Static versus Dynamic Presentation of Procedural Instruction: Investigating the Efficacy of Video-based Delivery

Lee Carroll
Eric N. Wiebe
North Carolina State University

In this study, a comparison of static (illustration and text annotation) and dynamic (video with audio annotation) delivery of procedural instructions on a highly spatial task is made. Two groups completed three origami paper-folding tasks, after which they answered a preferences questionnaire and then were asked to re-do the first origami task from memory. The results indicated a general advantage for the dynamic modality in both time for task completion and accuracy. However, complexity of the third origami task and video resolution mitigated the effectiveness of the dynamic modality. No significant difference was seen in instructional modality preference or in memory retention for the task.

INTRODUCTION

Instruction in procedural skills is probably as old as the human race, though how it is delivered has changed markedly with technology. As technology has evolved, this instruction has moved from being largely direct, person-to-person to including text, text and illustrations, and electronic multimedia. As more and more educational institutions are choosing to move away from direct instruction in favor of asynchronous modes (Serdiukov, 2001), technology-assisted means of delivery are being scrutinized more carefully. While creation of text and illustrations—much as has been done with traditional print media—for procedural instructions is an option for distance education delivery, video is increasingly popular as the bandwidth costs of Internet delivery become less significant. Rapid changes in technology has meant that videotaping an instructor performing a task and delivering this via the Internet can be done at a much lower cost and more quickly than creating a comparable sequence of static illustrations and accompanying text.

Asynchronous instruction in spatial tasks has typically demanded a heavy graphics content (Michas & Berry, 2000). This graphics content is usually paired with verbal or textual instructions to help support what is represented in the graphic. What varies in its delivery is the specific mode of representation. These modes can vary along the following dimensions:

- Object surfaces represented as lines (i.e., line drawings) versus shaded (e.g., photographs or computer rendered objects)
- Text versus audio annotation
- Static versus animated (dynamic) images

Ryan and Schwartz (1956) found that simplified line drawings were the most efficient means of conveying information of spatial forms for procedural tasks while Biederman and Ju (1988) found no significant differences for congruent line drawings and shaded still images. Both researchers noted the potential for strong interactions between the mode of representation and the geometry/function of the object being represented. For example, surface gradients revealed by shading can be critical for identifying convex vs. concave curved surfaces or a surface's orientation relative to other surfaces.

For the interplay of graphics and annotation, numerous researchers have hypothesized and confirmed the general advantage of dividing information streams between the visual and auditory channels (Moreno & Mayer, 1999; Paivio, 1983), but it is unclear whether there is a clear cut advantage to having the annotation in audio form rather than as text accompanying the graphic (Gyselinck, Cornoldi, Dubois, De Beni, & Ehrlich, 2002).

A parsimonious conclusion would be that animations for procedural instructions were superior in as much as they show directly what the learner is to perform. However, a review by Tversky, Morrison, and Betrancourt (2002) indicates that past studies have not shown a clear advantage of animations (including video) over static graphics. The reason, they concluded, for animations not showing a clear advantage is explained by two factors: 1) A lack of congruence between the animation and static graphic and/or the procedure to be performed, and 2) The animations are too fast or too complex—they fail to allow the viewer to perceive critical elements of the representation. In addition, Atlas, Cornett, Lane and Napier (1997) raised concerns that the tight correspondence between the animation and the procedure performed by the individual promoted shallow processing of the procedure and limited long term retention.

The present experiment attempts to further explore the potential of video as a substitute for traditional static presentations of procedural instruction. The study uses a highly structured and highly spatial task and presents it in both traditional text and illustration and as a video with audio annotation. Following the recommendations of Tversky et al., care was taken to make the video instruction to be as congruent as possible with the static form of instruction and to break the sequence of instruction into digestible segments.

It is hypothesized that in performing a series of spatial tasks, users of the video (dynamic) instruction will perform the tasks faster and with fewer errors than users of comparable static instruction. They will do so because of the more direct representation of the procedures to perform and the support that shading provides for directly perceiving surface shape and orientation. In addition, the use of the audio track in the video lessens the visual processing load. However, it is also hypothesized that the video users will not demonstrate better retention of their instruction.

METHOD

Twenty-four students were recruited for the study and asked to make three different origami figures using instructions that showed them how to fold the paper into the final form. After completing the task, they were given a short attitude survey concerning the mode of instructional delivery. Finally, the participants were asked to fold the first origami figure from memory.

The participants were split into two equal groups with one group using traditional illustrations and text for instruction, while the other group used a video of the researcher folding the figures (see Appendix 1). Participants in both groups had full control of the media and were allowed to pause or review previous steps. Video was taken of all of the participants folding their origami figures. None of the students considered themselves to have any expertise at origami.

The primary study employed a 2 x 3 mixed model design with two types of instructional format (dynamic and static) and 3 paper folding tasks (ranging from easiest to hardest) as the independent variables. Format was the between subjects variable while the paper folding tasks were completely crossed within subjects. Students were randomly assigned to the two format groups. One group was instructed through text and accompanying line illustrations (static instruction) while the other group was instructed via video (dynamic instruction) (see Appendix 1). Accuracy (error) and time for completion were the response measures for these three tasks.. Errors were counted as misfolds in the paper that had to be redone. In addition, all participants took a standardized test for visualization ability, the Paper Folding Test (Ekstrom, French, Harman, & Dermen, 1976), at the beginning of the session. The score on this test was used as a covariate in the analysis.

The attitude survey taken after the completion of the three folding tasks asked the participants to rate the instructional medium using six, seven-point bipolar scales. These were used to gauge the overall affective response to the instructional medium. The scale pairings were:

- Terrible – Wonderful
- Frustrating – Satisfying
- Dull – Stimulating
- Difficult – Easy
- Inadequate power – Adequate power
- Rigid – Flexible

After completing the survey, the participants were asked to fold the first origami figure using a new piece of paper and without referring to either the instructional materials or their previously folded figure. The number of steps remembered was the response measure for this final task

RESULTS

The third origami figure proved to be so complex that none of the participants in either group were able to complete the figure. A review of the video footage of the participants folding their figure revealed that they all essentially "got stuck" after the fifth step. Therefore, the analysis for this third folding task was done only through the completion of the fifth step of the task. For both response time and accuracy, the covariate (visualization test score) was found to be not significant and was not used in further analyses.

Response time was analyzed using a 2-way ANOVA with Group as a between subjects variable and Task as a within subjects variable (Figure 1). The Task X Group interaction was significant, $F(2, 40) = 4.85$, $p < 0.013$. There was also a significant simple effect for Group ($F(1,20) = 11.35$, $p < 0.003$) and a significant simple effect for Task ($F(2,40) = 34.44$, $p < 0.0001$). Post-hoc contrasts using Least Squares Means showed that the Dynamic group completed Task 1 in a significantly shorter time than the Static group ($p < 0.002$). The Dynamic group also completed Task 2 in a significantly shorter time that the Static group ($p < 0.046$). There was no significant difference in time to complete Task 3 for the Dynamic or the Static group ($p = 0.10$).

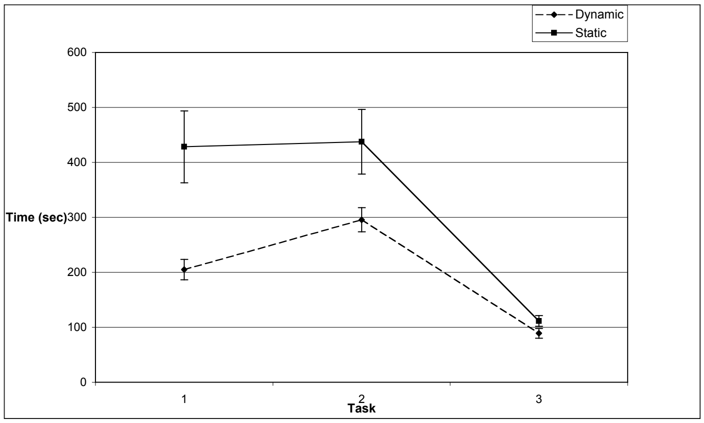


Figure 1. Response time—task versus format group. Error bars represent plus/minus one standard error.

Accuracy was analyzed using a 2-way ANOVA with Group as a between subjects variable and Task as a within subjects variable (Figure 2). There was a significant main effect for Group ($F(1,20) = 20.62$, $p < 0.0002$). There was also a significant main effect for Task ($F(2,40) = 11.95$, $p < 0.0001$). Post-hoc contrasts using Least Squares Means showed that the Dynamic group completed Task 1 with significantly fewer errors than the Static Group ($p < 0.0007$). The Dynamic group also completed Task 2 with significantly fewer errors than the Static group ($p < 0.002$). There was no significant difference in accuracy in Task 3 ($p = 0.31$).

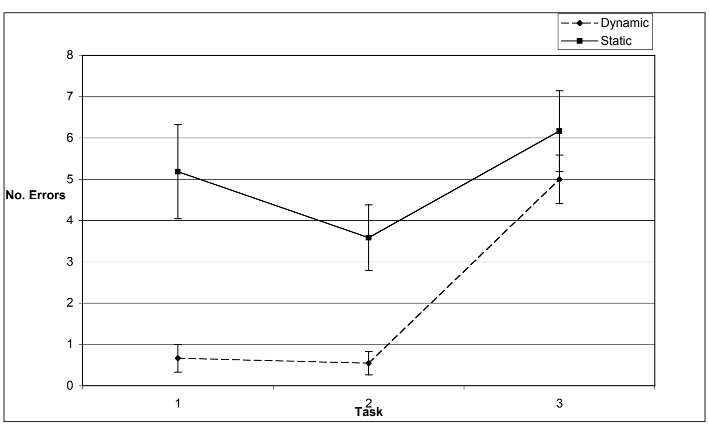


Figure 2. Errors—task versus format group.

The attitude survey was analyzed using a Kruskal-Wallis non-parametric test for each of the six scales. Using Group as the independent variable, no significant differences were found between the Dynamic and Static groups on any of these questions.

Using the number of steps remembered in the refolding task as the dependent measure and Group as the independent measure, results were analyzed using the Kruskal-Wallis non-parametric test. While the Dynamic group was able to remember more steps than the Static group ($M = 4.5$ versus $M = 2.9$), this differences proved to be not significant (ChSq(1,22) = 1.92, $p = .17$).

DISCUSSION

In this study of spatial procedural instruction, video was found to have a performance advantage over static line illustrations and text annotation when the two mediums contained congruent information. The results, therefore, supported the general conclusions of Tversky et al. (2002). By controlling for the congruence between the static and dynamic presentation of procedural instructions and segmenting the presentation into cognitively manageable chunks, the advantage of dynamic presentation was able to be demonstrated. In addition, using a highly spatial task that involved the interpretation of surface shape and orientation meant surface shading shown in the video provided valuable perceptual information.

Closer analysis of the third task also supports Tversky's conclusions. In retrospect, the third task was too advanced for beginning origami. The majority of the participants became "stuck" at a particularly advanced fold that required a high level of visualization ability to imagine the intermediate steps necessary to complete the step. Neither the paper-based nor video instruction provided adequate support for this operation. This lack of support violated Tversky's second principle of complexity and negated the video instruction's advantages. It is interesting to note that while the time to complete the third task was much shorter than the other two tasks (only five steps were analyzed), the error rate was generally higher than the other two tasks. This leads to the conclusion that the task was a particularly difficult one that participants stalled out on very quickly. In such an environment, neither media showed much difference in support for the procedures.

Biederman and Ju's (1988) conclusion that shaded representations could be critical for surface perception was also supported with this study. However, it is interesting to note that the researchers concluded that, at times, the lighting in the video made it difficult to discern individual surfaces on the origami figure, including in the very difficult third task. This would include perception of surface edges; something that line drawings represent very clearly. It would be interesting to find out if the video performance for the third task would have improved if the lighting contrast issues could have been resolved.

Contrary to the conclusions reached by Atlas et al. (1997), the video-based instruction did not seem to degrade long term retention of the procedure. In fact, the video-based group showed a small (but statistically insignificant) advantage over the static group. The attitude survey also did not show any significant difference in response between the two groups. No predictions were made as to which medium would be more favorably received. What is important was that a performance/affective response trade-off was not seen as is sometimes the case with designed visual stimuli. Probably the important lesson is the more universal one: Regardless of the medium, careful design and execution is needed if it is going to be positively received by your audience.

A number of researchers (e.g., Atlas et al., 1997; Najjar, 1998) have proposed instructional multimedia design heuristics based on cognitive and educational psychology research. The work presented here helps fill in some of the gaps; more concretely defining performance advantages to certain approaches and contexts. Still, this work leads to a number of questions that need to be addressed. Among them are more expansive studies on different types of spatial procedures that vary in both the type and complexity of manipulation. The level of quality of form resolution in digital media should also be addressed to see how severely it impacts performance. Finally, previous work on transfer of learning for spatial tasks (Atlas et al., 1997; Sims & Mayer, 2002) points to a need to more fully understand how modality of delivery may affect this transfer.

REFERENCES

Atlas, R., Cornett, L., Lane, D. M., & Napier, H. A. (1997). The use of animation in software training: Pitfalls and benefits. In M. A. Quinones & A. Ehrenstein (Eds.), *Training for a rapidly changing workplace: Applications of psychological research* (pp. 281-302). Hillsdale, NJ: Erlbaum.

Biederman, I., & Ju, G. (1988). Surface versus edge-based determinants of visual recognition. *Cognitive Psychology, 20*, 38-64.

Ekstrom, R. B., French, J. W., Harman, H. H., & Dermen, D. (1976). *Manual for kit of factor-refernced cognitive tests* (Manual). Princeton, NJ: Educational Testing Service.

Gyselinck, V., Cornoldi, C., Dubois, V., De Beni, R., & Ehrlich, M. F. (2002). Visuospatial memory and phonological loop in learning from multimedia. *Applied Cognitive Psychology, 16*(6), 665-685.

Michas, I. C., & Berry, D. C. (2000). Learning a procedural task: Effectiveness of multimedia presentations. *Applied Cognitive Psychology, 14*(6), 555-575.

Moreno, R., & Mayer, R. E. (1999). Cognitive principles of multimedia learning: The role of modality and contiguity. *Journal of Educational Psychology, 91*(2), 358-368.

Najjar, L. J. (1998). Principles of educational multimedia user interface design. *Human Factors, 40*, 311-323.

Paivio, A. (1983). The empirical case for dual coding. In J. C. Yuille (Ed.), *Imagery, memory, and cognition* (pp. 307-332). Hillsdale, NJ: Erlbaum.

Ryan, T., & Schwartz, C. (1956). Speed of perception as a function of mode of representation. *American Journal of Psychology, 69*, 60-69.

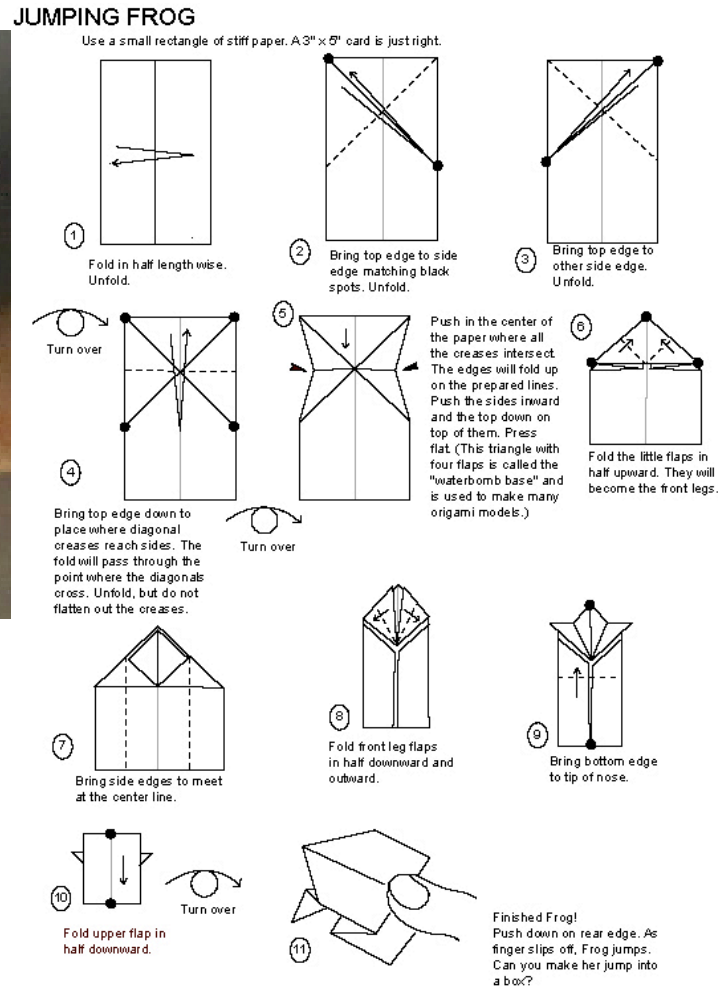
Serdiukov, P. (2001). Models of distance higher education: Fully automated or partially human? *Educational Technology Review, 9*(1), 15-25.

Sims, V. K., & Mayer, R. E. (2002). Domain specificity of spatial expertise: The case of video game players. *Applied Cognitive Psychology, 16*(1), 97-115.

Tversky, B., Morrison, J. B., & Betrancourt, M. (2002). Animation: Can it facilitate? *International Journal of Human-Computer Studies, 57*(4), 247-262.

APPENDIX 1

Example video screen shot for the Dynamic instructional format.



Example instruction sheet for the Static instructional format.