# Learning from Digital Video: An Exploration of How Interactions Affect Outcomes

Robb Lindgren, Roy Pea, Sarah Lewis, and Joe Rosen
Stanford University, 450 Serra Mall, Stanford, CA 94305
Email: {robblind, roypea, sarahl, joro}@stanford.edu

**Abstract:** The sinking costs of producing digital video and its growing presence on the Internet suggest that it has potential for use in web-based learning technologies. However, there have been few investigations into how the kind of interaction one has with video impacts subsequent learning. In this in-progress study participants are asked to watch video of an expert taking apart a toaster and describing how it works. The recorded event is the same for all participants, but the event is presented in one of three different modalities: (1) digital video shot from a free-standing camera (2) digital video shot from a free-standing camera that has been annotated in a video mark-up application called DIVER and (3) digital video shot from a head-mounted camera. A number of different assessment tasks are used to characterize the quantity and type of learning that is supported by a particular mode of video interaction.

## Introduction

Decreasing costs and increasing accessibility of digital video technologies has made it remarkably easy to create a persistent record of a significant event. There is tremendous potential for the use of these video records in learning technologies because they may contain vivid descriptions of expert knowledge or a demonstration of expert practice (e.g., a video of an art historian deconstructing the scenes in Picasso's *Guernica*). In certain visually-based domains a video is likely to have more instructional value than a text-based description of the same event, and yet there is still much to be understood about how we can best promote learning using video-based media. This proposal describes a study in-progress that investigates how the method of capture and the level of interactivity that one has with a video record can affect subsequent learning outcomes.

Digital video can be used to enhance online environments for collaboration by allowing users to participate in a distributed conversation by contributing a video clip rather than simply responding with text. Digital video may also be the focus of a collaborative effort where any number of remote participants each share their unique perspective on the events taking place in the clip—a process that has been referred to as Computer Supported Collaborative Video Analysis (CSCVA) (Pea, Lindgren, & Rosen, 2006). Software environments designed to support this type of video-based collaboration already exist (one in particular will be examined as part of this study). There is an assumption with all these environments that people are capable of learning from the video products of this collaboration; however there has been hardly any empirical research that examines how the mode of video presentation impacts one's ability to learn from that video.

## Video, Learning, and Interactivity

Previous research on video and learning has focused on people's ability to learn from relatively stable video formats, such as television, where the interaction with the video is generally at the level of passive transmission (for a review see Seels, Fullerton, Berry & Horn, 1996). Schwartz and Hartman (in press) discuss how to design digital video such that it promotes specific learning outcomes (e.g., "Doing" outcomes where one learns from a model of some human behavior). These accounts have a high utility for the design of learning technologies, however their focus has been primarily on the *content* of the video and not the ways in which the learner might interact with the content once it has been designed and recorded.

An intriguing finding from Smeeton, Williams, Hodges, and Ward (2005) suggests that the kinds of interactions that people have with video can affect their future performance. Participants in this study were intermediate-level tennis players who were shown video clips of players making different types of shots. Some of the participants were given explicit instruction from an expert on how to anticipate a shot based on the postural cues of their opponent. Another group was shown the same video clips, but was told simply to look for relationships between a player's posture and the subsequent shot. On a post-test using a tennis simulator, participants in the "explicit" group showed better reaction times than participants in the "discovery" group. However, when the conditions of the post-test were altered so that they provoked more game-like anxiety (e.g., participants were told that they would be ranked and that their performance would be reported to their coach) the participants in the explicit condition suffered a decrement in their performance while the players in the discovery group did not. This study suggests that different interaction conditions may be better suited for generating more *adaptive* performance. It may be that different aspects of digital video interaction put learners on a course for acquiring the more flexible and productive kinds of knowledge referred to as adaptive expertise (Hatano & Inagaki, 1996). One of the aims of this study is to begin exploring how a technology-based video environment with a particular set of interaction affordances might support the development of an adaptive expert.

## Study Design

The design of this study was intended to contrast three different modes of interacting with a video-recorded event while carefully controlling for the content presented in each of the interactions. While by no means comprehensive, we chose our three video conditions because we felt that they were representative of a class of video interactions that may have unique consequences for learning. The three viewing conditions are: digital video shot from a free-standing camera, digital video shot from a free-standing camera that has been annotated in a video mark-up application called DIVER, and digital video shot from a head-mounted camera. The subject matter event that we chose for this study is a video recording of an expert disassembling a toaster and describing how it works. Toasters are surprisingly complex instruments that employ several key concepts from physics and electricity. Very few people understand the functionality of a toaster and describing how one works is a highly visual task that is well suited for the video-based media used in this study.

### DIVER: A Software Environment for Video Analysis and Collaboration

The software application mentioned above that was used to create the annotated video in the second condition of this study is called DIVER (Digital Interactive Video Exploration and Reflection) (see Pea et al., 2004) which was developed at the Stanford Center for Innovations in Learning. DIVER is a video analysis application that allows users to associate text annotations with specific points in time and space within a video record; it is an authoring process that we call making a "dive." By controlling a viewing rectangle that is overlaid on the video, the user can direct the attention of others who view the dive to see what the author sees.

### Head Cam

In the last few years some researchers in the social sciences have begun equipping their participants with head-mounted cameras as a way to collect data about aspects of human behavior such as communication and collaboration (for an example, see Fussell, Setlock, & Kraut, 2003). We feel that viewing video recorded by a head cam is a unique opportunity to take on the perspective of another individual, and we were interested in how easily the adoption of an expert's perspective would translate to learning. For this study we fitted the toaster expert with a Sony Super HAD high-resolution camera attached to a headband.

### Study Procedures

When completed, a total of 30 participants will be run in this study with 10 participants in each of the three conditions. Each session begins with the participant completing a short survey that is used to assess the participant's prior knowledge of toasters and other related domains (e.g., physical mechanics). Next the participant is told that they will have the opportunity to view some video and that their task is to use the materials to learn as much as possible about how a toaster works in 15 minutes. Participants receive the video in one of three formats (see Figures 1(a-c) for a screen shot of each condition):

*1. Video from a free-standing camera + passage (Figure 1a)*: In this condition, participants are presented with the toaster video in a stand-alone digital video player on a laptop computer. The participant is also given a piece of paper with a passage that describes the functionality of a toaster. Some of the information in the passage is redundant information presented in the video, and some of the information is novel. Prior to starting, the participant is given brief instructions on how to use the video player.
*2. Video presented in a dive (Figure 1b)*: In this condition, participants are presented with a dive that was created in DIVER using the same video in the first condition. The dive is presented in a compact format that allows users to jump from section to section by clicking on a thumbnail to the right of the video. The text used to annotate the video is the same text used in the passage in the first condition. Prior to starting, the participant is given brief instructions on how to use DIVER.

*3. Video from a head-mounted camera + passage (Figure 1c)*: This condition is identical to the first condition except that the video used is recorded from a head camera. The head cam video was recorded at the same time as the free-standing camera so that the audio and the events captured in all three conditions are the same.

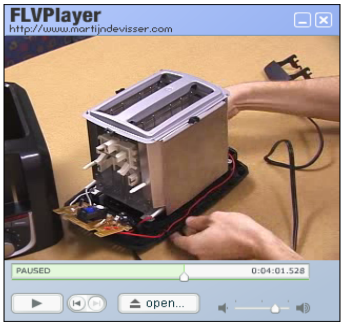


Figure 1a. Toaster video from a free-standing camera.


Figure 1b. Toaster video dive.


Figure 1c. Toaster video from a head-mounted camera.

After the participants in each of the conditions have reviewed their materials, a series of assessments are administered. Each of these assessments target a different form of knowledge or understanding in order to determine if different types of video interactions support different kinds of learning. Participants are asked to complete four tasks: (1) Answer a series of paper-based conceptual and vocabulary questions (e.g. "Why are the filaments in a toaster wound more tightly at the bottom than at the top?") (2) Describe how a toaster works to someone who doesn't know (3) Describe how they would troubleshoot a series of hypothetical toaster malfunctions (e.g. "The toaster is plugged in and the tray stays down, but my bread isn't heating up.") and (4) Draw a functional diagram of a toaster.

## Data Analysis

In addition to the paper-based assessments, we are videotaping the verbal assessments so that they can be coded and scored based on their level of understanding. Performance on each of the four assessments will be quantified so that comparisons can be made across assessments and across video conditions. We are also videotaping the learning phase of the study where participants interact with video and paper materials. We are working to create rich characterizations of learning behavior in each of the three conditions such that they can be related to the participants' outcome measures. Our hope is that a combination of quantitative and qualitative analyses will result in a clearer picture of how different types of digital video interactions contribute to the acquisition of expert knowledge.

## References

Fussell, S. R., Setlock, L. D. & Kraut, R. E. (2003). Effects of head-mounted and scene-oriented video systems on remote collaboration on physical tasks. In: *Proceedings of the CHI 2003 Conference on Human Factors in Computing Systems.* ACM Press.

Hatano, G., & Inagaki, K. (1986). Two courses of expertise. In H. Stevenson, H. Azuma, & K. Hakuta (Eds.), *Child development and education in Japan* (pp. 262–272). New York: Freeman.

Pea, R., Lindgren, R., & Rosen, J. (2006). Computer-supported collaborative video analysis. *Proceedings of the Seventh International Conference of the Learning Sciences (ICLS)*. Bloomington, Indiana. Mahwah, NJ: Erlbaum.

Pea, R., Mills, M., Rosen, J., Dauber, K., Effelsberg, W., & Hoffert, E. (2004). The DIVER™ project: Interactive digital video repurposing. *IEEE Multimedia, 11*(1), 54-61.

Schwartz, D. L., & Hartman, K. (in press). It is not television anymore: Designing digital video for learning and assessment. In R. Goldman, R. P Pea, B. Barron, & S. Derry (Eds.), *Video research in the learning sciences.*

Seels, B., Fullerton, K., Berry, L., & Horn, L. J. (2004). Research on learning from television. In D. H. Jonassen (Ed.), Handbook of research on educational communications and technology, 2nd edition. Mahwah, NJ: Erlbaum.

Smeeton, M. J., Williams, A. M., Hodges, N. J., Ward, P. (2005). The relative effectiveness of various instructional approaches in developing anticipation skill. *Journal of Experimental Psychology: Applied, 11*(2), 98-110.