However, movement of the agent around the computer screen did not significantly affect student learning.

Summary of Experiment 1

Analyses of the data from Experiment 1 showed that 4th and 5th grade students learned the proper usage of the apostrophe better from animated pedagogical agents when critical information was presented as speech rather than on-screen text. However, movement of the agent around the computer screen did not significantly affect student learning.
**Table 3. Apostrophe analysis of variance: repeated measures**

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>F</th>
<th>η²</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent animation</td>
<td>1</td>
<td>1.74</td>
<td>.014</td>
<td>.190</td>
</tr>
<tr>
<td>Modality</td>
<td>1</td>
<td>21.54</td>
<td>.147</td>
<td>.000</td>
</tr>
<tr>
<td>Grade</td>
<td>1</td>
<td>3.20</td>
<td>.025</td>
<td>.076</td>
</tr>
<tr>
<td>Agent animation * Modality</td>
<td>1</td>
<td>0.47</td>
<td>.004</td>
<td>.496</td>
</tr>
<tr>
<td>Agent animation * Grade</td>
<td>1</td>
<td>0.28</td>
<td>.002</td>
<td>.597</td>
</tr>
<tr>
<td>Modality * Grade</td>
<td>1</td>
<td>0.26</td>
<td>.002</td>
<td>.609</td>
</tr>
<tr>
<td>Agent animation * Modality * Grade</td>
<td>1</td>
<td>0.39</td>
<td>.003</td>
<td>.532</td>
</tr>
<tr>
<td>Error</td>
<td>125</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Post-Delay</td>
<td>2</td>
<td>22.29</td>
<td>.151</td>
<td>.000</td>
</tr>
<tr>
<td>Pre-Post-Delay * Agent animation</td>
<td>2</td>
<td>1.46</td>
<td>.012</td>
<td>.233</td>
</tr>
<tr>
<td>Pre-Post-Delay * Modality</td>
<td>2</td>
<td>5.02</td>
<td>.039</td>
<td>.007</td>
</tr>
<tr>
<td>Pre-Post-Delay * Grade</td>
<td>2</td>
<td>2.10</td>
<td>.017</td>
<td>.125</td>
</tr>
<tr>
<td>Pre-Post-Delay * Agent animation * Modality</td>
<td>2</td>
<td>0.08</td>
<td>.001</td>
<td>.922</td>
</tr>
<tr>
<td>Pre-Post-Delay * Agent animation * Grade</td>
<td>2</td>
<td>2.69</td>
<td>.021</td>
<td>.070</td>
</tr>
<tr>
<td>Pre-Post-Delay * Modality * Grade</td>
<td>2</td>
<td>0.09</td>
<td>.001</td>
<td>.913</td>
</tr>
<tr>
<td>Pre-Post-Delay * Agent animation * Modality * Grade</td>
<td>2</td>
<td>0.13</td>
<td>.001</td>
<td>.878</td>
</tr>
<tr>
<td>Error (within subjects)</td>
<td>250</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**EXPERIMENT 2**

The purpose of Experiment 2 was to determine whether this pattern of results could be repeated with 7th grade students learning the multiplication and division of fractions.

**Participants in Experiment 2**

The participants (n = 91) in this study were 7th grade students (12-15 years old) enrolled at a public High School in Eastern Canada with consent forms signed by a parent or guardian. They had not been introduced to multiplication or division of fractions. These participants were randomly assigned to one of the four treatment groups: Movement with Speech (M, S), No Movement with Speech (NM, S), Movement with Text (M, T), and No Movement with Text (M, T).

**Materials in Experiment 2**

A prototype was needed to help students learn and apply the basic rules of multiplication and division of fractions. Initially several mathematics software titles were reviewed for their math content. All of the programs reviewed were well done but did not deal exclusively with fractions, since fractions problems may only come-up every tenth problem. This is where the students needed the most practice. An original prototype was developed as a lesson enhancement to help students learn and apply the basic rules of multiplication and division of fractions using a tutorial loaded on the school’s website, that incorporated progressively more difficult fractions problems to solve and a talking pedagogical agent called “Peedy.”

The Peedy character has been used in Atkinson’s (2002) research with animated pedagogical agents.
The talking agent was used to activate associations between relevant prior knowledge and new information. The agent’s audio featured temporal speech cues (Mann, 1996, 2000). Several educators and students examined the prototype:

1. A Subject Matter Expert (SME) had been the Math Department Head for the past three years and had been a mathematics teacher for the past twelve years at Brother Rice Junior High School. The SME’s role in this quality review was to look at the content and grammar used in the fractions prototype.

2. This Instructional Design (ID) expert was currently employed as an environmental engineer, working at instructional design and web development on a regular basis. The ID expert’s role was to review the prototype with the following aspects in mind: ease of use of program, appearance and flow of program, coherency of prototype, and visible features including audio and animation.

3. Two students from the target audience from grade eight with average math background reviewed the prototype. The students’ role was to see if they liked the prototype, and if they could easily progress through the program on their own with little direction. The student reviews determined that the content would be of benefit to students aged 12-13 who had just introduced to the rules of multiplying and dividing fractions (see Figure 6).

In 1997, Lester and colleagues (Lester, Converse, Kahler, Barlow, Stone & Bhogal, 1997) developed a similar math prototype for 100 seventh graders. The animated pedagogical agent...
Sustained Learning in 4th and 5th Graders but not 7th Graders

was “Herman the Bug.” The students were randomly assigned to one of five groups. Each group consisted of 20 students with a cloned “Herman.” The five Hermans differed from one other with respect to their modes of expression and in the level of advice they offered in response to students’ problem-solving activities.

**Instruments in Experiment 2**

One fifteen-question test of student knowledge and skill with multiplication and division of fractions was administered in WebCT, and used as the pre-test, immediate post-test, and delayed post-test. The dependent variable was operationalized as a test of near-transfer items structurally-similar to items presented during instruction, in keeping with Atkinson (2002), namely: 1) concept learning: students were asked to recall the rule for the multiplication and division of fractions to the question “What is the rule for the division of fractions?” and, 2) rule-using: students’ application of the rule for the multiplication and division of fractions including conversion to simplest form, as in “What is the best answer for $1/3 \times 2/4 - 2/12, 1/6, 6/1, or 6 ? ”

**Design of Experiment 2**

The design of this study was the same as that in Experiment 1, namely: a repeated measures ANOVA with two independent factors, “agent animation” (movement or no movement) and “modality” (auditory or on-screen text information)
Sustained Learning in 4th and 5th Graders but not 7th Graders

The fractions test was administered prior to the treatment, pre-test, upon completion of the treatment, immediate post-test, and six weeks later, delayed post-test. The pre-test was similar to Atkinson’s (2002) 11 item pre-test.

Procedure in Experiment 2

The procedures in this study followed those suggested in previous studies of this kind (Brown & Mann, 2001; Mann, et al., 2002). A number of activities were done simultaneously and the final schedule established at the mutual convenience of the participants and the investigators. The entire process lasted just under an hour and comprised the following steps:

1. Written consent was obtained from the parents, from the School Board, the School Principal, from the Regional Superintendent, and from the Interdisciplinary Committee on Ethics in Human Research at the university.
2. Next, all grade 4 and 5 students with proper consent were treated as one sample with each student being randomly assigned to one of the four treatment groups: Movement with Speech, No Movement with Speech, Movement with Text, and No Movement with Text.
3. One week later, a test of the usage and interpretation of rules was administered in WebCT as a pre-test.
4. Four weeks later, students were administered the treatment and immediate post-test.
5. Six weeks later, students were administered the delayed post-test.

Results in Experiment 2

Whereas 96 students were present for the pre-test, 92 for the treatment and immediate post-test, and 96 for the delayed post-test, complete data were collected from only 79 students. Some students had missed one or more testing sessions. Scores were imputed for students who obtained the treatment and immediate post-test and missed either the pre-test or the delayed post-test—7 were imputed on the pre-test and 5 on the delayed post-test. Statistical imputation of the missing data yielded complete scores for 91 students (see Figure 7). Procedures for imputation were as described above.

Preliminary ANOVA on the fractions pre-test means of the four treatment groups indicated that they did not differ significantly, \( F(3, 87) = .37 \) (\( p = .772 \)). The groups were nearly equivalent in their prior knowledge of fraction multiplication and division. Figure 8 depicts the means for the four groups over the three testing administrations of the fractions test, and Figure 7 presents the descriptive statistics.

The lines in Figure 8 show that the changes between pre-test, immediate post-test and delayed post-test indicate substantial gains in learning from pre-test to immediate post-test, but that this learning was not maintained over the delayed post-test.

While it appears in Figure 8 that the speech movement group performed higher on the immediate post-test than did any of the other groups, this was not born out in the analysis; there was no significant interaction effects of modality or of modality and agent animation with pre-testing, immediate post-testing and delayed post-testing (see Figure 9).

The results were similar for the four groups of grade 7’s so that all four conditions could be readily collapsed to show the magnitude of change in student performance over time. The wide range of scores, particularly on the immediate post-test, served to obscure this apparent difference making

<table>
<thead>
<tr>
<th>Table 4. Research design to investigate modality</th>
</tr>
</thead>
<tbody>
<tr>
<td>Speech cues (S)</td>
</tr>
<tr>
<td>Speech cues (S)</td>
</tr>
<tr>
<td>Text cues (T)</td>
</tr>
</tbody>
</table>
it less than statistically significant. Since no interactions were significant, further analysis was conducted on all four groups combined to determine which pre-test, post-test delayed post-test differences were significant.

Figure 8 illustrates the deterioration in the abilities in all four treatment groups to multiply fractions on the delayed post-test. One plausible explanation for the deterioration in performance across the board may be attributed to events occurring in the school during the day the delayed post-test. That day, the whole school turned out to see their school friends leave to return home to France. A sad occasion to be sure. This event however, affected all four treatments equally and therefore did not account for the 15 point drop in the “speech movement” scores.

### Summary of Experiment 2

Learning fractions by the seventh graders in Experiment 2 was similar for all four conditions, namely: dramatic improvement in performance for all four groups, but dropping-off just as dramatically. Furthermore apparently student learning did not maintain over time, as indicated in the delayed post-testing.

### GENERAL DISCUSSION

The aim of this research was to determine whether convergent temporal speech cues spoken by a moving pedagogical agent would help young children learn from educational multimedia.
Sustained Learning in 4th and 5th Graders but not 7th Graders

Figure 8. Three administrations of a test on “multiplying fractions” with 7th graders

Figure 9. ANOVA indicating between- and within-subjects effects on the fractions test

<table>
<thead>
<tr>
<th>Source</th>
<th>df</th>
<th>MS</th>
<th>F</th>
<th>$\eta^2$</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Between subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Agent animation</td>
<td>1</td>
<td>98.40</td>
<td>0.28</td>
<td>.003</td>
<td>.598</td>
</tr>
<tr>
<td>Modality</td>
<td>1</td>
<td>4.29</td>
<td>0.01</td>
<td>.000</td>
<td>.912</td>
</tr>
<tr>
<td>Agent animation * Modality</td>
<td>1</td>
<td>184.90</td>
<td>0.52</td>
<td>.006</td>
<td>.471</td>
</tr>
<tr>
<td><strong>Error</strong></td>
<td>87</td>
<td>352.27</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Within subjects</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pre-Post-Ret</td>
<td>2</td>
<td>2827.18</td>
<td>20.67</td>
<td>.192</td>
<td>.000</td>
</tr>
<tr>
<td>Pre-Post-Ret * Agent animation</td>
<td>2</td>
<td>31.84</td>
<td>0.23</td>
<td>.003</td>
<td>.793</td>
</tr>
<tr>
<td>Pre-Post-Ret * Modality</td>
<td>2</td>
<td>112.86</td>
<td>0.83</td>
<td>.009</td>
<td>.440</td>
</tr>
<tr>
<td>Pre-Post-Ret * Agent animation * Modality</td>
<td>2</td>
<td>146.66</td>
<td>1.07</td>
<td>.012</td>
<td>.345</td>
</tr>
<tr>
<td>Error (within subjects)</td>
<td>174</td>
<td>136.80</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Data from the 4th and 5th Graders

Analysis of the data in Experiment 1 confirmed our expectations with 4th and 5th graders in previous studies, namely that cueing 4th and 5th graders with convergent temporal sound from a talking agent would help them to learn proper use of the apostrophe. In general, the results for cueing with speech are in line with those in Segers, Verhoeven & Hulstein-Hendrikse (2008) who found a modality effect with fifth graders, and akin to results in previous modality research with adults (Chandler & Sweller, 1991; Mann, et al., 2002; Mann, 1995; Mann, 1997a; Mayer, 1997; Moreno & Mayer, 1999).

The pre-test scores. Pre-test scores of the children in the moving agent text group in Experiment 1 were lower than those in the moving agent with speech group. However these differences were statistically non-significant. Similarly the pre-test scores of fourth and fifth graders in the non-moving agent text group were lower than those in the non-moving agent with speech group. Again the differences between the groups were not statistically significant. This finding of statistical non-significance means that the existing knowledge of apostrophe usage in all eight groups of children was uniform, and in fact, was uniformly low. Low prior knowledge of apostrophe usage was a necessary pre-condition in this study, as explained earlier.

The post-treatment results. Analyses of the post-treatment data in Experiment 1 showed that the children in both groups learned more grammar, with simple main effects that were statistically significant in both Speech and Text groups, though the patterns of change from pre- to post- and delayed posttest were different for Speech and Text. The pattern for the Speech group appears distinctly different from that of the Text group, the Speech group means changing from 10.95 on the pre-test to 15.73 on the posttest and 13.65 on the delayed posttest, whereas the corresponding Text group means changing from 8.80 to 10.60 and 9.56 (see Figure 3). Means for the speech group were highest on the posttest, followed by the delayed posttest, and then the pre-test. Thus, significant learning occurred in this group, and this learning was maintained to the delayed posttest but with some loss.

Agent movement. Results on agent motion however, were not supported. These results contradict previous research with adults in which agent movement and gesturing captured attention without breaking the learner’s train of thought (Craig, Gholson & Driscoll, 2002; Johnson, Rickel & Lester, 2000; Veronikas & Shaughnessy, 2005). Furthermore, the moving agent in this study did not get any better results from that of the agent that did not move.

Data from the 7th Graders

Analyses of the data from Experiment 2 yielded quite different results, yet also coincided with the results of a previous research with 7th graders using speech prompts in educational multimedia (Mann, et al., 2002). Agent movement was non-significant. These 7th graders did not learn any better from speech cueing than from text cueing. Whereas the children in the sound group appeared to be more absorbed in the experience, they were unable to formulate coherent understandings from the media mix. More research is needed to determine how 7th graders’ learn to control the frequency and strength of associations between brief sentences on the screen and the spoken prompts, and between the animations and the spoken prompts to formulate coherent understanding from the presentations.

Implications from Experiments 1 and 2

Several implications can be drawn from this research.
1. **Agent movement.** One implication of the data is that a moving agent didn’t get different results from that of an agent that did not move, despite speculation that agent motion and gesturing could capture student attention and offer feedback without breaking the student’s train of thought (Johnson, Rickel, & Lester, 2000). Similar results were found in previous research (Craig, Gholson, & Driscoll, 2002).

2. **Sub-vocalizing.** First, is it likely that the young participants in the Sound groups in the first study were better able than those in the Text groups to read from the computer because they could articulate the sound of the words, hearing themselves say the words and phrases, sub-vocalizing the sounds to their inner voice, consistent with Stanovitch’s interactive-compensatory hypothesis (Goldsmith-Phillips, 1989; Nickerson, 1991). Young children are not fully capable of mentally articulating instructions and feedback presented in text because their auditory memory consists of a phonological store without a phonological loop. Unarticulated material in young children is therefore analogous to extraneous cognitive load reported in adults (Kalyuga, Chandler, & Sweller, 1999; Mayer, Heiser, & Lonn, 2001; Sweller & Chandler, 1994). This reliance on reading context decreases as a function of reading development and ability (Goldsmith-Phillips, 1989; Swantes, 1991). Figure 2 illustrates the structure and process of working attention of students in the Sound groups, according to the Attentional Control Theory of Multimedia Learning (Mann, 2008).

   From watching the visual events on the screen, the children in the Sound groups would have formed their own associations between the spatial and visual data in visual-spatial memory and willfully re-sketch the graphic or animation through their own visual system like an *inner eye*. From listening to the hints, reminders and cautions spoken by the pedagogical agent, the children in the Sound groups would then have encoded its *gist* or meaning (Brainerd & Reyna, 1990; Hildyard & Olson, 1982; Reyna, 1992; Tannen, 1985) as *acoustic images* (Baddeley, 1986, p. 44) and ending-up in his or her phonological store as *coherent episodes* (Baddeley, 2002) also known as *integrated elements* (Sweller, Merrienboer, & Paas, 1998). Each child would continue to articulate the words in their mind’s ear to create more and more cohesive episodes, the result gradually changing the structure of their long-term memory. On these unfamiliar items and difficult tasks presented by the computer then, each child consciously controlled their own mental processing (Schneider & Shiffrin, 1977) to focus attention and listen, as they relegated any text and the graphical information to a “need to know only” priority. Under these difficult conditions, attention focusing became serial; only one task being processed at a time. Reading and understanding the difficult tasks or unfamiliar items became a priority for these 4th and 5th graders but not the 7th graders.

3. **Treatment effects.** The second implication is that random assignment of the children to treatment groups and their statistically non-significant pre-test scores implies that learning effects were attributable to the treatments and not to prior knowledge. Effective treatments had been developed using to the SSF model as a guide, based on the attentional control theory of multimedia learning, a conceptual framework of working attention during learning that integrates the functions of Brainerd and Reyna’s fuzzy-trace theory within the 4-component structure of Baddeley’s framework of working memory (Mann, 2008).

4. **Read-then-type.** The read-then-type requirement in this study did not disadvantage
speech-cued participants as suggested in a previous study (Segers, Verhoeven, & Hulstein-Hendrikse, 2008) because: 1) The speech-cued participants had their cognitive load shared between auditory and visual memory systems, and; 2) The speech-cued participants were required to process only the gist portion by listening to speech-cues, and the details portion only by reading the text.

5. Read the details, hear the gist. The fourth implication of the study, contrary to the view that the lack of time pressure in the self-paced treatments would not increase cognitive load sufficiently to show a modality effect (Witteman & Segers, 2010), is that the 4th and 5th graders in this study, like their adult counterparts, were able to extract details from reading, and acoustically encode the gist from listening. Clearly listeners to agent speech and readers of agent screen attend differently to information. In general, readers seek the surface features of the text aloud at word-level, as suggested by Halliday (1987). The readers adopted a message focus, in Tannen’s terms (1985), expressing themselves in complex, syntactic constructions and lexicalization for identification. Conclusions arising from the analyses of these data augment previous bimodal research (Brainerd, 1993; Halliday, 1987; Hildyard & Olson, 1982; Mann, 1993; Penney, 1989; Reyna, 1992).

6. Bimodal distribution. Fifth, the modality principle is well suited to the elementary student as it has been for undergraduate psychology students in previous research, but not for seventh grade students—a bimodal distribution.

Limitations

Limitations of the research must also be noted.

1. Experimental method. One limitation was that the experimental method itself restricted our investigations to a tightly controlled, narrowly defined curricular focus. All the teachers of these students were completely excluded from the experiments, quite unlike their daily multimedia learning experiences. The results indicate therefore, that whereas the nature of the presentations used and results from the experiment were valid, they may not be completely generalizable to other groups of school-age students.

2. Temporal speech cueing. In this research, only speech prompts were applied; differences between music, sound effects and speech as informational cues were not addressed. Temporal speech cueing by the agent was designated for empirical study with these students. No other function of sound was investigated. With temporal speech cueing, the instructional intent is that the pedagogical agent uses explanations of the reasoning process in stepwise fashion and context-sensitive program directions to enhance the activity-eliciting the potential of the multimedia learning environment. Temporal speech cueing is but one of six possible functions of sound that can assist students in learning difficult or unfamiliar tasks from a multimedia environment, as suggested in the Structured Sound Function (SSF) model. Instead, a speaking pedagogical agent could for example, be programmed to present one or more points-of-view, or establish a particular locale, atmosphere, feeling or mood. The origin (Mann, 2000, 1995a, 1992) and research application (Mann, 1995b, 1996, 1997a, 1997b; Mann, et al., 2002) of the SSF model is beyond the scope of this paper.

3. Single agent. A single pedagogical agent was used in both experiments. Multiple agents could quite readily be explored for split attention effects, especially with school-
Sustained Learning in 4th and 5th Graders but not 7th Graders

aged and perhaps even pre-school children. Possible dependent or moderating variables could include: reduced clicking backward and forward between web pages; conserved screen space; complementarity with website mathemagenics such as formatted text, graphics and static or moving visuals; aging, and; forgetting and reminiscence.

4. A singular purpose. This research had a particular purpose and therefore did not investigate whether or not these children attended to screen locations, not how they attended to cues, only whether they learned grammar better from convergent temporal sound or text cued by an animated agent character. If we had wanted to look at the psychological processes associated with attending to a cue during a grammar lesson (temporal speech vs. temporal text), we would have had to conduct an analysis of the cognitive episodes in their verbal protocols, as done previously with pre-service teachers (Mann, 1995).

5. Self-paced treatments. A fifth limitation was that the treatments were all self-paced, not system-paced, again a normal situation for the students in school.

6. Motivation. Sixth, motivation was left unplanned because the focus of this study was solely informational. Only simple learner controls on the interface were used instead of more sophisticated methods evident in children’s educational software, such as learner-control with advisement.

7. Sampling. A final limitation was that although they were randomly assigned to treatment groups, the children in both experiments were nevertheless, a population of convenience wherein whole (intact) classes of students were used.

8. The test items. A fourth limitation of the research was the method of testing. In this research, the children’s abilities were assessed with a performance test of multiple choice and short-answer items, in accordance with adult methods of assessing learning from multimedia. A method of testing more consistent with Mayer’s (1997) cognitive theory of multimedia learning would include an expanded inventory of tests of constructed response items or a protocol analysis of the children’s verbalisations to investigate the referential connections that young students make between verbal and spatial instructional materials.

Summary

This research has provided needed data about how school children learn with a web-based pedagogical agent, and how teachers in elementary schools might avail of the characteristics of web-based pedagogical agents in distance learning courses and homework websites. New research is needed on speaking pedagogical agents, and how school age and even pre-school children might learn from them. Spoken directions, instruction and feedback are critical elements of a pedagogical agent’s persona and should continue to be an integral part of the design of the learning environment. It is hoped that eventually, this kind of research will enable researchers and developers to create effective pedagogical agents to reach a wide range of learners from diverse backgrounds.

REFERENCES


Sustained Learning in 4th and 5th Graders but not 7th Graders


