

Gary R. Morrison and Steven M. Ross

Abstract

Instructional technology research is broad both in terms of topics and explorations of basic and applied research. In this chapter, we examine various types of stimulus materials that instructional technology researchers have used to study different phenomena. Specifically, we discuss and illustrate how the choice of stimulus material (e.g., actual lesson content, pictures, prose, etc.) directly influences the internal validity (rigor) and external validity (generalizability) of the findings. While randomized experiments are considered the so-called gold standard (Slavin, *Educational Researcher* 37(1):5–14, 2008) of educational research, particularly for evaluating the effectiveness of instructional strategies, these studies may employ artificial or novel stimulus materials that can limit generalization of the results. Since one goal of instructional technology research is to provide evidence that allows the instructional designer to generate heuristics easily applicable (i.e., generalized) to new situations, studies with strong external validity should be highly desired. Similarly, there are also instances where initial studies need to be designed with high internal validity, sometimes at the sacrifice of external validity, to control for extraneous variables. Using selected studies as illustrative examples, this chapter examines how validity has been addressed in instructional technology research.

Keywords

Research • Educational psychology • Educational technology • Stimulus materials • Internal validity • External validity

Introduction

The field of instructional technology began developing a rich knowledge base of research studies focusing on instructional technology with the start of *Audio-Visual Communications*

G.R. Morrison (✉)
Old Dominion University, 228 Education Building,
Norfolk, VA 23529, USA
e-mail: gmorriso@odu.edu

S.M. Ross
Johns Hopkins University, CRRE, 200 W. Towsontown Road,
Baltimore, MD 21204, USA
e-mail: sross19@jhu.edu

Review (AVCR) in 1954. While the field has mostly avoided meaningless media comparison studies in recent years (Clark, 1983), the breadth of topics continues to grow. Our earlier analysis of research methodologies employed in articles published in *AVCR*, *Educational Communications and Technology Journal (ECTJ)*, and *Educational Technology Research & Development (ETR&D)* (Ross & Morrison, 1996, 2004) found that the trends in the use of methodologies have changed over time. For example, time series studies dominated the first 10 years of *AVCR* publication, but have all but disappeared from *ETR&D*. In contrast, we have seen a steady increase of studies employing true experimental designs that was the dominant methodology from 1973 to 2001. More recently, we have examined the number

of studies classified as intervention research (Levin, 2004), that is, studies designed to compare two different instructional treatments such as immediate feedback compared to delayed feedback. We found a steady decline (Ross et al., 2008) in intervention studies in ETR&D similar to the trend in educational psychology journals (Hsieh et al., 2005).

In this chapter, we focus on how instructional technology researchers have designed the stimulus materials used in their studies to strengthen either the internal or external validity of findings. For readers who desire a more in-depth discussion of quantitative and qualitative methods, we suggest the various chapters in all four editions of this handbook. In the following section, we start with a brief discussion of internal and external validity issues in instructional technology research. Then, we examine the design of stimulus materials in studies with high internal validity and studies with high external validity. Last, we address the issue of generalization of results in instructional technology studies based on both the choice of stimulus materials and the degree to which the study participants mindfully engage with the material to be learned.

Validity Issues in Instructional Technology Research

Experimental research in education and psychology values studies establishing high internal validity to eliminate any unintended variables influencing the results (Ross & Morrison, 2004). According to Slavin (2008), researchers can further maximize high internal validity by using a randomized assignment of participants to treatments to eliminate systematic error. The quest for high internal validity orients researchers to design experiments in which treatment manipulations can be tightly controlled. In the process, using naturalistic conditions (e.g., real classrooms) becomes challenging, given the many extraneous sources of variance that are likely to operate in those contexts. For example, the extensive research conducted on verbal learning in the 1960s and 1970s largely involved associative learning tasks using simple words and nonsense syllables (Paivio, 1971; Underwood, 1996). With simplicity and artificiality comes greater opportunity for control of the variables.

This orientation directly supports the objectives of the basic educational psychology researcher whose interests lie in testing the generalized theory associated with treatment strategies, independent of the specific methods used in their administration. Educational technology researchers, however, are interested in the interaction of medium and method or instructional strategy, or simply the instructional strategy (Bernard et al., 2004; Bernard et al., 2009; Clark, 2001; Kozma, 1991, 1994; Ullmer, 1994). To learn about this interaction, realistic instruction rather than artificial or contrived instruction needs to be used. In other words, external validity becomes as important a concern as internal validity.

Discussing these issues brings to mind a manuscript that one of us was asked to review a number of years ago for publication in an educational research journal. The author's intent was to compare, using an experimental design, the effects on learning from programmed instruction and computer-based instruction (CBI). To avoid Clark's (1983) criticism of performing a media comparison, i.e., confounding media with instructional strategies, the author decided to make the two treatments as similar as possible in all characteristics except delivery mode. This task essentially involved replicating the exact programmed instruction design in the CBI condition. Not surprisingly, the findings showed no difference between treatments, a direct justification of Clark's position. But, unfortunately, this result (or one showing an actual treatment effect as well) would be meaningless for advancing theory or practice in educational technology. By stripping away the special attributes of a normal CBI lesson (e.g., interaction, sound, adaptive feedback, animation etc.), all that remained were alternative forms of programmed instruction and the unexciting finding, to use Clark's metaphor, that groceries delivered in different, but fundamentally similar ways still have the same nutritional value. Needless to say, this study, with its high internal validity but very low external validity, was evaluated as unsuitable for publication.

Stimulus Materials in Studies with High Internal Validity

Studies in instructional technology research that require high internal validity often focus on attributes of a medium such as on the legibility of projected materials (Adams, Rosemier, & Sleeman, 1965; Snowberg, 1973) or the design of CBI screens and materials (Acker & Klein, 1986; Grabinger, 1983; Morrison, Ross, Schultz, & O'Dell, 1989; Ross, Morrison, & Odell, 1988). Similarly, studies examining imagery (McManis, 1965; Noble, 1952) or exploring how individuals learn relationships from a diagram (Winn & Solomon, 1993) may use an experimental design with high internal validity to control for other variables. When designing these studies, the researchers must decide if internal or external validity is of greater importance. For example, consider the text in Fig. 3.1 which a researcher might use to investigate the emotional meaning of a particular typeface. In the first row, a real word is displayed in the two different



Fig. 3.1 Comparison of two types of stimulus materials

typefaces. If the participants indicated that the typeface on the left was light and elegant, reviewers might question the interpretation because of the word jewelry. Similarly, the word muscle printed in a bold, heavy font would also confound the interpretation of the meaning of the typeface. The second row uses nonsense words that have no meaning and allows the researcher to conclude that any meaning derived from the rating is due to the typeface. The first row of words has high external validity because of the use of real words; however, the words may influence the participants' rating of the typeface. The second row has high internal validity, but generalizing the results raise additional questions for application. Specifically, typefaces are rarely used in the absence of words and phrases having meaning. It thus seems highly probable that the emotional valences determined using nonsense words would vary (perhaps even considerably) when the same typefaces were employed with instructional text or popular literature.

Given these options, instructional technology researchers initially may decide to establish a theoretical construct (e.g., emotional connotation of type) by using the second row of stimuli. Thus, internal validity would be emphasized over external validity for this basic research study. After establishing the construct, they may design applied studies using materials with a high external validity to test the application of the construct in a more realistic context. In the following section, we illustrate these trade-offs by examining several studies that focus on media attributes and the type of stimulus materials they employed.

Using Artificial Materials in Studies of Media Variables

An example of a highly controlled study is one conducted by Snowberg (1973) examining the use of background colors in projected media. One of the concerns expressed by Snowberg was that the selection of colors as backgrounds for the slides offers almost limitless possibilities. To address this problem, Snowberg selected a range of color filters that allowed for replication. Additional neutral filters were combined with the color filters so that each background was of the same brightness or luminance, thus avoiding a difference between background colors. Ten letters for the stimulus materials were taken from a Snellen chart to create a chart similar to those used by optometrist to check visual acuity. By controlling the five colors, providing for brightness control, and using standardized letters; Snowberg was able to isolate the legibility of projected letters on various colored backgrounds. If real words were used, the participants could possibly have identified or guessed the word based on a few letters, thus reducing the number of possible answers. By using individual letters, the participant had to distinguish between letters

such H, D, N, O, and C. This controlled study allowed the researcher to examine the media attribute, the effect of background color on letter legibility, while controlling for confounding variables.

While these recommendations provide seemingly useful guidelines for selecting backgrounds for the best legibility, other color background variations could provide more aesthetically pleasing colors and larger more readable text. That is, one would seldom need to use a small (minimal legibility) font with black text on a white background, which was found to be the minimally legible combination. Thus, replicating this basic design using realistic materials and other background colors would be a logical extension of Snowberg's study. For a typical classroom, you might not need maximum legibility, but rather acceptable readability and an aesthetically pleasing display. An applied research study might also determine that attentiveness is also contextually (e.g., schools' colors) or gender (e.g., pink vs. blue as a preference) linked. Nonetheless, Snowberg's findings are valuable for establishing basic legibility principles that are minimally contaminated by extraneous variables.

In another study of legibility, Adams et al. (1965) studied the legibility of typewritten fonts projected on a white background. They also used letters from a Snellen chart and created stimulus slides consisting of five different type sizes ranging from 3/32 to 8/32 of an inch. Participants were elementary school students who were asked to judge the slides from distances of 20, 25, 30, and 40 feet from the screen in a darkened room. Adams et al. concluded the two smaller type sizes should be avoided, particularly if the viewing distance was beyond 20 feet. Findings indicated that letters at least 6/32 to 8/32 of inch (about 14–18 points) should be used.

These two studies (Adams et al., 1965; Snowberg, 1973) address questions of legibility of projected visuals. Both focused on recognizing individual letters (legibility) rather than words (readability) (Craig & Bevington, 2006). The results establish the color combination or letter size with the best legibility. Similarly, both Snowberg (1973) and Adams et al. (1965) have identified the *smallest* font one should use. These studies raise the question of whether a study using realistic words and sentences would produce similar results, especially if it examined larger font sizes rather than the minimum specified. How this question is answered directly bears on the external validity of the original (basic research) findings. For example, a typical classroom would not have the lighting controls used by Snowberg (1973) for either projection or ambient light. Thus, assuming that the brighter ambient lighting in a typical classroom would reduce the contrast between the words and lettering, we might find that a larger font size is needed.

An extension of this research (Aslan, Watson, & Morrison, 2011) is a study in progress in which participants use a paired-comparison technique to select the PowerPoint slide

design they most prefer. The slides were designed using 20-, 24-, 28-, and 32-point text with realistic material (bonsai art), but unrelated to the interests of the participants. The researchers were not interested in the smallest legible text, but rather an optimal-sized text. As the font size increases, the number of words and the length of each phrase on a line become shorter. Thus, the contextual support is also reduced (Ross & Morrison, 1989) when font size is increased. As an extension of the basic research studies reviewed, additional studies using realistic materials in natural settings are needed to find the balance between the smallest legible font and a readable font that provides adequate contextual support using aesthetically pleasing color combinations.

Using Artificial Materials to Study Learning

To control for prior knowledge, many studies examining serial learning and imagery have used nonsense words (McManis, 1965; Noble, 1952). Instructional technology researchers have adopted other approaches to control for internal validity in applied research. In a study of the effect of concrete-verbal and visual information on mental imagery, Clark (1978) selected abstract geometric figures for participants to reproduce. Participants were presented either (a) picture only, (b) printed instructions for creating the drawing, (c) audio only instructions, (d) audio with pictures, (e) audio and video of the instructor giving directions, or (f) audio instructions while showing the instructor. Participants then reproduced from memory the drawing described in the stimulus materials. The general hypothesis was that dual channel presentations would be more effective. By using abstract geometric figures that were the equivalent of nonsense words, Clark could increase internal validity by controlling for prior knowledge of the image.

When studying the effectiveness of objectives, overviews, or inserted questions, the stimulus materials require one or more pages of meaningful textual information so the participant can answer test questions. However, the meaningful text introduces a confounding variable that can threaten internal validity as the participants may have relevant prior knowledge. Consider the study by Hannafin, Phillips, Rieber, and Garhart (1987) who examined two different types of orienting strategies on learning. Participants received either a behavioral orienting strategy that directed them to focus on a specific name, place, or date; or a cognitive strategy that directed them to focus on a broader topic such as culture. The control group was advised simply to pay attention to the material. Given the nature of this study, careful consideration was needed for selection of the stimulus material. For example, if they were to select a chapter from a science textbook on the solar system, some students might have prior knowledge they could use to answer the items on the pretest. The use of nonsense words or

even a foreign language as used in Ho's (1984) or Winn, Li, and Schill (1991) studies is not practical when students must learn from textual materials.

To reduce the threat to internal validity, Hannafin et al. (1987) used a fictitious story that included realistic scientific, cultural, political, and geographic elements to create a plausible story line. This contrived story allowed participants to apply intact scientific knowledge to a novel topic. Results indicated that the behavioral and cognitive strategies were more effective for factual learning while the control group showed superior performance for inferential learning. Two explanations of these results were offered. First, students revert to their own preferred approach for learning and ignore the recommended strategy. Second, the orienting activities were ineffective because the materials included sound design features that reduced the effectiveness or need for an orienting strategy. By using a fictitious, but realistic scenario Hannafin et al. were able to reduce the threat to internal validity from prior knowledge and increase the external validity by using contrived, but realistic appearing materials.

While artificial stimulus material allows the researcher to control for other variables such as prior knowledge, generalization of the results therefore is more limited. To the degree that instructional technology research is expected to inform practice, an impact that some researchers have questioned (Ross, Morrison, & Lowther, 2010), it would seem the use of realistic material in natural settings would be more valuable in using technology as a teaching and learning tool.

Stimulus Materials in Studies with High External Validity

Examples of progressing from highly controlled to more realistic application contexts come from CBI research. CBI tends to present information on individual screens with the learner having the capability to navigate between screens rather than scrolling through the instruction as one might do with electronic text.

From Basic to Applied Research: Contrasting Internal and External Validity

When an individual screen design (or frame) is used to present the stimulus material or the instruction, there is a limited number of characters or words the designer can include on a single frame much like we are limited to how many characters or words we can type on a single sheet of paper with 1 in. margins and 12 point Times Roman font. Grabinger (1983) was one of the first to study screen design layout for CBI. To control for confounding variables, Grabinger created stimulus screens consisting of x's and o's

(e.g., XxxooxxxxooXxxooxxxxoo) to control for any meaning the message might include that could influence the participants preference for the design. Participants were shown two different designs on identical monitors side-by-side and asked to indicate which one they preferred. Results were similar to those for printed instruction (Dair, 1967), indicating a preference for large amounts of white space and screens with sparse amounts of text.

Using Grabinger's (1983) research as a starting point, we conducted several studies to extend the original research to realistic materials. In the first study, Morrison et al. (1989) used realistic stimulus materials to test Grabinger's findings. Several authors in addition to Grabinger suggested the use of white space for CBI screen design as the designer was no longer constrained by properties of the printed page (Allessi & Trollip, 1985; Bork, 1987; Hooper & Hannafin, 1986). However, as the amount of white space increases on the screen, the amount of information decreases requiring the reader to read additional screens to obtain the same amount of information. The first study by Morrison et al. examined learner preferences for screen density when realistic instructional materials were used. A lesson from a unit on measures of central tendency was selected. To allow for replication and application, we used a measure of screen density that calculated the maximum number of characters that could be displayed on a screen and then divided the actual number of characters to arrive at a screen density percentage creating four different density levels. Two designs were shown one at a time in a random order for a total of six pairings.

The results indicated that participants preferred the 31% density screen over the others. It appears that participants desired greater contextual support when viewing realistic materials than when viewing artificial designs that lacked meaning. The Morrison et al. (1989) study extended Grabinger's work through the use of high external validity materials to test the assumptions in a realistic setting. Importantly, it supported somewhat different design principles, namely, that density reduction and contextual support need to be balanced to maximize readability.

Comparing Internal and External Validity in a Single Study

The results of the two previous studies raised additional questions. For example, as the density (i.e., number of words on the screen) increases, the number of screens needed to read the same materials decreases. At first glance, it would seem logical to have the participant review *all* the screens for each density level (one to four screens depending on the density level). However, if the participants tended to select the higher density screens, one might conclude it was the easier choice since they only had to review one or two screens.

To determine if the number of screens viewed would influence the preference, two additional treatments were added. In the first treatment, participants *only* viewed the first screen for each density level. In the second treatment, participants were required to review all screens for a density level before making a choice. In this study, Ross, Morrison, and Schultz (1995) compared realistic materials, approximation to English (ATE) (nonsense words with same letter pattern as English), and nonsense notation (x's and o's) used by Grabinger (1983). The realistic materials were the same used by Morrison et al. (1989). Four different screen designs consisting of 53, 31, 26, and 22% density were employed, with each requiring 1, 2, 3, and 4 screens, respectively, to present the full content. The resulting design consisted of three types of text, four density levels, and two screen conditions (first screen only or all screens of the density level). The six comparisons of four density levels for a specific text type (realistic lesson, ATE, or nonsense) were presented in a random order and rated until all three text types were judged by each participant. Overall, the higher density screens were preferred for realistic materials while the lower density screens were preferred for the artificial text (ATE and nonsense). The results confirmed our hypotheses that students wanted more information on a single screen when viewing realistic materials, but preferred more white space when viewing nonrealistic or nonsense materials.

Using Realistic Learning Material to Increase External Validity

Tessmer and Driscoll (1986) investigated the effectiveness of a concept tree and narrative text for learning coordinate concepts with high school students taking physics. Stimulus materials that had multiple related concepts were needed for the study. It would have been extremely difficult to create fictitious stimulus materials of this complexity. Therefore, Tessmer and Driscoll selected a physics unit that the classroom teacher judged as unfamiliar to the students based on past performance. The stimulus materials were then created for each treatment based on realistic materials. The participants were given 20 min to read the treatment materials and then completed an immediate posttest followed by a delayed posttest. Participants in the concept tree treatment performed better on concept classification. Although using realistic material increased the risk that students' prior knowledge and experiences in the physics course would bias treatment effects, it significantly increased the external validity of the study and the implication that the concept tree could be a useful applied instructional strategy.

Another example of a study with high external validity is one conducted by Ross and Anand (1987) which used realistic instructional materials and personalized those materials for

one treatment group. Participants were fifth- and sixth-grade students who received stimulus materials that taught the procedures for dividing by fractions. The abstract treatment group received examples and problems that referred to items as quantity, fluid, liquid, and so forth. The concrete treatment group received examples and problems that substituted hypothetical concrete referents such as Bill, Joe, English, artist, etc. In the personalized treatment group, personal information collected from a biographical survey was inserted into the examples and problems so the participant saw his or her name, best friends' names, birth date, pet's name, and favorite candy. Participants in all three treatments received the same examples and problems; only the context used for presenting the examples and problems was modified by substitution of words. The results indicated that students in the personalized treatment performed significantly better on the context subtest and transfer test. By using realistic materials, the researchers provided evidence of the potential effectiveness of the personalization strategy for applied classroom use.

More recent examples of the use of realistic materials include the use of an existing problem-based learning unit from science (Song & Grabowski, 2006), a math unit on addition subtraction developed by the researchers (Kopcha & Sullivan, 2008), and the use of two different math units of which one was a commercial product (Roschelle et al., 2009). By using experimental or quasi-experimental designs, these studies combine moderate to high levels of both internal and external validity.

Realistic Materials and Incentives: Are They Adequate?

As researchers, it is easy (and comforting) to assume that if we use realistic materials that are relevant to our participants, such as a unit on momentum for students in a physics class or a unit on writing objectives for pre-service teachers, they will put forth the same effort to learn as they would if studying for a class. Given this assumption and the contradictory results in the research literature on feedback, we decided to explore whether feedback strategies would operate differently under varied incentive conditions for learning (Morrison, Ross, Gopalakrishnan, & Casey, 1995).

The 246 participants in the feedback study were drawn from two pre-service teacher education courses (Morrison et al., 1995). The instructional materials were designed to be relevant to students' academic preparation by focusing on writing behavioral objectives, the three domains of objectives, and the taxonomy of behavioral objectives. Students from each of the two classes were randomly assigned to one of five feedback treatments including knowledge of correct response (KCR), delayed with immediate knowledge of

response (e.g., correctness of answer), answer until correct (AUC), questions with no feedback, no questions or feedback. Participants from the first course were in the performance incentive group as they could use the score from the treatment to receive credit for a required unit on objectives. Participants in the second course were classified as the task incentive group as they received five bonus points for participating in the study. It was predicted that participants in the performance incentive group would show greater motivation to learn and mindfully use the feedback, particularly in the more complex (i.e., KCR and AUC) feedback treatments. This assumption was only partially supported. The performance incentive group did learn more and made greater use of the review opportunities after answering a question. However, differences between groups for selecting the option to review were not significant. When participants complete an artificial learning task as the task-incentive treatment, they may show little interest in mastering content or in using the instructional resources such as feedback. One concern for researchers is how to motivate them to go beyond surface processing of the content and engage in a deeper level of processing that produces meaningful learning (or at least emulates real-life learning processes). While the performance incentive (substitute study performance for a course assignment) in the above study did appear to motivate the performance-incentive group to perform well, it was not enough to promote a deeper level of processing or make extensive use of the feedback. Thus, generalizability to real-life instructional contexts, where there is greater accountability for achievement, may be limited.

Conclusion

In this chapter, we have examined the use of stimulus materials in instructional technology research. Depending on the purpose of the research, the stimulus materials can range from artificial using nonsense symbols, to contrived materials using real words or text, and ultimately to realistic using actual lesson content. The selection of the type of stimulus materials is determined primarily by the focus of the research—verifying basic laws and principles of learning using technology or evaluating the effectiveness of applied instructional strategies using technology. Underlying the particular focus and concomitant selection of stimulus materials is the researcher's emphasis on addressing different types of validity concerns. Basic research studies rely primarily on materials that foster high internal validity by controlling extraneous variables relating to the learner characteristics and the learning context. Applied studies place a greater emphasis on external validity to allow for generalization of the results to real-life learning contexts. It is this trade off that often requires researchers to begin a new area of inquiry with a study emphasizing high

internal validity to isolate variables and phenomena. As a subsequent step, the laws and principles supported in the initial basic research are tested in realistic settings to determine their utility for different application contexts.

While the design of stimulus material directly influences the absolute and relative strengths of internal and external validity in a research study, the meaningfulness of the evidence obtained also depends on the degree to which the study participants mindfully engage with the instruction. That is, whether the material to be learned consists of nonsense symbols or material straight from the textbook currently being used, if participants' primary incentive is to earn extra credit points that are noncontingent on performance, both internal validity (i.e., appropriate treatment induction) and external validity (realistic learning conditions) are likely to be compromised. Instructional technology research needs to continue to focus on relevant and quality research that addresses issues relevant to the field and to education in general. Studies are needed that help practitioners solve practical problems. But unless the research designs employed establish sufficient rigor, the results may not accurately reflect the uses and impacts of the technology applications examined (Ross et al., 2010).

References

- Acker, S., & Klein, E. (1986). Visualizing spatial tasks: A comparison of computer graphic and full-band video displays. *Educational Technology Research and Development*, 34(1), 21–30. doi:10.1007/bf02768359.
- Adams, S., Rosemier, R., & Sleeman, P. (1965). Readable letter size and visibility for overhead projection transparencies. *Educational Technology Research and Development*, 13(4), 412–417. doi:10.1007/BF02766846.
- Allessi, S. M., & Trollip, S. R. (1985). *Multimedia for learning: Methods and development* (3rd ed.). Needham, MA: Allyn & Bacon.
- Aslan, A., Watson, G. S., & Morrison, G. R. (2011). *Determining the optimum font size for readability in PowerPoint presentations*. Unpublished manuscript, Instructional Design & Technology Program, Old Dominion University, Norfolk, VA.
- *Bernard, R. M., Abrami, P. C., Borokhovski, E., Wade, C. A., Tamim, R. M., Surkes, M. A., et al. (2009). A meta-analysis of three types of interaction treatments in distance education. *Review of Educational Research*, 79, 1243–1289. doi: 10.3102/0034654309333844
- Bernard, R. M., Abrami, P. C., Lou, Y., Borokhovski, E., Wade, A., Wozney, L., et al. (2004). How does distance education compare with classroom instruction? A meta-analysis of the empirical literature. *Review of Educational Research*, 74(3), 379–439.
- Bork, A. (1987). *Learning with personal computers*. New York: Harper & Row.
- Clark, R. (1978). Media, mental imagery, and memory. *Educational Technology Research and Development*, 26(4), 355–363. doi:10.1007/bf02766371.
- *Clark, R. E. (1983). Reconsidering the research on media. *Review of Educational Research*, 53(4), 445–459.
- *Clark, R. E. (Ed.). (2001). *Learning from instructional media: Arguments, analysis, and evidence*. Greenwich, CT: Information Age Publishers, Inc.
- Craig, J., & Bevington, W. (2006). *Designing with type: A basic course in typography* (4th ed.). New York: Watson Guptill.
- Dair, C. (1967). *Design with type*. Toronto: University of Toronto Press.
- Grabinger, R. S. (1983). *CRT text design: Psychological attributes underlying the evaluation of models of CRT text displays*. Unpublished Doctoral Dissertation, Indiana University, Bloomington, IN.
- Hannafin, M., Phillips, T., Rieber, L., & Garhart, C. (1987). The effects of orienting activities and cognitive processing time on factual and inferential learning. *Educational Technology Research and Development*, 35(2), 75–84. doi:10.1007/bf02769433.
- Ho, H.-K. (1984). Two experiments on the effects of mnemonic strategies: Is it mode or cognitive function that influences learning? *Educational Technology Research and Development*, 32(2), 89–100. doi:10.1007/bf02766668.
- Hooper, S., & Hannafin, M. (1986). Variables affecting the legibility of computer generated text. *Journal of Instructional Development*, 9, 22–29.
- *Hsieh, P., Acee, T., Chung, W., Hsieh, Y., Kim, H., Thomas, G., et al. (2005). Is educational intervention research on the decline? *Journal of Educational Psychology*, 97(4), 523–529.
- Januszewski, A., & Molenda, M. (2007). *Educational technology: A definition with commentary*. Retrieved from <http://www.aect.org>
- Kopcha, T., & Sullivan, H. (2008). Learner preferences and prior knowledge in learner-controlled computer-based instruction. *Educational Technology Research and Development*, 56(3), 265–286. doi:10.1007/s11423-007-9058-1.
- *Kozma, R. B. (1991). Learning with media. *Review of Educational Research*, 61(2), 179–211.
- *Kozma, R. B. (1994). Will media influence learning? Reframing the debate. *Educational Technology Research and Development*, 42(2), 7–19. doi: 10.1007/BF02299087
- *Levin, J. R. (2004). Random thoughts on the (in)credibility of educational-psychological intervention. *Educational Psychologist*, 39(3), 173–184.
- McManis, D. L. (1965). Position-cues in serial learning. *The American Journal of Psychology*, 78(4), 668–671.
- *Morrison, G. R., Ross, S. M., Gopalakrishnan, M., & Casey, J. (1995). The effects of feedback and incentives on achievement in computer-based instruction. *Contemporary Educational Psychology*, 20(1), 32–50. doi: 10.1006/ceps.1995.1002
- Morrison, G. R., Ross, S. M., Schultz, C. W., & O'Dell, J. K. (1989). Learner preferences for varying screen densities using realistic stimulus materials with single and multiple designs. *Educational Technology Research and Development*, 37, 53–60. doi:10.1007/BF02299056.
- Noble, C. E. (1952). Analysis of meaning. *Psychological Review*, 59, 421–430.
- Paivio, A. (1971). *Imagery and verbal processes*. New York: Holt, Rinehart, & Winston.
- Roschelle, J., Rafanan, K., Bhanot, R., Estrella, G., Penuel, B., Nussbaum, M., et al. (2009). Scaffolding group explanation and feedback with handheld technology: Impact on students' mathematics learning. *Educational Technology Research and Development*, 58(4), 399–419. doi:10.1007/s11423-009-9142-9.
- Ross, S., & Anand, P. (1987). A computer-based strategy for personalizing verbal problems in teaching mathematics. *Educational Technology Research and Development*, 35(3), 151–162. doi:10.1007/bf02793843.
- *Ross, S. M., & Morrison, G. R. (1989). In search of a happy medium in instructional technology research: Issues concerning external validity, media replications, and learner control. *Educational Technology Research and Development*, 37(1), 19–33. doi: 10.1007/BF02299043
- Ross, S. M., & Morrison, G. R. (1996). Experimental research methods. In D. Jonassen (Ed.), *Handbook of research on educational communications and technology* (pp. 1148–1170). New York: Simon & Schuster Macmillan.

- *Ross, S. M., & Morrison, G. R. (2004). Experimental research methods. In D. J. Jonassen (Ed.), *Handbook of research on educational communications and technology* (2nd ed., pp. 1021–1043). Mahwah, NJ: Lawrence Erlbaum Associates.
- Ross, S. M., Morrison, G. R., Hannafin, R. D., Young, M., van den Akker, J., Kuiper, W., et al. (2008). Research designs. In J. M. Spector, M. D. Merrill, J. V. Merriënboer, & M. P. Driscoll (Eds.), *Handbook of research on educational communications and technology* (3rd ed., pp. 715–761). New York: Taylor Francis.
- Ross, S. M., Morrison, G. R., & Lowther, D. (2010). Educational technology research past and present: Balancing rigor and relevance to impact school learning. *Contemporary Educational Technology, 1*(1), 17–35.
- Ross, S. M., Morrison, G. R., & Odell, J. K. (1988). Obtaining more out of less text in CBI: Effects of varied text density levels as a function of learner characteristics and control strategy. *Education Technology Research and Development, 36*(3), 131–142. doi:10.1007/BF02765460.
- Ross, S. M., Morrison, G. R., & Schultz, C. W. (1995). Preferences for different CBI text screen designs based on the density level and realism of the lesson content viewed. *Computers in Human Behavior, 10*(4), 593–603. doi:10.1016/0747-5632(94)90049-3.
- Slavin, R. (2008). What works? Issues in synthesizing education program evaluations. *Educational Researcher, 37*(1), 5–14. doi:10.3102/0013189X08314117.
- Snowberg, R. (1973). Bases for the selection of background colors for transparencies. *Educational Technology Research and Development, 21*(2), 191–207. doi:10.1007/BF02768947.
- Song, H.-D., & Grabowski, B. (2006). Stimulating intrinsic motivation for problem solving using goal-oriented contexts and peer group composition. *Educational Technology Research and Development, 54*(5), 445–466. doi:10.1007/s11423-006-0128-6.
- Tessmer, M., & Driscoll, M. (1986). Effects of a diagrammatic display of coordinate concept definitions on concept classification performance. *Educational Technology Research and Development, 34*(4), 195–205. doi:10.1007/bf02767401.
- Ullmer, E. J. (1994). Media and learning: Are there two kinds of truth? *Educational Technology Research and Development, 42*(1), 21–32. doi:10.1007/BF02298168.
- Underwood, E. J. (1996). *Experimental psychology*. New York: Appleton-Century-Crofts.
- Winn, W., Li, T.-Z., & Schill, D. (1991). Diagrams as aids to problem solving: Their role in facilitating search and computation. *Educational Technology Research and Development, 39*(1), 17–29. doi:10.1007/bf02298104.
- Winn, W., & Solomon, C. (1993). The effect of the spatial arrangement of simple diagrams on the interpretation of english and nonsense sentences. *Educational Technology Research and Development, 41*(1), 29–41. doi:10.1007/bf02297090.