Designing Contrasting Video Case Activities to Facilitate Learning of Complex Subject Matter

Brian D. Beitzel and Sharon J. Derry

University of Wisconsin-Madison

Paper presented at the annual meeting of the
American Educational Research Association

San Diego, CA
April 2004

Please address correspondence regarding this paper to Brian Beitzel (brian@edpsych.net).

This material is based upon work supported by the National Science Foundation under Grant No. 107067 (ROLE).

Instructional use of video cases has become quite popular in recent years, particularly in teacher education (see, for example, Abell, 1996; Bransford et al., 1986; Copeland & Decker, 1996; Derry, Siegel, Stampen, & STEP Team, 2002; Goldman & Barron, 1990; Hannah, 1995; Lampert & Ball, 1998; Liedtka, 2001; Merseth & Lacey, 1993; Risko, Yount, & Towell, 1991; Stephens, Leavell, Fabris, Buford, & Hill, 1999; Stigler & Hiebert, 1999). However, little empirical evidence demonstrating the benefits of instruction with video over more traditional methods of instruction has accompanied this proliferation of video usage. Given the recent calls for evidenced-based practice and credibility in educational research (e.g., Levin, 2003; Levin & O'Donnell, 1999; O'Donnell & Levin, 2001), this lack of empirical justification for such a widely adopted instructional medium is somewhat troubling. In addition to the lack of empirical support, no theoretical framework has been established to guide the instructional design of activities that use video cases. The experiment reported in this paper was designed to address both of these issues.

Although the kinds of instructional activities in which video cases may potentially be used are unlimited, we have chosen a specific kind of activity, called a "contrasting-cases" activity (Schwartz & Bransford, 1998), with which to investigate the use of video cases in education. First we describe Schwartz and Bransford's implementation of contrasting cases and their associated theoretical explanation of its benefits; then we draw upon some other theoretical perspectives to inform the predictions for our study with video cases. Finally, we present our experimental design, along with a portion of our data, and suggest some future directions for video research.

The Contrasting-Cases Approach

Schwartz and Bransford (1998) showed that, when students have limited prior knowledge or experience related to a complex to-be-taught concept, having them examine the similarities and differences among cases representing various real-world instantiations of that concept can help them generate prior knowledge that prepares them to derive greater benefit from a subsequent lecture or reading on that topic. Schwartz and Bransford used the term "contrasting cases" to refer to this instructional method.

To illustrate the notion of contrasting cases, consider the concept *knife*. Many types of knives exist; some (e.g., a bread knife) are very specialized to perform specific tasks, while others (e.g., a pocket knife) are intended for general purpose use. All knives have some type of blade and handle; the shapes and sizes of the blades and handles, however, vary a great deal, as would be apparent if we were to compare various examples of knives side by side. However, most real-world concepts are more complicated than might be apparent using the knife example. Now consider the concept *understanding*. Students in teacher preparation programs are expected to state instructional goals and design assessments and activities to help learners acquire *enduring understandings* in various subject disciplines. But what is *understanding*? How does one learn or teach about what it means to *understand*? Abstract concepts such as *understanding*, *remembering*, *beauty*, *truth*, *justice*, etc., are particularly difficult to teach because they cannot be defined or explained in a way that will cover all, or even most, cases that students will encounter.

Schwartz and Bransford (1998) conducted three experiments, in the context of their own classroom instruction, to investigate the use of contrasting cases in teaching complex cognitive concepts related to memory and remembering. For each concept to be taught, they constructed a set of cases; each set comprised several cases consisting of simplified presentations (a few lines

of text) of recall data based on some classic cognitive psychology experiments. For each set of cases, they also prepared an integrated summary of the cases in that set, representing what people had recalled during the original experiments from which the cases were drawn.

The basis of Schwartz and Bransford's methodology is as follows: Students were given either a set of contrasting cases to "analyze" or a summary of the cases to study. Analyses of the cases consisted of carefully examining the cases and noticing similarities and differences among them; students recorded their observations in a graph depicting the frequencies of the observed features in the cases. After studying either the cases or the summary as a homework assignment, students attended a lecture that provided in-depth coverage of theoretical explanations accounting for the patterns observable in the cases (and inducible from the summary). Schwartz and Bransford hypothesized that when students compared the cases, they would notice patterns for which they had no theoretical explanation; they would then be better prepared to learn from a lecture explaining the significance of the distinctions they had observed in the cases. It was expected that the contrasting-case analysis activity would foster the creation of a sufficiently differentiated prior knowledge base so that students who had analyzed the cases would benefit more from the lecture than students who had read the summary of the cases. About a week after the lecture, students were asked to make predictions regarding the outcome of a hypothetical experiment that drew on the concepts from the cases. This prediction task served as a measure of transfer—students who had contrasted cases and thus constructed their own "well differentiated" knowledge of the concepts prior to being informed of the theoretical significance of those distinctions were expected to make more accurate predictions than those who understood the concepts superficially.

Schwartz and Bransford's results supported their claim that when students analyze contrasting cases surrounding a previously unfamiliar theme and subsequently attend a lecture that provides a theoretical explanation for the distinctions present in the cases, their understanding of the concepts represented by the cases aligns more closely with expert knowledge than the understanding of students who had read a summary of the cases and then attended the same lecture. Similar results were obtained when students contrasted cases and then read a text (similar to a chapter in a typical undergraduate textbook) rather than attending a lecture. Alternative explanations for these findings, such as an active-learner hypothesis (since the students who read the summary of the cases were not engaged in an "active" learning activity), were ruled out by carefully designed follow-up studies. In addition, the contrasting-case activity was shown to be insufficient in itself to create advanced subject-matter knowledge—additional information, such as that provided by a lecture or text, was required in order for students to gain in-depth knowledge of the topic. According to Schwartz and Bransford, contrasting-case analysis can be a method of activating differentiated cognitive structures based on prior knowledge in order that more advanced instruction may build upon this knowledge base; it is not a stand-alone activity. They referred to this explanation of their results as the "knowledge-differentiation hypothesis."

Contrasting Video Cases

Video cases allow the transmission of much more verbal information than can be conveniently recorded in a text case. Probably the greatest advantage of video cases, however, is their ability to portray *nonverbal* information, such as people's gestures, tone of voice, and physical movements around the environment. This nonverbal information would be impractical,

if not impossible, to include in a text case. Such information could be crucial to one's understanding of the case if no equivalent verbal cues exist. Additionally, when video cases are used for training purposes, such as teacher education, they may not only convey more information than text cases, but they may also be a more accurate representation of reality in the field of practice; indeed, J. J. Gibson (1979) argued that learning from film activates perceptual processes similar to those that are active in comparable real-world situations. Presumably, the use of video cases in training programs provides experiences similar to those encountered in professional practice, thus enhancing the probability of transfer to one's future practice in the field.

Schwartz and Bransford's (1998) knowledge-differentiation hypothesis was formulated and examined in an instructional context that used text cases. Their cases were brief (only a few lines of text) and clear; the major distinctions among them were not difficult to discern, although the *reasons* for those distinctions were less obvious. Therefore, any differentiated knowledge that students constructed from comparing the cases was rather straightforward and "correct" with respect to the following text or lecture; that is, there was little room for interpretation because of the clear-cut nature of the cases. Consider once again the complex concept *understanding*. Suppose that several video cases representing the concept *understanding* could be constructed and presented to students for them to compare and contrast. Would the differentiated knowledge that they construct align as closely with a forthcoming text or lecture about the concept *understanding* as the students' knowledge in Schwartz and Bransford's study did? It is possible that when students analyze cases representing real-world complex concepts, they may develop differentiated knowledge structures that are very different from—or even in conflict with—the ideas they will later encounter in reading about those concepts; in such a situation, learning may be hindered rather than aided (cf. Hewson, Beeth, & Thorley, 1998; Posner, Strike, Hewson, & Gertzog, 1982). Consequently, extending Schwartz and Bransford's contrasting-cases methodology to video cases may not result in an outcome consistent with their predictions. We next consider some alternative theories that may be helpful in guiding predictions in the context of instructional designs that employ a contrasting-cases approach with *video* cases.

Other Theoretical Perspectives on Contrasting Cases

Research involving examination of the similarities and differences among a set of examples is not a new endeavor. For example, perceptual theorists have studied for decades how people discriminate among stimuli with differing appearances (e.g., Garner, 1962, 1974; E. J. Gibson, 1969; J. J. Gibson & Gibson, 1955; Royer, 1966). Past and contemporary research on learning by analogy (e.g., Gentner, Loewenstein, & Thompson, 2003; Gick & Holyoak, 1980, 1983; Kurtz, Miao, & Gentner, 2001; Thompson, Gentner, & Loewenstein, 2000) also indicates that comparing multiple examples of a concept may contribute to a more advanced understanding of the concept. Schwartz and Bransford's (1998) unique contribution to this line of work is the notion that engaging in an activity designed to elucidate the distinctions among several examples of a concept can help to activate some relevant prior knowledge—and might even perhaps *create* some prior knowledge—that, if sufficiently "differentiated," provides a rich knowledge base that subsequent instruction can build upon.

Schema Elaboration

In professional training, such as teacher education, might students be better off if they are introduced to a concept by first reading a relevant text, and then directed to examine real-world concept variations by analyzing contrasting video cases? Students who read first to activate an anchoring knowledge structure and then analyze video cases may use the cases to elaborate schemas that were activated in the reading. This is essentially the argument advanced by Derry (in press), which we call here the schema-elaboration hypothesis. The basic idea is that, when video cases are employed, contrasting-cases treatments would best be presented after some introductory reading in which fundamental ideas about the complex concepts are learned. After some learning takes place at a generalized level, various contrasting cases can be presented that criss-cross (Spiro, Vispoel, Schmitz, Samarapungavan, & Boerger, 1987) the landscape of practice (e.g., classroom instruction), helping students elaborate their original schemas about the concept.

The schema-elaboration view suggests that mere differentiation is not the only important factor in learning from contrasting cases; more stringently, the differentiation must be analogically compatible with the reading. Thus, the schema-elaboration hypothesis seems more in concert with the analogical reasoning research of Gentner and her colleagues (e.g., Gentner, 1983; Gentner et al., 2003; Kurtz et al., 2001) than with Schwartz and Bransford's (1998) differentiation hypothesis. If contrasting-cases effects are merely due to more differentiated structures (i.e., regardless of the pattern of that differentiation), contrasting-cases treatments should work even if the students develop their own *unique* differentiated knowledge structures. In studying cases that contain more complexity, students are unlikely to generate contrasts that match the forthcoming reading or lecture. Thus, even though students who contrast cases may develop more differentiated knowledge structures, those structures could be very different from, even in conflict with, ideas in a subsequent reading or lecture and consequently may create confusion rather than support, especially in early learning.

Reconstructive Processing

Another perspective on cognitive processing during video case learning is that it may resemble a reconstructive memory process in the sense of Bartlett (1932) and Spiro (1977). Briefly, these theorists suggested that the process of remembering text ideas is a reconstructive one that is heavily influenced by experiences that occur subsequent to reading. Consequently, the reconstructive hypothesis predicts that when students recall a text that they have read, any relevant knowledge they had prior to reading the text, as well as other related material encountered *after* reading the text, will be integrated with their text memories. Because of the diverse contexts provided by these various knowledge sources, this integration is expected to enhance the likelihood of transfer. It might not, however, support accurate text recall if, after reading, other sources are encountered that reinterpret the text ideas from the point of view represented by that later context.

Therefore, according to the reconstructive hypothesis, after students have read a text and are engaging in a contrasting-cases activity, they are not elaborately updating previously acquired text memories (as the schema-elaboration hypothesis suggests) so much as they are actively re-processing and possibly even *distorting* text memories from the personal point of view activated by the case analysis. Hence, when video cases are analyzed after reading, text

ideas are actively "imported" into the case-comparison activity along with other prior knowledge. As a result, although some text details may be lost to future recall due to the integration of prior knowledge and ideas from the contrasting-cases activity, stronger transfer capability should result. However, when a text is read *after* a contrasting-cases activity, the resulting knowledge structures will contain relatively few importations from prior knowledge, will not encourage transfer, but will support text recall.

Indexical Hypothesis

Glenberg and Robertson's (1999) indexical hypothesis proposes that perceptual background knowledge contributes to a reader's understanding of a text. For example, a person can comprehend the instructions in an owner's manual for a newly acquired device much more easily when he or she has first seen the device. Glenberg and Robertson proposed that a two-step process is responsible for this enhanced comprehension following a relevant perceptual experience. First, words or phrases in the instruction manual are "indexed" (or mapped, in one-to-one correspondence) to an object (or to a mental representation of an object). Second, affordances (Gibson, 1979) are derived from the object that guide the comprehension of the text. Thus, perceptual experiences facilitate indexing, which makes affordances available, which in turn increases text comprehension. Glenberg and Robertson stressed that assessments of comprehension should include authentic, real-world tasks that are better measures of transfer than paper-and-pencil instruments.

Thus, by extension, the indexical hypothesis seems to suggest that when students contrast video cases prior to reading a related text, both their text comprehension and their transfer of the ideas in the text should be enhanced, relative to students who have not contrasted video cases before reading the text. Accordingly, if students read a text without having first contrasted video cases, there will be no perceptual background knowledge to draw upon as the text is read, and lower text comprehension and poorer transfer of the text ideas to real-world situations should result.

In sum, the knowledge-differentiation hypothesis supported by Schwartz and Bransford (1998) claims that a contrasting-cases exercise prepares students to read a text (or hear a lecture) by creating and activating relevant differentiated knowledge structures. However, when using video cases to teach complex concepts, and particularly in situations where the reading itself is not extremely difficult to comprehend, the Schwartz and Bransford knowledge-differentiation hypothesis may not apply. The schema-elaboration hypothesis predicts that a more useful understanding will be developed by first reading or hearing introductory material to develop an initial schema and then contrasting realistic video cases to elaborate on that schema (Derry, in press; Spiro et al., 1987). The reconstructive memory view (e.g., Bartlett, 1932; Spiro, 1977) suggests that text recall will be better if the text is read *after* contrasting some video cases, but that transfer of text ideas will be enhanced if the text is read first, followed by a contrasting-cases activity. Finally, Glenberg and Robertson's (1999) indexical hypothesis favors an instructional sequence in which students first engage in a contrasting-cases activity that provides some perceptual background knowledge; armed with this perceptual experience, students will then be better able to comprehend a related reading. These hypotheses were examined in a laboratory-based experiment, described next.

Method

Participants

The participants in this experiment were 153 (21 male, 132 female) students recruited from undergraduate educational psychology courses at the University of Wisconsin-Madison in exchange for extra course credit.

Materials

Two sets of materials were used in an effort to provide internal replication of any findings. The first set of materials was related to the concept *understanding*; the second set of materials was related to the concept *assessment*. These concepts were selected because they illustrate nicely the varied appearance that complex concepts can assume; additionally, video cases are frequently used to teach these particular concepts in teacher professional development courses.

Video Cases

The video cases were drawn from various sources, including classroom-based footage and appropriate segments of Hollywood movies. Each case was approximately 1 min or less in length. Twelve cases illustrated different aspects of the concept *understanding*, and 12 cases represented *assessment*. For example, one of the *understanding* clips depicted a high school science class discussion of static electricity, illustrating the ability to "explain" one's knowledge—a form of understanding discussed in the related reading. One of the *assessment* cases depicted a physical education teacher evaluating the gymnastic skills of her students; this clip illustrated "performance assessment," one of the types of assessment described in the related reading.

Readings

Two text passages were prepared to provide depth of understanding of the concepts illustrated in the video cases. The difficulty of these readings represented that of a typical college textbook. The reading related to *understanding* described Wiggins and McTighe's (1998) six "facets" of understanding and gave an example of each. The reading related to *assessment* described various purposes and types of classroom assessment. The *understanding* reading was 1,127 words in length; the *assessment* reading contained 2,145 words.

Online Environment

All activities were developed within and guided by STELLAR, a Web-based authoring system that allows developers and researchers to design step-by-step online learning activities (Derry, in press). As an example of one of the steps in the experiment, when participants reached the step containing the contrasting-cases activity, they read some brief instructions and were asked to click on a hyperlink that opened a new browser window in an adjacent computer monitor; the new window displayed hyperlinks to the video cases that participants were then

asked to view and sort into six groups containing two or three cases apiece. Each group was supposed to represent a different aspect of the concept in focus (e.g., *understanding*). Participants were directed to enter their case groupings into a standard Web form that was displayed below the instructions for that step; they were also asked to label each group and briefly explain how the cases in each group fit the label.

Dependent Measures

Assessments were administered, within the STELLAR environment, to evaluate participants' memory for, understandings of, and ability to transfer and apply, the two conceptual systems studied: *understanding* and *assessment*. Three dependent measures were administered for each of these two conceptual systems; however, only two of these dependent measures are reported in this paper, as scoring is still in progress.

Recall tasks. These were basic memory tasks in which participants were asked to list the main ideas of the related reading (i.e., either *understanding* or *assessment*).

Application tasks. The purpose of these tasks was to measure participants' transfer of knowledge from the study materials to a new situation. In the *understanding* version of this task, participants were asked to list and describe all the learning goals that they would like second-graders to achieve by the end of a three-week unit on rainforests. In the *assessment* version, participants were asked to write an assessment plan describing all the types of assessment that they would use in a three-week rainforest unit for second-graders.

Design

Participants were randomly assigned in approximately equal numbers to three experimental conditions in a between-subjects design with repeated measures on two topics: *understanding* and *assessment*. The experimental conditions, described more fully below, varied in both the study materials provided and the sequence in which they were presented. Briefly, participants in the "cases-text" condition first contrasted some video cases and then read a related reading; participants in the "text-cases" condition read the reading and then contrasted video cases; participants in the "text-notes" condition read the text and then took notes on that text. The video cases and readings used in all conditions were identical.

All participants experienced both the *understanding* and the *assessment* materials; the order of presentation of the two sets of materials was counterbalanced within each condition.

Procedure

This experiment was conducted within a laboratory setting. As mentioned earlier, all participants were guided through the experiment in the STELLAR learning environment. Participants completed this experiment in two sessions, approximately two (M = 1.98, SD = .41) days apart.

Session One

In the first session, participants studied the materials for both *understanding* and *assessment*. The order of the materials was counterbalanced within each condition—i.e., half of

the participants studied the *understanding* materials first, and the remaining participants studied the *assessment* materials first.

In the video-based conditions, the sequence of the cases and reading was also manipulated. Participants in the cases-text condition first viewed and analyzed the contrasting cases; their analyses were entered into a standard Web-based form and consisted of grouping the cases into six groups of two to three cases each and labeling the groups. After participants submitted the form containing their analyses, they were presented with the reading. Participants in the text-cases condition encountered the same materials as the participants in the cases-text conditions, except that they studied the reading prior to analyzing the contrasting cases. Although participants were unrestricted in the amount of time they were allowed to study the materials, time data were recorded.

In the text-notes condition, participants first read the text; then, the same text was redisplayed and participants were asked to take notes on it.

A demographic questionnaire was administered after the first set of materials had been studied.

Following the questionnaire, the second set of materials was presented. The sequence of cases and reading (or reading and note-taking) was held constant for each participant—i.e., if a participant contrasted the *understanding* cases before reading the *understanding* text, s/he contrasted the *assessment* cases before reading the *assessment* text.

Session Two

In the second session, approximately two days later, several dependent measures were administered. The dependent measures were grouped by conceptual system (*understanding* and *assessment*) and presented in the order in which each participant had studied the conceptual systems; for example, if a participant had studied the *understanding* materials first, s/he was first presented with the dependent measures related to *understanding*, followed by the dependent measures related to *assessment*. Again, time on task was not limited, although time data were recorded.

Within each set of dependent measures, the recall task was presented first, followed by the application task (and then a third task that is not discussed in this paper).

At the end of the experiment, the general feedback questionnaire was administered.

Results

As described earlier, there were two overarching goals for this study: 1) to provide some empirical data comparing video-based and non-video-based learning activities; and 2) to explore various theories that may be fruitful in explaining the effects of learning activities involving video cases and providing some guidance for designing learning activities that involve video cases. Our *a priori* predictions, based on some pilot data, suggested that the reconstructive hypothesis may possibly be the theoretical perspective that best predicts the results of a contrasting video case activity.

Our pairwise comparisons, described in detail below, allowed us to examine the predictions associated with the theories described in the introduction of this paper; however, instead of merely accumulating evidence consistent with any of these theories, we decided to subject them to a more rigorous test by placing our theory of interest (e.g., the reconstructive

hypothesis) in the null hypothesis and trying to show that there is evidence *against* this theory by trying to reject the null hypothesis. Since no theory can ever be proven to be true (i.e., there may always be some undiscovered counterevidence right around the corner), we chose to severely test our theory of interest by placing it in the null hypothesis. This configuration constrains the conclusions we can draw from our results; namely, if we reject the null hypothesis, which contains our theory of interest, we can say that we have evidence against our theory; conversely, if we fail to reject the null hypothesis, we must be content to say that at least the theory was not wrong—but in this case of non-rejection, we must also be careful not to conclude that the theory was correct. This logic underlies all of the pairwise comparisons that will be presented in the remainder of this paper.

It is also important to note that our study involved text passages that were straightforward and fairly easy to understand without the aid of additional materials; this type of reading is unlike that of Glenberg and Robertson (1999), for example, where a perceptual device was needed to help understand the associated reading. We believe that in our situation, the video is an "enhancement" of the text, providing supplementary information and examples, but is not required for understanding the text. As with any scientific research, the conclusions or generalizations that we draw from our study must be limited to contexts similar to our own, rather than making broad claims regarding *any* type of instruction using video cases.

A subset of the data from each dependent measure was scored by two independent raters; if the criterion inter-rater reliability of .80 or greater was not met, the process was continued iteratively with additional subsets of the data; after the inter-rater reliability criterion was reached, the remainder of each dependent measure was scored by a single rater. The summary data for each of the conditions, collapsed across the two sets of materials, are shown in Table 1.

All statistical tests were conducted with a familywise Type I error probability of .05. The total number of minutes that participants spent studying the text and video cases during the first experimental session was used as a covariate in all analyses in order to statistically control for the differing amounts of time that participants spent studying the materials.

Recall Tasks

Participants' responses to the recall tasks for the *understanding* and *assessment* materials were scored for both term recall and gist recall. For term recall, points were awarded according to the following scale for each concept discussed in the related reading: 0 = "no mention"; 1 = "paraphrase of term"; 2 = "exact recall of term." For gist recall, points were given according to the following scale for each concept from the related reading: 0 = "no recall"; 1 = "minimal gist recall"; 2 = "partial gist recall"; 3 = "excellent gist recall."

Since there were theoretical predictions associated with all three possible pairwise comparisons, a Fisher LSD approach (Levin, Serlin, & Seaman, 1994) was used for both term recall and gist recall.

To evaluate whether participants' performance differed between the two sets of materials, an ANCOVA model was computed using percentage data (since the total number of points possible for each set of materials was different, due to the fact that there were more assessment concepts in the *assessment* reading than facets of understanding in the *understanding* reading). This analysis also allowed us to check for order effects. We found no differences in term recall between the *understanding* and *assessment* materials, $F(1, 146) = 0.14$, $p = .14$; therefore, the data from the two sets of materials were combined for the reporting of results. Also, there was

Table 1. *Group means*[1] *(and standard errors) for the recall and application tasks*

	Condition		
Task	**Cases-Text**[2]	**Text-Cases**[3]	**Text-Notes**[4]
Term Recall	2.71 (0.42)	7.53 (0.42)	5.35 (0.44)
Gist Recall	2.17 (0.31)	4.28 (0.31)	4.45 (0.33)
Transfer	2.60 (0.16)	3.37 (0.16)	3.30 (0.17)

[1]Adjusted for minutes spent studying the materials during Session One
[2]$n = 52$
[3]$n = 51$
[4]$n = 50$

no order effect, $F(1, 146) = 0.62$, $p = .43$, and no condition × order interaction, $F(2, 146) = 0.36$, $p = .70$; we may therefore conclude that participants' term recall was not affected by the order in which they studied the two sets of materials. To analyze group differences in recall we then used the same ANCOVA model, but with the "raw" scores as the dependent variable, rather than percentage scores.

The omnibus ANCOVA for differences in term recall among the three conditions (with study time as the covariate) was statistically significant, $F(2, 146) = 36.20$, $p < .001$, allowing us to conduct the three pairwise comparisons, with a Type I error probability of .05 for each comparison.

The omnibus ANCOVA for group differences in gist recall (with study time as the covariate) was also statistically significant, $F(2, 146) = 16.06$, $p < .001$, permitting us to conduct the three pairwise comparisons, with a Type I error probability of .05 for each comparison. Again, there were no differences between the two sets of materials, $F(1, 146) = 0.59$, $p = .45$; therefore, the gist recall data were combined for the reporting of results. There was no order effect for gist recall, $F(1, 146) = 0.24$, $p = .62$, and no condition × order interaction, $F(2, 146) = 0.74$, $p = .48$.

Cases-Text vs. Text-Cases

The reconstructive hypothesis predicts that the text will be better remembered in the cases-text condition than in the text-cases condition, since the intervening contrasting-cases activity in the text-cases condition may tend to reinterpret the previously read text and thus cause some distortion in memory for the text. The indexical hypothesis also predicts that better recall will occur in the cases-text condition, since engaging in the contrasting video case activity (a perceptual experience) prior to reading the text is likely to increase comprehension of, and consequently memory for, the text. However, the schema-elaboration hypothesis predicts that building a schema by first reading the text, followed by the contrasting-cases activity that encourages elaboration of that schema, will yield better recall than first building a schema while

contrasting some cases and then reading a text that may conflict with the schema developed during the contrasting-cases activity.

Therefore, using the hypothesis-testing logic described earlier, if recall is greater in the text-cases condition, relative to the cases-text condition, then neither the reconstructive hypothesis nor the indexical hypothesis is supported (the null hypothesis is rejected), but the conflicting hypothesis (schema elaboration) still stands.

The null hypothesis was rejected for the term recall data, $t(101) = 14.25$, $p < .001$, effect size = 1.63, indicating that the reconstructive and indexical hypotheses failed to predict participants' term recall for this comparison. Participants in the text-cases condition recalled more terms from the text than participants in the cases-text condition, a result consistent with the schema-elaboration hypothesis.

The null hypothesis was also rejected for the gist recall data for this comparison, $t(101) = 11.17$, $p < .001$, effect size = 0.96, indicating that the reconstructive and indexical hypotheses also failed to predict gist recall between the cases-text and text-cases conditions. Participants in the text-cases condition had better gist recall of the text than participants in the cases-text condition, again consistent with the schema-elaboration hypothesis.

Cases-Text vs. Text-Notes

The indexical hypothesis predicts that, due to the perceptual experience provided by the contrasting video case activity before reading the text, recall is likely to be higher in the cases-text condition than in the text-notes condition. The other theories under consideration have no particular prediction for this comparison. Consequently, if recall is greater in the text-notes condition than in the cases-text condition, the indexical hypothesis is not supported (the null hypothesis is rejected).

The null hypothesis was rejected for term recall, $t(100) = 7.74$, $p < .001$, effect size = 0.89, indicating that the indexical hypothesis failed to predict term recall for this comparison. Participants in the text-notes condition recalled more terms from the text than participants in the cases-text condition.

The null hypothesis was also rejected for gist recall, $t(100) = 11.94$, $p < .001$, effect size = 1.03, indicating that the indexical hypothesis also failed to predict gist recall between the cases-text and text-notes conditions. Participants in the text-notes condition had better gist recall of the text than participants in the cases-text condition.

These mean differences, especially noting the large effect sizes, are certainly worthy of the attention of educators who think that the mere inclusion of video cases in their instructional activities will encourage better learning than more standard study techniques such as taking notes on a text! It is, of course, possible to challenge this conclusion with arguments that save the indexical hypothesis in light of this evidence. For example, one might argue that the text activated sufficient perceptual knowledge to allow indexing, even though a video was not employed. These and other such arguments might be addressed in future research. In this particular reading situation, however, the cognitive processing does not appear to require engineering of a related perceptual experience.

Text-Cases vs. Text-Notes

Similar to the previously described rationale predicting poorer text recall for the text-cases condition under the reconstructive hypothesis, the text-cases condition was also expected to be inferior to the text-notes condition, due to the fact that the text-notes condition had no intervening activity to potentially reinterpret the text. As before, the schema-elaboration hypothesis predicts that reading a text first, followed by contrasting some video cases, encourages development and elaboration of a sophisticated schema, which will be superior to a schema that is built from a single source (i.e., reading a text and then taking notes on that same text). The indexical hypothesis has no particular prediction in the absence of an enhanced perceptual experience and so was not considered in this comparison.

If recall is greater in the text-cases condition, relative to the text-notes condition, then the reconstructive hypothesis is not supported (the null hypothesis is rejected), but the competing hypothesis (schema elaboration) still stands.

For term recall, the null hypothesis was rejected, $t(99) = 6.43$, $p < .001$, effect size = 0.74, indicating that the reconstructive hypothesis again failed to predict term recall in this situation. Participants in the text-cases condition recalled more terms from the text than participants in the text-notes condition, again consistent with the schema-building hypothesis.

The null hypothesis was not rejected for gist recall, $t(99) = 0.87$, $p = .36$, indicating that the results may or may not provide support for the reconstructive hypothesis.

Application Tasks

To quantify participants' transfer of text ideas to a new situation, the learning goals (*understanding*) and assessment plan (*assessment*) data were examined for evidence that participants were using ideas from the related reading in their responses (e.g., the *understanding* application task was inspected for the facets of understanding); 1 point was awarded for each unique concept used appropriately in context, regardless of whether the specific concept name was correctly recalled. No points were given for concepts used incorrectly. Partial credit (.5 points) was allowed on this task for participants whose responses demonstrated limited knowledge or inadequate transfer of the concept.

Again, a Fisher LSD approach was used for the analysis. In order to check for differential performance on the two sets of materials, and to detect any effects in the order of presentation of the two sets of materials, an ANCOVA model was used with the transfer data converted to percentages. No inter-material differences in transfer were found, $F(1, 146) = 1.10$, $p = .30$; therefore, the data from the two sets of materials were combined for reporting the results. Oddly, an order effect was found, $F(1, 146) = 4.71$, $p = .03$, with transfer performance being higher when the *assessment* materials were presented first ($M = .32$, $SE = .01$) than when the *understanding* materials were presented first ($M = .28$, $SE = .01$). This order effect disappears when the *understanding* and *assessment* materials are analyzed separately: *understanding,* $F(1, 146) = 1.35$, $p = .25$; *assessment,* $F(1, 146) = 3.29$, $p = .07$. There was no condition × order interaction, $F(2, 146) = 0.66$, $p = .52$.

The omnibus ANCOVA for between-group differences in transfer (with study time as the covariate) was statistically significant, $F(2, 146) = 7.35$, $p = .001$, allowing us to conduct the three pairwise comparisons, with a Type I error probability of .05 for each comparison. The

predictions for each of these comparisons for this task, along with the results, will be presented next.

Cases-Text vs. Text-Cases

The reconstructive hypothesis predicts that transfer of the ideas in the reading will be better in the text-cases condition than in the cases-text condition, because video cases would activate prior knowledge as an integrative context for recalling text ideas, making them a more unified part of a student's prior knowledge. The schema-elaboration hypothesis also predicts that more transfer will occur in the text-cases condition; however, the mechanism responsible for this prediction is the elaboration of the schema that was developed during the reading of the text. As before, the indexical hypothesis predicts that better performance—including transfer of text ideas to new situations—will consistently be observed in the cases-text condition, where the contrasting video case activity provides a perceptual experience that enhances comprehension of a subsequently read text.

If transfer is greater in the cases-text condition, relative to the text-cases condition, then neither reconstructive theory nor schema-elaboration theory is supported (the null hypothesis is rejected), but the competing hypothesis (the indexical hypothesis) still stands.

The null hypothesis was not rejected for this comparison, $t(101) = 8.06$, $p = .99$, indicating that there is no evidence against either the reconstructive hypothesis or the schema-elaboration hypothesis.

Cases-Text vs. Text-Notes

Again, the indexical hypothesis predicts that transfer in the cases-text condition will be better than in the text-notes condition, where participants have no perceptual background knowledge to enhance their comprehension of the text. The other hypotheses under consideration have no obvious predictions for this pairwise comparison. Therefore, if transfer is greater in the text-notes condition than in the cases-text condition, the indexical hypothesis is not supported.

The null hypothesis was rejected for this comparison, $t(100) = 7.27$, $p = .001$, effect size = 0.49, indicating that in this situation, the indexical hypothesis failed to predict the outcome of this comparison.

Text-Cases vs. Text-Notes

The reconstructive hypothesis predicts that participants in the text-cases condition will be better able to transfer ideas from the reading than will participants in the text-notes condition, due to the intertwining of text ideas with video case material, through a cognitive mechanism in which what has been previously experienced is interpreted by what is currently being experienced. In the text-notes condition, the text is the only viewpoint and thus does not encourage as much integration with prior knowledge, according to the reconstructive view. The schema-elaboration hypothesis also predicts that more transfer will occur in the text-cases condition than in the text-notes condition because of the elaborated schemas developed by the contrasting-cases activity. The indexical hypothesis has no particular prediction for this comparison, since no perceptual experience precedes the text in either of these conditions.

Our predictions for this comparison specified that if transfer is greater in the text-notes condition, relative to the text-cases condition, then neither reconstructive theory nor schema-elaboration theory is supported (the null hypothesis is rejected).

The null hypothesis was not rejected for this comparison, $t(99) = 0.70$, $p = .97$, indicating that there is no evidence here against either the reconstructive hypothesis or the schema-elaboration hypothesis.

Discussion

The results of this study provide some interesting insights into the burgeoning educational literature on video cases. In the present discussion, we will first briefly consider some implications for instructional design suggested by our results, and then re-examine the theories that we thought likely to be helpful in predicting the outcome of contrasting video case activities.

The primary take-home message regarding the use of video cases in instruction is that if video-case activities are designed appropriately, they can foster better learning than more traditional methods of study, such as taking notes on a text. This was demonstrated in our comparisons of term recall between the text-cases and text-notes conditions. However, this endorsement of video cases is not carte blanche: The students in our study who took notes on a text *outperformed* the students who contrasted video cases and then read the same text. Regardless of the reason(s) for this outcome, the message emerging is that including video cases in instructional activities *must not* be done in a mindless or random fashion if learning some text material or the effective transfer of text ideas are among the goals of instruction. Perhaps the cases-text condition could have been modified in some way to promote the kinds of learning that we wanted to encourage; nevertheless, the point remains that teachers must not assume that just because they teach with video cases, students' learning will be enhanced.

Even though students in the cases-text condition did not recall the text ideas as well as students in the text-cases condition, the students in these two video-based conditions overwhelmingly reported that the case-comparison activity was interesting (98%), enjoyable (93%), and helpful (96%)—not even one person reported being bored with the video-case activity, disliking it, or that it interfered with their learning. Yet despite their positive attitude toward the activity, some of these students (specifically, the students in the cases-text condition) were outperformed by their peers who took notes on a text. One of the implications of this finding for education is that even though students may enjoy a particular learning approach, it may not pay off in the long run, relative to other instructional approaches.

We now move on to consider the theories that we subjected to severe testing under the null hypotheses of our various pairwise comparisons. We wish to make it very clear at this point that we are in no way evaluating the "absolute" worth of these theories; indeed, we fully recognize that each of them is very promising in the context in which it was developed. Our goal is simply to examine these theories in terms of their fruitfulness for predicting the outcomes of contrasting video case activities in the kind of context represented in our work, and, consequently, for guiding the design of other learning activities that include video cases in similar contexts. The specific context to which these findings apply is when a readable text is used and the instructional goals include better memory for text and application of text ideas.

The indexical hypothesis does not offer guidance in this context. In every test where the indexical hypothesis was included in the null hypothesis, the null hypothesis was rejected,

indicating a failure of the indexical hypothesis to predict the results of the comparison. Again we wish to emphasize that this is not a limitation of the indexical hypothesis; perhaps there were some implicit assumptions in this hypothesis that were not met in our study. One such assumption might be that the indexical relationship between the perceptual object and the text should be obvious; this was certainly not the case in our study, as we (intentionally) made no effort to point out the conceptual correspondence between the video cases and the text. Thus, students in the cases-text condition may not have been indexing in a manner that was consistent with the forthcoming text, and that may have hindered their learning. Another explanation for why the indexical hypothesis did not predict the outcome in our study is that the perceptual basis for indexing might have been supplied by the text so that an enhanced perceptual treatment was not needed.

The reconstructive hypothesis, originally our favored predictor of learning outcomes for video case activities, based on a pilot project, survived only somewhat better than the indexical hypothesis. The null hypothesis was rejected in three out of six tests where the reconstructive hypothesis was represented in the null hypothesis. In two of the three tests where the reconstructive hypothesis was not rejected, the schema-elaboration hypothesis was also not rejected. However, in none of the tests we conducted across the entire experiment can we claim that we have results that are consistent with the reconstructive hypothesis.

The schema-elaboration hypothesis, which predicts both transfer and enhanced text memory in the text-cases condition, does, however, look promising. Although it was not subjected to severe testing (in the null hypothesis) as frequently as the other two hypotheses just discussed, the schema-elaboration hypothesis emerged from our analyses completely unscathed. The null hypothesis was not rejected in either of the two tests where the schema-elaboration hypothesis was represented in the null hypothesis (together with the reconstructive hypothesis in both of these tests). Moreover, in three additional tests the null hypothesis *was* rejected when a prediction consistent with the schema-elaboration hypothesis was represented in the alternative hypothesis. So what went right? Could it be that in the cases-text condition, the schema that was developed while contrasting the video cases was conceptually in conflict with the text? This would explain the superior recall (both term and gist) in the text-cases condition relative to the cases-text condition. More rigorous tests are needed in future studies to fully examine the validity of the schema-elaboration hypothesis in a contrasting video cases situation.

In sum, carefully designed video-case activities can result in significant learning. The participants in our study who read a text and then contrasted some video cases were better able to remember the details of the text than participants who had taken notes on the text. However, not all video-case activities are created equal; when a contrasting-case activity is used before an easy-to-understand text is presented, students may not benefit from the video cases, and may actually perform *worse* on tests of recall and transfer of text ideas, relative to peers who simply take notes on a text. Our quest for a theory has led us to consider the schema-elaboration hypothesis as the best candidate for predicting the outcomes of a contrasting video case activity, at least under conditions that match those of our study: readable text, and instructional goals that include recall and transfer of text ideas. Nevertheless, when working with materials as complex as video cases, much has yet to be discovered.

References

Abell, S. K. (1996). Interactive video cases developed for elementary science methods courses. *TechTrends, 41,* 20-23.

Bartlett, F. C. (1932). *Remembering: A study in experimental and social psychology*. Cambridge, England: The University Press.

Bransford, J. D., Goin, L. I., Hasselbring, T. S., Kinzer, C. K., Sherwood, R. D., & Williams, S. M. (1986). Learning with technology: Theoretical and empirical perspectives. *Peabody Journal of Education, 64,* 5-26.

Copeland, W. D., & Decker, D. L. (1996). Video cases and the development of meaning making in preservice teachers. *Teaching and Teacher Education, 12,* 467-481.

Derry, S. J. (in press). eSTEP as a case of theory-based web course design. To appear in A. M. O'Donnell & C. Hmelo (Eds.), *Collaboration, reasoning, and technology*.

Derry, S. J., Siegel, M., Stampen, J., & STEP Team (2002). The STEP system for collaborative case-based teacher education: Design, evaluation & future directions. In G. Stahl (Ed.), *Proceedings of Computer Support for Collaborative Learning (CSCL) 2002* (pp. 209-216). Hillsdale, NJ: Lawrence Erlbaum Associates.

Garner, W. R. (1962). *Uncertainty and structure as psychological concepts*. New York: John Wiley and Sons.

Garner, W. R. (1974). *The processing of information and structure*. New York: Lawrence Erlbaum Associates.

Gentner, D. (1983). Structure-mapping: A theoretical framework for analogy. *Cognitive Science, 7,* 155-170.

Gentner, D., Loewenstein, J., & Thompson, L. (2003). Learning and transfer: A general role for analogical encoding. *Journal of Educational Psychology, 95,* 393-408.

Gibson, E. J. (1969). *Principles of perceptual learning and development*. New York: Appleton-Century-Crofts.

Gibson, J. J. (1979). *The ecological approach to visual perception*. Boston: Houghton Mifflin.

Gibson, J. J., & Gibson, E. J. (1955). Perceptual learning: Differentiation or enrichment? *Psychological Review, 62,* 32-41.

Gick, M. L., & Holyoak, K. J. (1980). Analogical problem solving. *Cognitive Psychology, 12,* 306-355.

Gick, M. L., & Holyoak, K. J. (1983). Schema induction and analogical transfer. *Cognitive Psychology, 15,* 1-38.

Glenberg, A. M., & Robertson, D. A. (1999). Indexical understanding of instructions. *Discourse Processes, 28,* 1-26.

Goldman, E., & Barron, L. (1990). Using hypermedia to improve the preparation of elementary teachers. *Journal of Teacher Education, 41*(3), 21-31.

Hannah, C. L. (1995). Self-study: Students evaluate the use of video "cases" in an educational psychology course with a field component. *Journal of Technology and Teacher Education, 3,* 267-279.

Hewson, P. W., Beeth, M. E., & Thorley, N. R. (1998). Teaching for conceptual change. In B. J. Fraser & K. G. Tobin (Eds.), *International handbook of science education* (pp. 199-218). London: Kluwer Academic Publishers.

Kurtz, K. J., Miao, C.-H., & Gentner, D. (2001). Learning by analogical bootstrapping. *Journal of the Learning Sciences, 10,* 417-446.

Lampert, M., & Ball, D. L. (1998). *Teaching, multimedia, and mathematics: Investigations of real practice*. New York: Teachers College Press.

Levin, J. R. (2003, SpringRandom thoughts on the (in)credibility of educational-psychological intervention research. *Newsletter for Educational Psychologists, 26,* 1, 10-11.

Levin, J. R., & O'Donnell, A. M. (1999). What to do about educational research's credibility gaps? *Issues in Education, 5,* 177-229.

Levin, J. R., Serlin, R. C., & Seaman, M. A. (1994). A controlled, powerful multiple-comparison strategy for several situations. *Psychological Bulletin, 115,* 153-159.

Liedtka, J. (2001). The promise and peril of video cases: Reflections on their creation and use. *Journal of Management Education, 25,* 409-424.

Merseth, K. K., & Lacey, C. A. (1993). Weaving stronger fabric: The pedagogical promise of hypermedia and case methods in teacher education. *Teaching and Teacher Education, 9,* 283-299.

O'Donnell, A. M., & Levin, J. R. (2001). Educational psychology's healthy growing pains. *Educational Psychologist, 36,* 73-82.

Posner, G. J., Strike, K. A., Hewson, P. W., & Gertzog, W. A. (1982). Accommodation of a scientific conception: Toward a theory of conceptual change. *Science Education, 66,* 211-227.

Risko, V. J., Yount, D., & Towell, J. (1991). Video-based CASE analysis to enhance teacher preparation. In T. V. Rasinski, N. D. Padak, & J. Logan (Eds.), *Reading is knowledge* (pp. 87-96). Pittsburg, KS: College Reading Association.

Royer, F. L. (1966). Figural goodness and internal structure in perceptual discrimination. *Perception and Psychophysics, 1,* 311-314.

Schwartz, D. L., & Bransford, J. D. (1998). A time for telling. *Cognition and Instruction, 16,* 475-522.

Spiro, R. J. (1977). Remembering information from text: The "state of schema" approach. In R. C. Anderson, R. J. Spiro, & W. E. Montague (Eds.), *Schooling and the acquisition of knowledge* (pp. 137-165). Hillsdale, NJ: Lawrence Erlbaum Associates.

Spiro, R. J., Vispoel, W. P., Schmitz, J. G., Samarapungavan, A., & Boerger, A. E. (1987). Knowledge acquisition for application: Cognitive flexibility and transfer in complex content domains. In B. K. Britton & S. M. Glynn (Eds.), *Executive control processes in reading* (pp. 177-199). Hillsdale, NJ: Lawrence Erlbaum Associates.

Stephens, L., Leavell, J., Fabris, M., Buford, R., & Hill, M. (1999). Producing video-cases that enhance instruction. *Journal of Technology and Teacher Education, 7,* 291-301.

Stigler, J. W., & Hiebert, J. (1999). *The teaching gap: Best ideas from the world's teachers for improving education in the classroom*. New York: Free Press.

Thompson, L., Gentner, D., & Loewenstein, J. (2000). Avoiding missed opportunities in managerial life: Analogical training more powerful than individual case training. *Organizational Behavior and Human Decision Processes, 82,* 60-75.

Wiggins, G., & McTighe, J. (1998). *Understanding by design*. Alexandria, VA: Association for Supervision and Curriculum Development.