Multi-Modal Interaction in Digital Instructional Media

Seungoh Paek*, Daniel L. Hoffman*, Antonios Saravanos**, Miwha Kim*, John B. Black*
*Teachers College Columbia University, 525 West 120 St., New York, NY 10027
**Oxford University, 62 Banbury Road, Oxford OX2 6PN, UK,
Email: sop2102@tc.columbia.edu

Abstract: The current study examines the impact of various multi-modal combinations on the effectiveness of digital instructional media designed to introduce the concept of multiplication. It compares touch and mouse input methods in conjunction with audio and visual feedback in an effort to improve young children's learning in a virtual environment. One hundred forty-one (N = 141) first and second grade students played *Puzzle Blocks*, a virtual manipulative designed to introduce students to the concept of multiplication through repetitive addition. The results show that having both auditory narration and a touch experience, in addition to visual feedback, is beneficial for young children's math learning.

Introduction

Instructional multimedia can be defined as "presenting words and pictures that are intended to promote learning" (Mayer, 2005, p. 3). The relationship between instructional multimedia and learning has been heavily influenced by research in cognitive science. Research has shown that instructional designers working with instructional multimedia should draw on knowledge about human cognition (Bransford, 2000; Low & Sweller, 2005), by optimizing for working memory and the transfer of information into long-term memory (Mayer & Moreno, 2010). A well-established multimedia design principle, the modality principle, also based on models of working memory, states that people learn better from graphics and narration then from graphics and printed text (Low & Sweller, 2005). This principle has been demonstrated by many studies (Kalyuga, Chandler, & Sweller, 2000; Moreno & Mayer, 2000; Mousavi, Low, & Sweller, 1995). While the modality principle mainly emphasizes information processing through visual and auditory channels, there is a movement in cognitive science to grant the body a more central role in shaping the mind (Wilson, 2002). This suggests the modality principle might need to be expanded to include the kinesthetic/motor channel. What are the implications of including the body in the design of multimedia learning environments?

Theoretical Background

Working Memory and Multimedia Learning

In recent years, a great deal of research has focused on how multimedia learning environments can be used to promote learning. This work has been largely influenced by current theories of working memory, although the relationship between working memory and long-term memory systems is not fully understood and "continues to drive memory research today" (Thorn & Page, 2009, p. 2). Researchers generally agree that working memory is a limited capacity system, which temporarily maintains and stores information, and supports human thought processes by providing an interface between perception, long-term memory and action (Baddeley, 2003). Recent models of working memory postulate different subsystems that process verbal and non-verbal information separately and independently from one another (Baddeley, 1986; 2000). Baddeley's multi-component model of working memory involves three basic processes including a brain network for the maintenance of auditory and verbal information, a separate network for the maintenance of visual and spatial information, and a central executive network for attentional control and the manipulation of items in working memory (Baddeley, 1986).

The Modality Effect

The modality effect is "well-established" (Brunken, Plass, & Leutner, 2004, p. 116) and is based on the notion that working memory has two modality-specific slave systems: one for processing visual and spatial information and one for processing acoustic information (Baddeley, 1992). The effect assumes that the manner in which information is presented will affect how well it is learned and remembered. It has been demonstrated that mixed mode presentations of information are more effective than when the same information is presented in a single mode (Low & Sweller, 2005). Low and Sweller (2005) go on to explain that the "instructional version of the modality effect derives from the split-attention effect, a phenomenon…[that] occurs when multiple sources of information that must be mentally integrated before they can be understood have written (and therefore visual) information presented in spoken (and therefore auditory) form" (p. 147). Brunken and colleagues (2004) write that the underlying assumption is that the "cognitive processing of visually and acoustically presented materials takes place in two separate subsystems of working memory that command separate, independent cognitive resources" (p. 116). Being able to process information using the resources of both subsystems can explain learning outcomes. In other words, when information is presented in two sensory modalities (visual and auditory) rather than one, both slave systems are addressed and total working memory capacity is increased.

Embodied Cognition

While research in multimedia learning has focused largely on how visual and auditory information processing influences learning, research on embodied cognition suggests that the body and its interactions with the world deeply affect cognition. Such research emphasizes sensory and motor systems, suggesting that these systems are essential for successful interaction with the environment (Mahon & Caramazza, 2008; Wilson, 2002). Accordingly, studies in with theories of embodied cognition argue that interaction with the environment through "embodiment" (i.e., physical actions or bodily activity) is an essential part of cognition (Anderson, 2003; Glenberg, 1997; Lakoff, 1987). In recent years, researchers have examined the potential of technology to support embodied activities and ultimately cognition supported by movement.

Purpose of the Study

This study was designed to examine the possible role of the kinesthetic modality on learning. Can the inclusion of a touch-based kinesthetic modality extend the modality effect? More specifically, do kinesthetic interactions in addition to visual and auditory information support young children when it comes to learning a concept such as multiplication in a virtual manipulative learning environment?

Method

Participants and Design

The participants were recruited from public and private schools in a large city located in the northeast of the United States of America. Prior to participation, all participants were required to take a paper-based pre-test to determine their knowledge of addition, a prerequisite skill for multiplication, and their knowledge of multiplication. Students who demonstrated knowledge of multiplication by achieving a score of 50% or higher on the pre-test were dropped from the study. The remaining 141 participants were randomly assigned to one of four groups defined by a 2 x 2 experimental design.

	Audio Narration (A)	No Audio Narration (N)
Finger (F)	N = 39	N = 35
Mouse (M)	N = 34	N = 32

***Table 1*: Experimental group design with treatment.**

Research Instrument

To explore the role of modality on learning in an educational game, a video game called *Puzzle Blocks* was designed based on multimedia learning principles. The goal of *Puzzle Blocks* is to reveal a hidden scene by combining groups of blocks. For example, to create a group of six blocks, players build the group of six by adding two-blocks three times (2+2+2=6). While players move the blocks, they receive visual feedback about the value of the blocks and the sign of the operator. When a group is complete, players are shown the underlying equation including both factors, the equal sign, and the resulting product (e.g. 2x3=6). In addition to visual feedback, *Puzzle Blocks* also provides auditory feedback. When present, the audio feedback is played at the same time as the visual feedback, providing a one-to-one reinforcement of the information shown visually. The audio consists of a male voice-over that counts and reads the appropriate equations aloud. By experiencing the value of the blocks through both the visual and auditory channels, participants might be able to make the connection between the symbolic blocks, their actions, and the underlying mathematical concepts of grouping and multiplication. How they move the blocks, whether directly on a touch-screen (F), or indirectly with a traditional computer mouse (M), may also impact learning outcomes by mediating the other modalities.

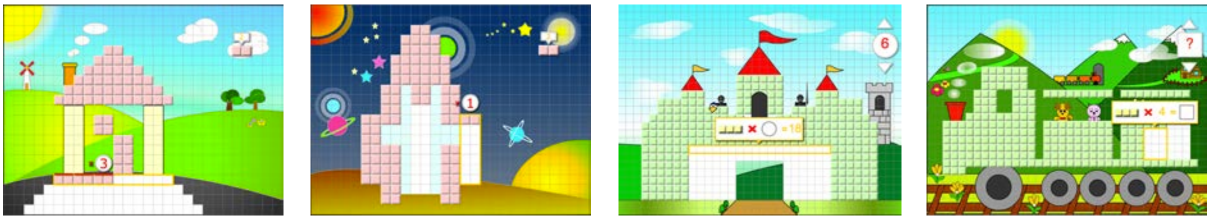

Figure 1: **The *Puzzle Blocks* Interface*.***

Procedure

After the pre-test, all students were assigned to one of four experimental groups. Once the groups were assigned, students played *Puzzle Blocks* for five sessions focusing on the two-times table. In each session students played two or three levels, which lasted, on average, between fifteen to twenty minutes. After the first five sessions of game-play, students took a mid-test. Following the mid-test, students played *Puzzle Blocks* again for five more sessions focusing on the three-times table. After the last play session, students took the post-test.

Materials

Two electronic tests were administered throughout the experiment (mid-test and post-test). The electronic mid-test and post-test consisted of custom software designed by the researchers. Like the virtual manipulative, these digital assessments were deployed on two platforms: a laptop or an iPad. No audio was included. The mid-test electronic assessment presented twelve single-digit whole number multiplication problems that asked for either a missing factor or the product of an equation, such as “2 x ? = 8” or “2 x 6 = ?”. Three problems were from the two-times table and had been “practiced” by participants in the experimental groups. The remaining nine problems were transfer problems in that they had not been seen before in the virtual manipulative environment. All subjects saw the same problems in the same order. The electronic post-test included the twelve questions from the electronic mid-test along with six new transfer questions. No time limit was given and students had three attempts to answer each question.

For each problem presented, a non-interactive group of blocks was shown on the screen as a visual aid (see Figure 2a and 2b).

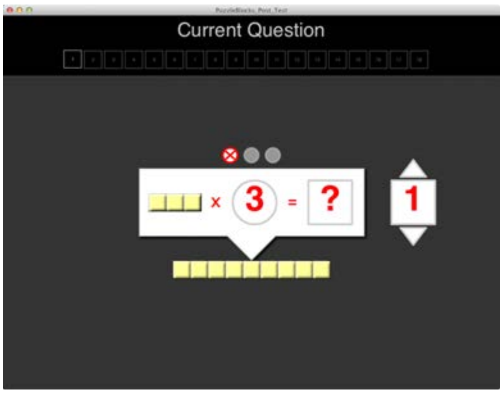

Figure 2a**:** A “studied” three-times table problem presented in the electronic mid- and post-tests. The non-interactive blocks below the problem provide a visual scaffold to help students find the product.

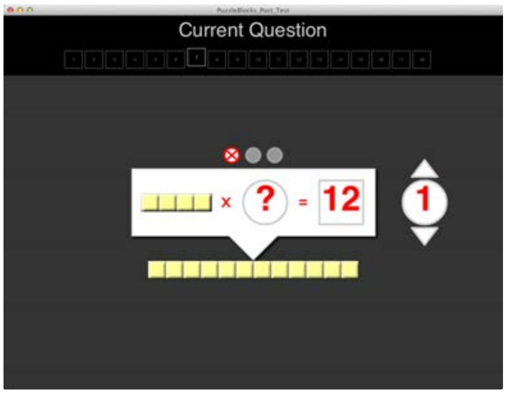

Figure 2b**:** A “non-studied” four-times table problem presented in the mid- and post-tests. The non-interactive blocks below the problem provide a visual scaffold to help students find the missing factor.

Results

Students’ mid-test results show that students in all four groups were able to solve multiplication questions without the game environment.

	N	Mean	Std. Dev
AF	39	10.15	2.25
AM	34	8.97	2.43
NF	35	9.92	1.93
NM	32	8.72	2.53

Table 2: **Mid-test Results by Group**

The results of a Profile Analysis indicate that a significant difference exists on the near-transfer mid-test scores between groups, *F* (3, 136) = 3.838, *p* = .011.

Source	SS	df	MS	F	Sig.
Groups	57.677	3	19.226	3.838	0.011*
Error	681.259	136	5.009		
Total	738.936	139			

* *p* < .05

Table 3: **Mid-test Results by Group**

In addition, the results of pairwise comparisons between groups found significant differences on the near-transfer mid-test between the AT group and the AM group, between the AT group and the NM group, between the NF group and the AM group, and the NF group and the NM group.

Students' post-test results, however, showed that students' post-test mean scores between groups were not significantly different. The differences were marginal, *F* (3, 136) = 2.408, *p* = .070.

	N	Mean	Std. Dev
AF	39	17.44	1.447
AM	34	16.26	2.937
NF	36	16.53	2.762
NM	32	15.88	3.024

Table 4: **Post-test Results by Group**

	SS	df	MS	F	Sig.
Groups	47.914	3	15.971	2.408	0.070
Error	901.879	136	6.631		
Total	949.793	139			

Table 5: **Post-test Results by Group**

The results of pairwise comparisons revealed significant differences only between the AF and NM groups.

Following these results, the impact of two variables, using fingers (F) vs. a computer mouse (M), and having audio narration (A) vs. not having audio narration (N), were examined. The results of two factor comparisons indicated that the presence of audio narration while playing *Puzzle Blocks* was not a significant factor, *F* (1, 136) = .178, *p* = .674, but using a finger versus a computer mouse was a significant factor, *F* (1, 136) = 11.302, *p* = .001, for the near-transfer mid-test score. Similarly, having audio narration was not a significant factor, *F* (1, 136) = 1.881, *p* = .172, for the near-transfer post-test score, but using a finger on a touchscreen was, *F* (1, 136) = 4.863, *p* = .029. There was no interaction effect between the two factors.

Source	SS	df	MS	F	Sig.
Audio	.891	1	.891	.178	.674
Touch	56.615	1	56.615	11.302	.001***
Audio* Touch	.294	1	.294	.059	.809
Error	681.259	136	5.009		

*** p < .001

***Table 6:* Two-Way Analysis of Variance Results for Effects of Audio Narration and Touchscreen on Mid-test Mean Scores**

Source	SS	df	MS	F	Sig.
Audio	12.474	1	12.474	1.881	.172
Touch	32.250	1	32.250	4.863	.029*
Audio* Touch	1.518	1	1.518	.229	.633
Error	901.879	136	6.631		

* $p < .05$

***Table 7:* Two-Way Analysis of Variance Results for Effects of Audio Narration and Touchscreen on Post-test Mean Scores**

Discussion

Recall that the purpose of this study was to examine the impact of various combinations of modalities on learning in a virtual manipulative environment. To explore this topic a virtual manipulative that could be controlled with a mouse or a touchscreen was developed. The results show that all students who used the virtual manipulative gained from the experience. However, the students manipulating the blocks on the touchscreen with their finger showed significantly larger learning gains than students that manipulated the blocks with a computer mouse. In addition, the presence of audio narration seemed to complement, or perhaps supplement, the touchscreen interface, generating even larger learning gains. Ultimately, students manipulating the blocks with their finger and hearing simultaneous narrative feedback showed the largest gains. This finding seems to suggest that multimedia learning researchers should broaden the scope of the modality principle to include not just visual and aurally presented material, but material that can be experienced physically in some way. To the best of the authors' knowledge, this is the first empirical study that shows that combining audio and kinesthetic channels of information results in greater learning outcomes.

References

Baddeley, A. (1992). Working memory: The interface between memory and cognition. *Journal of Cognitive Neuroscience, 4*(3), 281-288.

Baddeley, A. (2003). Working memory: Looking back and looking forward. *Nature Reviews Neuroscience, 4*(10), 829-839.

Baddeley, A. D. (1986). *Working memory*. Oxford: Oxford University Press, Clarendon Press.

Baroody, A. J. (2003). The develiopment of adaptive expertise and flexibility: The integration of conceptual and procedural knowledge. In A. J. Baroody & A. Dowker (Eds.), *The development of arithmetic concepts and skills: constructing adaptive expertise* (pp. 1-34). Mahwah, NJ: Lawrence Erlbaum Associates.

Baroody, A. J., & Ginsburg, H. P. (1986). The relationship between initial meaningful and mechanical knowledge of arithmetic. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 75-112). Hillsdale, NJ: Lawrence Erlbaum Associates.

Bransford, J. D., Brown, A. L., & Cocking, R. R. (Eds.). (2000). *How people learn: Brain, mind, experience, and school* (2nd ed.). Washington, D.C.: National Academies Press.

Brunken, R., Plass, J. L., & Leutner, D. (2004). Assessment of cognitive load in multimedia learning with dual-task methodology: Auditory load and modality effects. *Instructional Science, 32*(1), 115-132.

Hiebert, J., & Lefevre, P. (1986). Conceptual and procedural knowledge in mathematics: An introductory analysis. In J. Hiebert (Ed.), *Conceptual and procedural knowledge: The case of mathematics* (pp. 1-27). Hillsdale, NJ: Lawrence Erlbaum.

Kalyuga, S., Chandler, P., & Sweller, J. (2000). Incorporating learner experience into the design of multimedia instruction. *Journal of Educational Psychology*, *92*(1), 126-136.

Low, R., & Sweller, J. (2005). The modality principle in multimedia learning. In R. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 147-158). Cambridge: Cambridge University Press.
Mahon, B.Z., & Caramazza, A. (2008). A critical look at the Embodied Cognition Hypothesis and a new proposal for grounding conceptual content. *Journal of Physiology-Paris*, *102*(1-3), 59-70.
Mayer, R. E. (2005). Principles for reducing extraneous processing in multimedia learning: coherence, signaling, redundancy, spatial contiguity, and temporal contiguity principles. In R. E. Mayer (Ed.), *The Cambridge Handbook of Multimedia Learning* (pp. 183-212). New York, NY: Cambridge University Press.
Mayer, R. E., & Moreno, R. (2010). Techniques That Reduce Extraneous Cognitive Load and Manage Intrinsic Cognitive Load During Multimedia Learning. In J. Plass, R. Moreno & R. Brunken (Eds.), *Cognitive Load Theory* (pp. 131-152). New York, NY: Cambridge
Moreno, R., & Mayer, R. E. (2000). A coherence effect in multimedia learning: The case for minimizing irrelevant sounds in the design of multimedia instructional messages. *Journal of Educational Psychology, 92*(1), 117-125.
Mousavi, S., Low, R., & Sweller, J. (1995). Reducing cognitive load by mixing auditory and visual presentation modes. *Journal of Educational Psychology, 87*(2), 319–334.
Sweller, J., Ayres, P., & Kalyuga, S. (2011). *Cognitive load theory* (Vol. 1). New York, NY: Springer.
Thorn, A., & Page, M. (2009). *Interactions between short-term and long-term memory in the verbal domain*. New York, NY: Psychology Press.
Wilson, M. (2002). Six views of embodied cognition. *Psychonomic Bulletin and Review*, *9*(4), 625-636.

Acknowledgments

The authors would like to thank the Games for Learning Institute and the Institute for Learning Technologies at Columbia University’s Teachers College for its intellectual and material support.