Journal of Teacher Education
61(4) 302–312

DOI: 10.1177/0022487110369556
http://jte.sagepub.com
SAGE

Effects of a Computerized Professional Development Program on Teacher and Student Outcomes

Joseph B. Fisher[1], Jean B. Schumaker[2], John Culbertson[3], and Donald D. Deshler[2]

Abstract

Effects of a computerized professional development (PD) program for a concept teaching routine were investigated in two studies. For each, teachers were randomly assigned to either a virtual workshop group that used a multimedia software program for PD or an actual workshop group that participated in a live PD session. In Study 1, the teachers' knowledge about the routine and planning for the routine significantly improved after completing either workshop; no significant differences were found between the groups. Both teacher groups were satisfied with the PD. In Study 2, the teachers' performance of the routine in their classrooms improved, as did student performance on tests of concept knowledge. Students were satisfied with the instruction provided by both groups of teachers. No differences were found between the posttest scores earned by the teacher groups or by students of the teachers. Implications regarding computerized PD for teachers are discussed.

Keywords

instructional practices, professional development, teacher education and development, technology

The availability and use of computerized programs for the professional development of teachers are rapidly expanding (Appana, 2008; Laferriere, Lamon, & Chan, 2006). Departments of education, school districts, universities, foundations, professional organizations, and even broadcasters have all produced computerized programs in various forms (e.g., podcasts, webinars, online courses, and multimedia software programs) for teachers (for reviews, see Ginsburg, Gray, & Levin, 2004; Kleiman, 2004). Many organizations have embraced computerized programs because they make professional development accessible to teachers (Kleiman, 2004; Walker, Downey, & Sorensen, 2008; Wells, Lewis, & Greene, 2006) and affordable for schools (Abbott, Greenwood, Buzhardt, & Tapia, 2006; Wentling et al., 2000), factors that have proven to be barriers to professional development in the past (Archibald & Gallagher, 2002; Elges, Righettini, & Combs, 2006).

Nevertheless, although professional development should be accessible and affordable, more importantly, it must be effective. That is, it must improve teacher classroom practice and, by extension, improve student outcomes (Gersten & Dimino, 2001; Snow-Renner & Lauer, 2005). Unfortunately, little is known about the effectiveness of computerized professional development programs (i.e., both online programs and multimedia software programs) in relation to the improvement of teacher classroom practice (Dede, Ketelhut, Whitehouse, Breit, & McCloskey, 2009; Ginsburg et al., 2004; Harlen & Doubler, 2007; Laferriere et al., 2006; Liaupsin, 2003).

This fact becomes especially clear when the research literature on such programs for teacher professional development is examined using an evaluation model outlined by Kirkpatrick (2006) that focuses on four levels of outcomes: reaction, learning, behavior, and results. Level 1, reaction, simply refers to participant satisfaction with a professional development program. Level 2, learning, refers to the knowledge and skills participants gain as a result of participation in a professional development program. Level 3, behavior, refers to change in participant actions as a result of a professional development program. In education, this means change in how instruction is provided by a teacher in the classroom. Level 4, results, refers to change in others' behavior as a result of change in participant behavior at Level 3. In education, this means improvement in student learning as a result

[1]Grand Valley State University, Grand Rapids, MI, USA
[2]University of Kansas, Lawrence, KS, USA
[3]American Society of Radiologic Technologists, Albuquerque, NM, USA

Corresponding Author:
Joseph B. Fisher, Grand Valley State University, College of Education, 301 W. Fulton St. Suite 920, Grand Rapids, MI 49504
Email: fisherj@gvsu.edu

of change in how instruction is provided by a teacher who participated in a professional development program.

To date, most of the published studies in this area have reported the effects of computerized professional development for teachers relative to the first two levels of Kirkpatrick's model: reaction and learning. Generally, these studies have compared the effects of computerized professional development (i.e., online programs and multimedia software programs) to the effects of live, face-to-face professional development. These studies have reported that teachers have expressed positive reactions to the instruction they received through computerized professional development (e.g., De La Paz, Hernandez-Ramos, & Barron, 2004 [multimedia software]; Fisher, Deshler, & Schumaker, 1999 [multimedia software]; Warren & Holloman, 2005 [online]), that teachers in both computerized professional development and face-to-face professional development were equally satisfied with the instruction received (e.g., Fisher et al., 1999; Warren & Holloman, 2005), that teachers have reported that their understanding of classroom issues and instructional practices improved as a result of computerized professional development (e.g., Fisher et al., 1999; Walker et al., 2008 [multimedia software]), and that the ability of teachers to prepare items such as lesson and unit plans, PowerPoint presentations, and professional mission statements was equivalent following both computerized professional development and face-to-face professional development (e.g., Fisher et al., 1999; Peterson & Bond, 2004 [online]; Warren & Holloman, 2005).

Few studies have been found that examine computerized professional development for teachers relative to the last two levels of Kirkpatrick's model: behavior and results. For example, Whitehouse, Breit, McCloskey, Ketelhut, and Dede (2006) reported, in a review of 400 articles about teacher professional development published after 2000, that 40 met their criteria for high-quality empirical research and only 2 reported on the effects of computerized professional development regarding teacher behavior and student results (Harris & Grandgenett, 2002; Leach et al., 2004). In these studies, results were based on self-report measures of teachers (i.e., surveys, questionnaires, and interviews). According to Desimone (2009), self-report measures can provide valuable data about the effects of professional development on teachers' classroom practice and students' learning and are commonly used by researchers because they are cost-effective to employ. Likewise, Dede et al. (2009) also indicated that self-report measures offer valuable insight into teachers' perspectives of their own teaching; however, these same researchers argued that too much of the research examining the effects of professional development on teacher classroom behavior and student learning is based on self-report measures, and this overemphasis needs to be balanced with studies using measures of change that are more objective.

One study that was conducted to provide this kind of balance (Fisher et al., 1999) employed classroom observation to measure directly the effects of computerized professional development on teacher behavior. Teachers who were randomly assigned to the computerized professional development group received all of their instruction from a multimedia software program stored on a compact disc. In contrast, teachers who were randomly assigned to a live, face-to-face group received all of their instruction in small groups from an experienced and knowledgeable professional developer. Regarding teacher behavior, results indicated teachers in both groups correctly performed a greater number of the practice's targeted behaviors in their classrooms after participation than before, and teachers in both groups made similar gains.

To summarize, studies have shown that computerized professional development programs appear to be acceptable to teachers (reaction) and to produce significant gains in teacher knowledge on paper-and-pencil tests (learning). Though the number of these studies is few, even fewer have directly measured the actual classroom instruction (behavior) of teachers, and no studies have been found that have directly measured student learning (results). Clearly, studies are needed that show that actual teacher behavior and student learning improve along with teacher learning of information and satisfaction with the professional development program.

The purpose of the two studies reported here was to strengthen and extend the existing literature base on computerized programs for teacher professional development. The purpose of Study 1 was to test the effects of a multimedia software program relative to Kirkpatrick's first two levels: reaction and learning. Specifically, Study 1 addressed the research question: How do the effects of a virtual workshop (a multimedia software program) and an actual workshop (a live, face-to-face program) compare with regard to teacher knowledge of the intervention (learning), teacher skill in preparing to use the intervention (learning), and teacher satisfaction (teacher reaction) with the professional development received. The purpose of Study 2 was to measure the effects of the same multimedia software program with regard to Kirkpatrick's last two levels: behavior and results. In addition, student satisfaction with their teachers' use of the instructional practice was also measured as an additional level of evaluation. Specifically, Study 2 addressed the research question, how do the effects of a virtual workshop and an actual workshop compare with regard to teacher implementation of an instructional practice in the classroom (behavior), student learning (results), and student satisfaction with the instruction they received (student reaction).

Study 1

Method

Participants. A total of 59 certified teachers who were enrolled in a graduate-level course on increasing access to the general education curriculum for students with disabilities volunteered to participate in the study. They were randomly selected into the experimental or control group. Of the 30 teachers in the experimental group, 19 were female and 11 were male. These teachers' ages ranged from 23 to 53 years (M = 32.72 years), and they had an average of 6.53 years of teaching experience**.** Of the 29 teachers in the control group, 21 were female and 8 were male. Their ages ranged from 24 to 56 years (M = 37.06 years), and they had an average of 8.89 years of teaching experience.

Settings

Virtual Workshop. The virtual workshop (VW; the workshop involving a multimedia software program) took place in a computer lab at a Midwestern high school. The classroom had 25 Macintosh computers on tabletops arranged in rows. The lab was also outfitted with an instructor's computer, a data projector, a screen, and white boards.

Actual Workshop. The actual workshop (AW; the workshop involving a live, face-to-face program) took place in a classroom at a Midwestern high school. The classroom was outfitted with tables and 25 to 30 chairs arranged in rows. The classroom was also outfitted with an instructor's computer, a data projector, a screen, and white boards.

The Instructional Practice. The instructional practice about which the teachers received professional development was the Concept Mastery Routine (CMR; Bulgren, Schumaker, & Deshler, 1993). This routine is a research-validated inclusive practice (Bulgren, Schumaker, & Deshler, 1988) designed to help teachers teach students how to process and understand information related to key concepts (e.g., colonialism, socialism, poetry, phylum) in subject-area courses. It not only allows teachers to focus on subject-area content but also allows them to focus on how students learn that content. The routine includes a series of procedures teachers use to co-construct knowledge with students about a single concept. During the co-construction process, the teacher and students record information about the concept on a graphic device called a Concept Diagram. The instructional routine is based on concept-acquisition theory and empirically validated principles of effective concept instruction as derived through the literature. In addition, it combines the use of advance organizers, graphic organizers, and interactive discussion into one simple yet powerful routine for teaching complex concepts.

The Professional Development Programs

Virtual Workshop. The VW for the CMR was a multimedia software program (Fisher & Schumaker, 2008) created using Macromedia Authorware. For this study, the software was distributed to participants on two compact discs. Through the coordinated use of text, video, audio, and animated graphics, Disc 1 explicitly instructed users about the routine's purpose, instructional sequence, and the Concept Diagram. Moreover, it prompted users through the process of preparing an initial draft of a Concept Diagram. Disc 2 was a classroom simulator. Through the coordinated use of multimedia, it guided teachers in the application of the routine. Specifically, it allowed the user to access a lesson plan, interact with virtual students, receive support from a virtual coach, and record information on a virtual Concept Diagram. The VW was designed to be a stand-alone package to reduce the costs associated with professional development and increase teacher access to professional development.[1]

Actual Workshop. The AW for the CMR was composed of face-to-face instruction that was divided into two parts, which corresponded to the parts of the VW described above. In Part 1, participants were explicitly instructed about the routine and on how to prepare drafts of Concept Diagrams for classroom application. To ensure content consistency between the workshops, the text and graphics from Disc 1 of the VW were translated into PowerPoint slides for Part 1 of the AW, and the video clips from Disc 1 of the VW were placed on a DVD and viewed in Part 1 of the AW.

In Part 2 of the AW, users had an opportunity to apply the routine in a simulated lesson. That is, participants taught a practice lesson. They were provided a lesson plan, a blank Concept Diagram, students with whom to practice, and a coach to prompt their application of the routine as needed.

Overall, the difference between the AW and the VW was that the AW involved face-to-face professional development and the VW involved computerized professional development. Both workshops integrated the same content and known principles of effective teacher development including a focus on content instruction that responds to how students learn, models of a specific instructional practice by expert teachers, opportunities for active learning, guidance during preparation, and coaching during application (e.g., Boudah, Logan, & Greenwood, 2001; Desimone, 2009; Knight, 2004, 2007; Snow-Renner & Lauer, 2005).

Measurement Instruments and Measures

Teacher knowledge test. This test was composed of seven short-answer questions developed to measure a teacher's recall and understanding of the CMR's components and procedures. For example, one question asked, "What are the three ways that characteristics of a concept are classified in the Concept Mastery Routine?" Teachers were provided 30 minutes to answer the questions on this paper-and-pencil test both before and after instruction.

To score teacher answers, evaluation guidelines specifying acceptable responses for each question were created. Different point values were awarded according to the number of items specified for each answer. Teachers could earn a maximum score of 38 points on the test. The percentage of points earned was calculated for each teacher by dividing the number of points earned by 38 and multiplying by 100. This percentage was called the teacher knowledge score.

Concept Diagram test. For this test, teachers completed a blank Concept Diagram for a common concept that was familiar to all of them ("automobile"). This test measured the teachers' knowledge of what type of information (e.g., characteristic, example, etc.) belongs in each section of the Concept Diagram as well as their ability to create and place that information. Teachers were provided 10 min to fill in the 26 blanks on the diagram both before and after they had received training.

To score participants' completed Concept Diagrams, evaluation guidelines specifying acceptable responses for each blank were created. For 21 of the Concept Diagram's 26 blanks, participants earned 5 points for an acceptable response. For the remaining blanks, participants received 1 point for each acceptable response. Overall, each teacher could earn a maximum score of 110 points. The percentage of points earned was calculated for each teacher by dividing the number of points earned by 110 and multiplying by 100. This percentage score was called the Concept Diagram score.

Teacher satisfaction questionnaire. This questionnaire assessed the teachers' opinions about the training they received. Each of the 14 items included a 7-point Likert-type scale ranging from *strongly disagree* (1) to *strongly agree* (7). The questionnaire items were designed to determine (a) how enjoyable teachers found the training, (b) how engaged teachers felt during the training, (c) how understandable teachers found the content, and (d) how applicable teachers found the content. Teacher ratings were averaged for each item on the questionnaire for each group of teachers. The teacher ratings on these items were called the teacher satisfaction ratings.

Interscorer reliability. Interscorer reliability was determined by having two scorers independently score 20% of the teacher knowledge tests and Concept Diagram tests. For each measure, points awarded by the scorers were compared item by item. The percentage of agreement was calculated by dividing the number of agreements by the number of disagreements and multiplying by 100. For the teacher knowledge tests, the scorers agreed 407 times out of 456 opportunities to agree (total percentage of agreement = 89.25%). For the Concept Diagram tests, the scorers agreed 1,089 times out of 1,320 opportunities to agree (total percentage of agreement = 82.50%).

Procedures

Common procedures. The teachers participated in their respective assigned workshop. The same session leader (the first author) was present in each workshop. Immediately before each workshop, the session leader administered the Teacher Knowledge and Concept Diagram tests. If participants asked about a question's answer, they were instructed to answer the question as best they could. Once the teachers' respective workshop was complete, they were given the Teacher Knowledge and Concept Diagram tests. Also, they completed the teacher satisfaction questionnaire.

VW procedures. Once teachers in the VW group completed these pretests, the session leader used the instructor station computer, data projector, and screen to provide a 5-min demonstration of how to navigate the VW program's pages. Following this demonstration, teachers turned on their computers on which Disc 1 of the VW had been loaded and began to navigate the program. Once teachers completed Disc 1, the session leader loaded the second compact disc into the computer and the teachers continued to work through the VW. Teachers were provided a maximum of 3 hr to complete the VW. The session leader remained in the lab to provide technical support to ensure the teachers completed the intervention. If teachers had questions about the content, they were told the VW contained all the information they needed to understand the CMR. (Please note: The VW is designed such that a session leader is not needed, and teachers can complete the program independently. The session leader was present during this study only to ensure that the teachers fully completed the program and did not run into technical difficulties with their computers.)

AW procedures. Once teachers in the AW group completed the pretests, the session leader began the workshop. Using the instructor station computer, data projector, and screen located in the classroom, the session leader directed the AW. To provide content consistency, the session leader presented the text and graphics taken directly from the VW. Moreover, the session leader played all the same video clips used in the VW at the appropriate times in the presentation. The teachers completed the same practice activities as teachers in the VW; however, the session leader, not the computer program, provided corrective feedback. Like teachers in the VW, teachers in the AW were also provided an opportunity to practice implementing the routine; however, unlike the VW, the session leader prompted teacher implementation, not the computer program. The AW lasted for 3 hr. Any questions asked by teachers about the CMR were answered by the session leader.

Experimental Designs and Data Analysis. Two experimental designs were used. A pretest–posttest control-group design (Campbell & Stanley, 1963) was used to compare the knowledge scores and Concept Diagram scores of teachers who participated in the VW and the AW. The mean scores and standard deviations were calculated for each measure for the pretest and posttest for both treatment groups. To compare the differences between the pretest and posttest scores within each treatment group, t tests were performed. To determine whether the two training methods had differential effects on the teachers' performance on each test, analyses of covariance (ANCOVAs) were employed using the teachers' posttest scores as the dependent variable and their pretest scores as the covariate.

Second, a posttest-only control-group design (Campbell & Stanley, 1963) was used to compare the satisfaction scores of teachers participating in the VW and the AW. The mean ratings were calculated for each item as well as an overall mean rating for each group of teachers. To illuminate differences between the satisfaction scores of each treatment group for each item, an analysis of variance (ANOVA) was performed for each questionnaire item. For each test, the criterion for significance was set at .05.

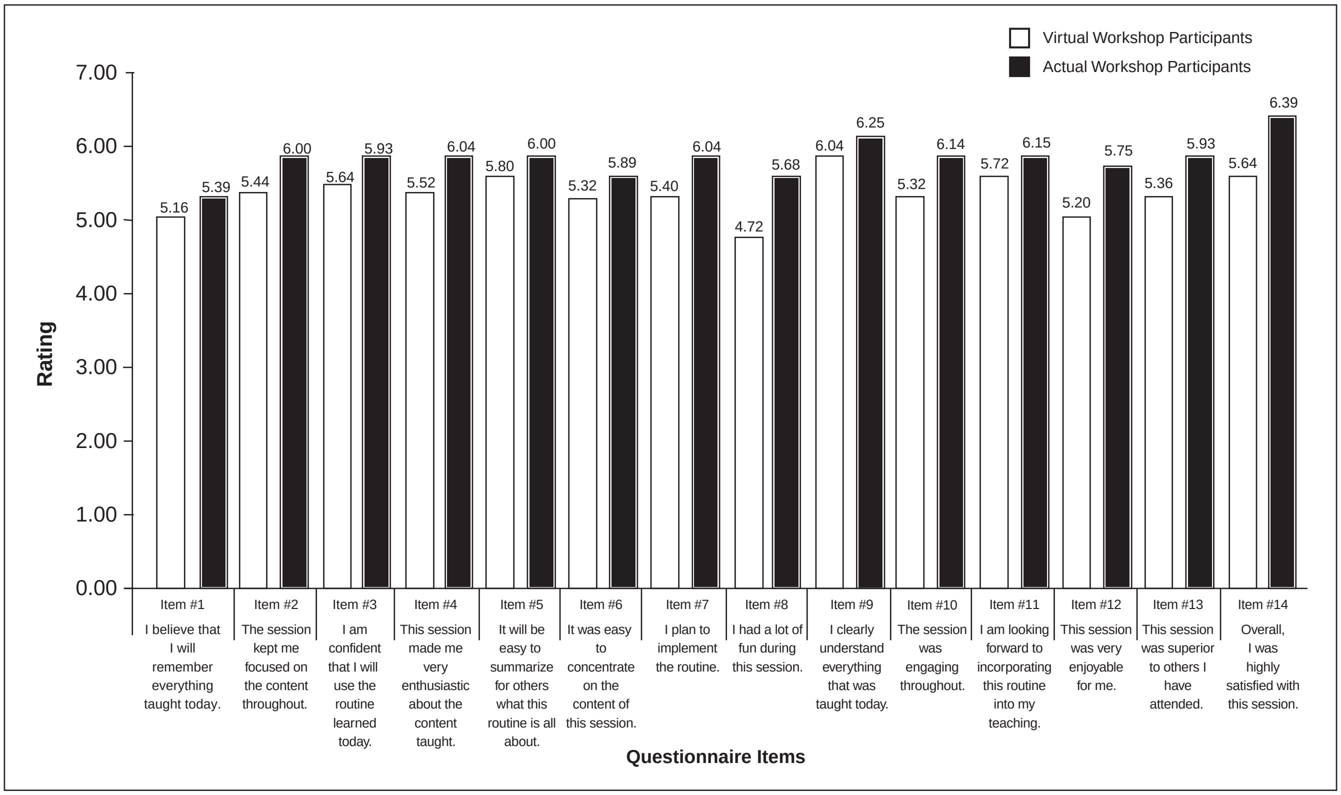

Figure 1. Teacher satisfaction questionnaire results

Results

Knowledge Results. Knowledge scores earned by the AW teachers ranged from 0% to 8% before training ($M = 0.93\%$, $SD = 2.03$) and from 47% to 82% after training ($M = 65.72\%$, $SD = 11.77$). Likewise, knowledge scores earned by VW participants ranged from 0% to 11% before training ($M = 0.53\%$, $SD = 2.17$) and from 47% to 92% after training ($M = 68.26\%$, $SD = 10.35$). t tests indicated that the posttest scores of VW teachers were significantly different than their pretest scores, $t(29) = -35.15$, $p < .00$, as were those of the AW teachers, $t(28) = -29.67$, $p < .00$. An ANCOVA revealed no significant differences between the posttest scores of AW and VW participants, $F(1, 58) = 1.28$, $p = .263$.

Concept Diagram Results. Concept Diagram scores for the AW Group ranged from 0% to 32% on the pretest ($M = 6.13\%$, $SD = 7.46$) and from 73% to 100% on the posttest ($M = 87.51\%$, $SD = 8.98$). Concept Diagram scores for the VW Group ranged from 0% to 25% on the pretest ($M = 4.50\%$, $SD = 5.46$) and from 45% to 100% on the posttest ($M = 85.80\%$, $SD = 10.86$). t tests indicated that the posttest scores of VW teachers were significantly different than their pretest scores, $t(28) = -29.67$, $p < .00$, as were the pretest and posttest scores of AW teachers, $t(28) = -35.39$, $p < .00$. An ANCOVA revealed no significant differences between the posttest scores of AW and VW participants, $F(1, 58) = 0.60$, $p = .440$.

Teacher Satisfaction Results. Across all 14 questionnaire items, the mean satisfaction scores of the AW teachers ranged from 5.39 to 6.39 (overall $M = 5.97$, $SD = 1.09$). Similarly, mean satisfaction scores of the VW teachers ranged from 5.22 to 6.15 (overall $M = 5.44$, $SD = 1.30$). Figure 1 displays the mean rating for each item for each group.

Although the AW mean rating was higher than the VW mean rating for all 14 items, ANOVAs revealed significant differences between the satisfaction scores of the AW and VW participants on the following four items: Item 7, $F(1, 51) = 4.18$, $p = .05$, which inquired about participant willingness to implement the routine, Item 8, $F(1, 52) = 6.94$, $p = .01$, which inquired about how enjoyable participants found the professional development, Item 10, $F(1, 52) = 5.37$, $p = .02$], which inquired about how engaged participants felt during the professional development, and Item 14, $F(1, 52) = 6.12$, $p = .02$, which inquired about participants overall satisfaction with the professional development received.

Study 2

Method

Participants

Teachers. Eight teachers volunteered to participate in this study. Four teachers were randomly selected to serve in the experimental (VW) group; the four remaining teachers served in the control (AW) group. In the experimental group, all four teachers were female. Three were seventh-grade teachers, and one was a sixth-grade teacher. These teachers' ages ranged from 23 to 53 years ($M = 34.64$), and they averaged 10.75 years of teaching experience, with a range of 1 to 32 years. In the control group, three teachers were female and one was male. Three were seventh-grade teachers, and one was an eighth-grade teacher. These teachers' ages ranged from 45 to 55 years ($M = 47.73$), and they averaged 22.25 years of teaching experience, with a range of 12 to 27 years.

Students. A total of 125 students with permission from their parents or guardians participated in the study. All of these students were enrolled in one of the eight teachers' classes at three middle schools in a large Midwestern city. Five classes were observed in School A; four of these classes participated in the experimental group, and one participated in the control group. School A had a total enrollment of 515 students, including 173 sixth-grade students, 168 seventh-grade students, and 174 eighth-grade students. Of School A's student population, 63% was White and 37% was non-White, and 48% qualified for free or reduced-price lunches. Two classes were observed in School B, and both of these classes participated in the control group. School B had a total enrollment of 410 students, including 144 sixth-grade, 121 seventh-grade, and 145 eighth-grade students. Of School B's student population, 39% was White and 61% was non-White, and 31% qualified for free or reduced-price lunches. One class was observed in School C, and this class participated in the control group. School C had a total enrollment of 446 students, including 158 sixth-grade students, 148 seventh-grade students, and 140 eighth-grade students. Of School C's student population, 48% was White and 52% was non-White, and 31% qualified for free or reduced-price lunches.

The age of the 76 students whose teachers were in the experimental (VW) group ranged from 11.0 to 13.4 years ($M = 12.0$). There were 32 males and 44 females, with 43% of the students representing minority populations. The age of the 49 students whose teachers were in the control (AW) group ranged from 12.1 to 13.5 years ($M = 12.7$). There were 36 males and 13 females, with 61% of the students representing minority populations.

Settings. All of the teacher instruction took place in a room equipped with chairs and tables. The student instruction took place in the teachers' and students' regularly assigned classrooms.

Measurement Instruments and Measures

Implementation checklist. This observational checklist was used in Study 2 to assess teacher instruction during a classroom lesson about a concept. The teachers were asked to identify which class periods they would be teaching students about a concept and to name the concept. The observers used the checklist during those identified class periods. The checklist was composed of three sections that corresponded to parts of the CMR's instructional sequence.

A total of 39 teacher behaviors (e.g., teacher cueing that a concept was going to be taught, teacher naming of the concept) were assessed. If a behavior was performed, the teacher could earn 1 to 5 points for the behavior, depending on the complexity of the behavior. If a behavior was not performed, the teacher received 0 points for that behavior. Overall, each time a concept was taught, a teacher could earn a maximum score of 165 points. Once a teacher had been observed, the points for that class period were totaled, divided by 165, and multiplied by 100. These percentage scores were called the implementation scores.

Student concept acquisition test. This test assessed students' understanding of a concept that the students' teacher had identified as being taught in a given class period. This 22-point, short-answer test was designed to test students' knowledge of a concept's definition, characteristics, and examples. Also, it was designed to test students' ability to analyze whether a new item was an example or a nonexample of the concept. Items on the test were adapted to the particular concept that had been taught. The test was administered to the students after their teachers had taught them about a specified concept: once before their teachers had received training and once after they received training. Student responses were scored using written evaluation guidelines specifying acceptable responses for each question. The number of correct items was divided by 22 and multiplied by 100. This percentage score was called the student concept acquisition score.

Student satisfaction questionnaire. This questionnaire was used in Study 2 to measure students' satisfaction with the teachers' use of the CMR. There were 11 items on the questionnaire, and each item included a 7-point Likert-type scale ranging from *extremely satisfied* (7) to *extremely dissatisfied* (1). The questionnaire items were designed to determine, for example, (a) how satisfied the students were that they understood the Concept Diagram, (b) how satisfied the students were that they could use the Concept Diagram to study for tests, (c) how satisfied the students were with participating in the creation of the Concept Diagram, and (d) how satisfied the students were that they understood the lesson being taught with the Concept Diagram. The students' ratings were averaged for each item for each group of students. The student ratings were called the student satisfaction scores.

Interscorer reliability. Interscorer reliability was determined by having two scorers independently score 20% of the student concept acquisition tests and by having two observers simultaneously record information in 20% of the classroom observations of teachers' implementation of the routine. For each measure, points awarded by the scorer–observers were compared item by item. The percentage of agreement was calculated by dividing the number of agreements by the number of agreements plus disagreements and multiplying by 100.

For the student concept acquisition tests, the scorers agreed 218 times out of 264 opportunities to agree (total percentage of agreement = 82.57%). For the implementation checklist, the scorers agreed 1,691 times out of 1,980 opportunities to agree (total percentage of agreement = 85.40%).

Procedures. The teachers in Study 2 completed their respective workshop following the same procedures as those described for Study 1 above. Both groups of teachers learned about the CMR through the AW or the VW, depending on their assigned group; however, before beginning their respective workshop, these teachers were observed delivering three or four lessons in their classrooms, depending on their assignment within the experimental design. During each of these lessons, each participating teacher delivered instruction on a concept of his or her choice within the subject area of the course that the teacher was teaching.

During each lesson, trained observers scored the middle school teachers' presentations using the implementation checklist. Once the baseline data were stable or showed decreasing trends for each of these teachers, they participated in either the VW or the AW, which were the same workshops as described for Study 1. After training had been completed, the teachers were observed presenting lessons in which they indicated a concept would be taught. During each lesson, observers again scored the teachers' presentations using the implementation checklist.

After each participating teacher's last baseline lesson, his or her class of students completed the concept acquisition test. Then, following each participating teacher's very last observed lesson, his or her class of students again completed the concept acquisition test.

Experimental Designs and Data Analysis. Three experimental designs were used. To determine the effects of the workshops on teacher implementation scores, a multiple-baseline across-teachers design (Baer, Wolf, & Risley, 1968) was employed. Eight teachers participated in this design, with two teachers participating in each replication of the design. The teachers were all observed teaching concepts to their students several times before and several times after training. Implementation scores were graphed for visual analysis.

A pretest–posttest control-group design (Campbell & Stanley, 1963) was also used to compare the concept acquisition scores of students in the two groups. To compare the differences between the pretest and posttest scores of students within each treatment group, *t* tests were performed. To illuminate differences between the concept acquisition scores of each group, an ANCOVA was performed with the pretest scores serving as the covariate and the posttest scores serving as the dependent variable.

A posttest-only control-group design (Campbell & Stanley, 1963) was used to compare the satisfaction scores of the student groups. The student ratings on the satisfaction questionnaire were handled descriptively, with a mean rating calculated for each item on the questionnaire as well as an overall mean rating for each student group. To illuminate differences between the satisfaction scores of each student group for each item, ANOVAs were performed. For each test, the criterion for significance was set at .05.

Results

Implementation Results. Figure 2 shows the performance of the eight teachers in their classrooms as reflected by their implementation scores. Each teacher's implementation scores are shown both before (baseline) and after the workshop (after training). Baseline scores earned by AW participants ranged from an average of 0.00% by Teacher 2 to 3.00% by Teacher 1. Overall, the mean baseline score earned by the AW group was 1.79% (SD = 0.97). Similarly, baseline scores earned by VW participants ranged from an average of 0.00% by Teacher 5 to 2.25% by Teachers 7 and 8. Overall, the mean baseline score earned by the VW group was 1.85% (SD = 1.28). After-training scores for AW participants ranged from an average of 69.33% by Teacher 2 to 84.00% by Teacher 4. Overall, the mean after-training score earned by the AW group was 75.20% (SD = 6.16). Of the 10 lessons observed after training for AW teachers, two scores exceeded the mastery level of 80%. After-training scores for VW participants ranged from an average of 85.33% by Teacher 6 to 95.00% by Teacher 7. Overall, the mean after-training score earned by the VW group was 88.51% (SD = 4.36). Of the nine lessons observed after training for the VW teachers, eight exceeded the mastery level of 80%.

Concept Acquisition Test Results. The concept acquisition test was administered to students over the concept taught during the last baseline lesson and over the concept taught during the last after-training lesson. Scores only for those students who took both the pretest and the posttest are included in the results. Scores earned by students taught by AW teachers ranged from 0% to 21% correct on the pretest (M = 10.25%, SD = 9.17) and from 44% to 88% correct on the posttest (M = 62.00%, SD = 16.81). Scores earned by students taught by VW teachers ranged from 3% to 31% (M = 14.75%, SD = 11.78) correct on the pretest and from 63% to 73% correct on the posttest (M = 67.75%, SD = 4.25). Dependent t tests indicated that the posttest aggregated mean scores of students whose teachers participated in the VW were significantly different than their pretest scores, $t(7) = 9.93$, $p = .002$, as were the means of students whose teachers participated in the AW, $t(7) = 8.35$, $p = .004$.

An ANCOVA was used to compare the aggregated classroom means of the student groups. Posttest scores served as the dependent variable and pretest scores as the covariate. These analyses revealed no significant differences between the posttest scores of students whose teachers participated in the VW and AW, $F(1, 7) = 0.30$, $p = .606$.

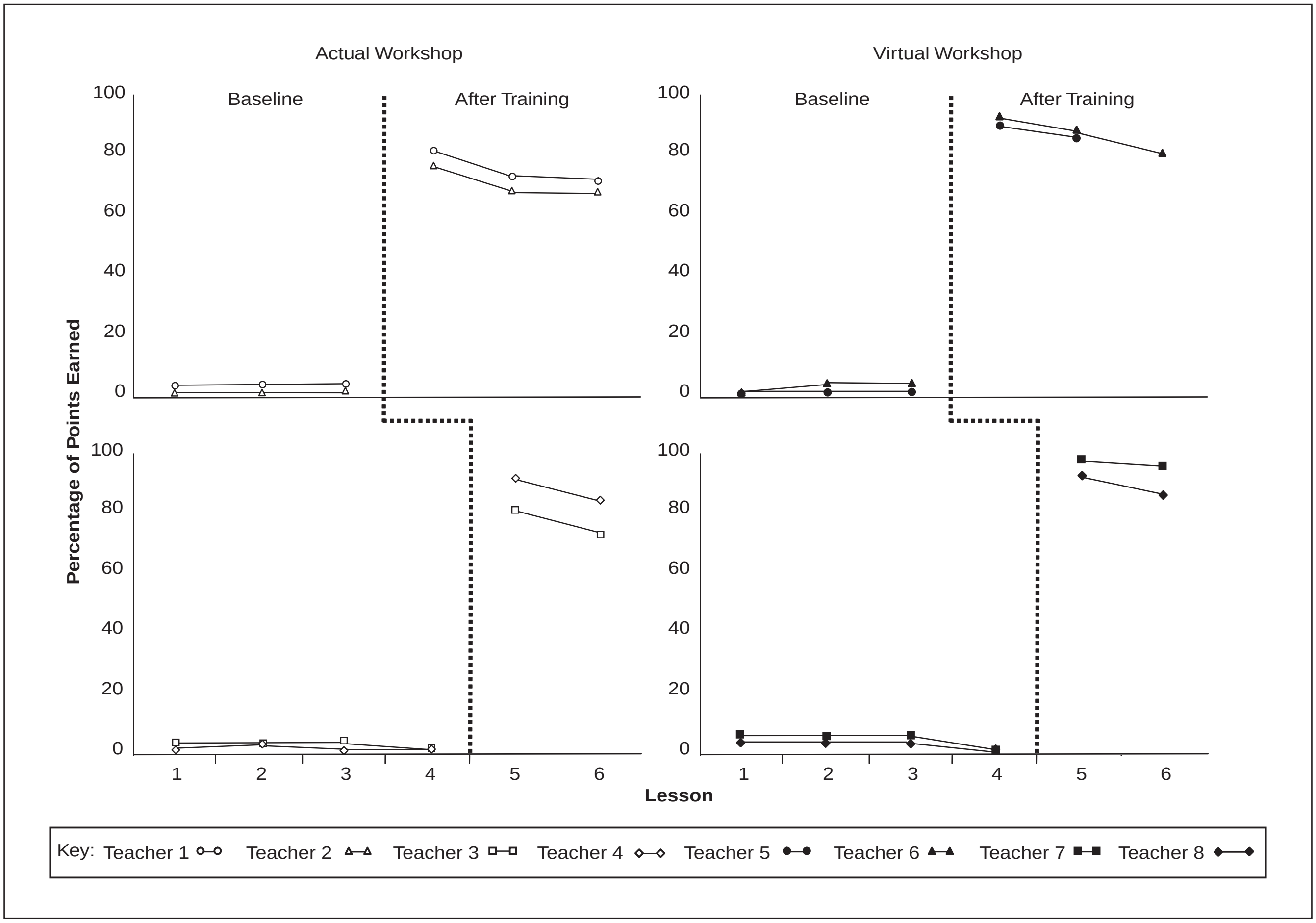

Figure 2. Implementation scores of Teachers 1–4 (actual workshop participants) and Teachers 5–8 (virtual workshop participants)

Student Satisfaction Results. Student satisfaction scores on individual items ranged from 5.59 to 6.18 (overall $M = 5.79$, $SD = 1.30$) for AW teachers' instruction and from 5.58 to 6.29 for VW teachers' instruction (overall $M = 5.86$, $SD = 1.18$). ANOVAs revealed no significant differences between the satisfaction scores of the student groups for each item.

Discussion

The purpose of this investigation was to measure the effects of a computerized professional development program in comparison to face-to-face instruction relative to all four levels of Kirkpatrick's (2006) evaluation model plus student satisfaction results. Study 1, which focused on Levels 1 and 2 (learning and reaction), showed that the teachers' scores on the knowledge and Concept Diagram tests significantly improved following participation in either the VW or AW. Moreover, both the VW and AW had similar effects on the scores teachers earned on the knowledge and the Concept Diagram tests, suggesting that the programs were equally effective. Results from the satisfaction questionnaire indicated that teachers rated both the VW and AW favorably; however, teachers who participated in the AW expressed somewhat higher satisfaction ratings than teachers who participated in the VW. In fact, on Item 7, which inquired about willingness to implement, Item 8, which inquired about enjoyment with the professional development, Item 10, which inquired about engagement during the professional development, and Item 14, which explicitly inquired about overall satisfaction with the professional development, participants in the AW rated their overall satisfaction significantly higher than participants in the VW. These findings suggest that making it more engaging and fun might improve the VW for participants, which may result in improved implementation and overall satisfaction ratings.

With regard to Study 2, the implementation results suggest that teachers in both groups performed a substantially greater number of the targeted instructional behaviors in their classrooms after participation than before, and their posttraining scores represented a high level of fidelity. Moreover, results indicate that teachers who completed the VW implemented the CMR in a manner similar to the teachers who completed the AW. With regard to the student results, their posttest scores on the concept acquisition test significantly improved following teacher participation in either the VW or AW. Also, the concept instruction provided by teachers who completed either the VW or AW had similar effects on the scores students earned on the concept acquisition test. Furthermore, the two groups of students were similarly satisfied with their teachers' use of the routine. Thus, the computerized professional development program used in this study was as effective as face-to-face professional development relative to Levels 3 and 4 of Kirkpatrick's model (behavior and results) plus a new factor: student satisfaction.

Regarding both teacher reaction and learning, the results of Study 1 support the results of previous studies on computerized professional development. Specifically regarding reaction, results from Study 1 correspond to those previously reported indicating that teachers express positive reactions to computerized professional development programs (e.g., De La Paz et al., 2004; Fisher et al., 1999; Warren & Holloman, 2005); however, unlike previous studies in which teachers were equally satisfied with the instruction they received (Fisher et al., 1999; Warren et al., 2005), in Study 1 teachers who participated in face-to-face professional development rated their satisfaction significantly higher on 4 of the 14 items than teachers who participated in computerized professional development. This finding is not inconsistent with other research findings. For example, Laiw (2008) reported that learners often desire contact with instructors and active discussion with other participants, and the absence of such social interaction in computerized programs such as the VW that do not involve human facilitation may negatively affect satisfaction ratings.

Regarding teacher learning, results from Study 1 correspond to those of previous studies showing that teacher knowledge of instructional methods (e.g., Fisher et al., 1999; Walker et al., 2008) and ability to prepare for instruction (e.g., Fisher et al., 1999; Peterson & Bond, 2004; Warren & Holloman, 2005) improved following computerized professional development. Also, results of Study 1 were like those from previous studies (Fisher et al., 1999; Peterson & Bond, 2004; Warren & Holloman, 2005) in that learning outcomes were similar for teachers who participated in face-to-face professional development and those who participated in computerized professional development.

Regarding teacher behavior, student learning, and student reaction, the results of Study 2 extend the research base on computerized professional development. Specifically, Study 2 demonstrated that computerized programs not only can change teacher behavior but also can change teacher behavior in a way that results in improved student outcomes plus student satisfaction. Previous research has rarely examined the impact of computerized professional development for teachers on student learning (Whitehouse et al., 2006), nor has it focused on student satisfaction. When studies have focused on student learning, questionnaires completed by students or their teachers (e.g., Harris & Grandgenett, 2002; Leach et al., 2004) have been the main method of gathering research data (for a review, see Shih, Feng, & Tsai, 2008). Using objective measures helps to bring needed balance to the research literature (Dede et al., 2009).

Nevertheless, the current studies are limited in several ways. First, the number of teachers in Study 2 was limited. This relatively small sample of middle school teachers may not be representative of teachers in general, which limits the generalization of the results. Second, this small sample size may have limited the power of the statistical tests to detect real differences between the groups, especially related to implementation. Although the VW group consistently earned implementation scores above the AW group scores, the differences were not significant. It may have also limited the possibility that the groups of students would be equivalent demographically. Also, in both studies, participating teachers were volunteers and may not be representative of the general population of teachers.

These limitations underscore the need for additional research regarding the use of computerized professional development for teachers. First, studies need to be conducted that focus on larger numbers of teachers and that include all four levels of Kirkpatrick's evaluation model gathered from teachers and their students over a longer period of time. In addition, future research might explore in more detail the impact of the VW and AW on implementation scores. As mentioned above, although there was not a statistically significant difference between the groups' implementation scores in Study 2, a visible pattern was apparent. This same pattern was observed in the Fisher et al. (1999) study. With a larger sample size, the power of a statistical test would increase and may detect any real difference. Also, future studies might explore the differential impact of teachers' implementation on different types of students' performance. Currently, nothing is known about how different types of students react to changes in teacher performance as a result of professional development. Finally, future research might explore critical features of professional development rather than its structure (e.g., face-to-face professional development or computerized professional development). Desimone (2009) indicated that critical features of professional development include content focus, active learning, coherence, duration, and collective participation and that these features explain change in teacher knowledge, skill, and classroom practice, not the structure of professional development.

This investigation and other future endeavors like it that document the effects of computerized professional development regarding Kirkpatrick's four dimensions of effectiveness have the potential to affect teacher education in meaningful ways at both the inservice and preservice levels.

Today, inservice teachers rarely participate in effective professional development because it requires expertise, time, funds, and coordination beyond the means of many schools and/or districts (e.g., Archibald & Gallagher, 2002; Elges et al., 2006). Computerized professional development programs, like the VW however, remove these access barriers. Through computerized professional development programs, teachers could easily access, on their own schedules, effective professional development produced by leading experts in the field (Liaupsin, 2003). Moreover, professional development could be provided to teachers at less cost than face-to-face approaches (Abbott et al., 2006). At the preservice level, computerized professional development programs could be used in place of instruction provided in courses on instructional practices. These programs could provide teacher candidates a degree of depth and breadth of instruction not possible in the limited time of most university courses. Face-to-face course time, in turn, could be used to provide teacher candidates a setting in which to practice and receive additional coaching on the instructional practice above and beyond that provided in the computerized program.

In summary, in light of Kirkpatrick's evaluation model, the current research studies demonstrate that the field has the capacity to engineer software programs that have the power to provide effective professional development to teachers. These two studies demonstrate that computerized professional development programs can be designed in ways that teachers gain a great deal of knowledge (i.e., learning) about an instructional practice and express high levels of satisfaction (i.e., reaction) with what they have learned and how they have learned it. More importantly, such programs have the power to change teacher classroom practice (i.e., behavior) in ways that significantly improve student learning (i.e., results) and that are acceptable to students. For the field, the potential implications of these findings are significant. That is, through the use of computerized programs, the field of education now has a medium though which to make effective professional development on innovative instructional practices available on a broad scale in a format that teachers can easily access.

Acknowledgements

We thank Dr. Janis Bulgren for her leadership role in the initial creation of the Concept Mastery Routine as well as the teachers, graduate students, and middle-school students who kindly volunteered to participate in the studies. We also thank administrators in School District #501 for their support of this project.

Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interests with respect to the authorship and/or publication of this article.

Funding

This research was supported by grant #R44 HD36173 awarded to Edge Enterprises, Inc. from the National Institute for Child Health and Human Development, Small Business Innovation Research Program. The authors received no financial support for the authorship of this article.

Note

1. The workshop could have been formatted as an online teacher professional development (oTPD) program (e.g., webinar or online course), and doing so was considered. However, oTPD programs involve online human facilitation, which reduces access and increases costs when compared to programs that require no human facilitation. That is, human facilitators must be paid, which increases the cost of PD. Also, human facilitators must have expertise in both the content to be taught and the technology used to teach it online. Because such individuals are rare, access to PD is reduced.

References

Abbott, M., Greenwood, C. R., Buzhardt, J., & Tapia, Y. (2006). Using technology-based teacher support tools to scale up the classwide peer-tutoring program. *Reading and Writing Quarterly*, *22*(1), 47-64.

Appana, S. (2008). A review of benefits and limitations of online learning in the context of the student, the instructor, and the tenured faculty. *International Journal on E-Learning*, *7*(1), 5-22.

Archibald, S., & Gallagher, H. A. (2002, May 31). A case study of professional development expenditures at a restructured high school. *Education Policy Analysis Archives*, *10*(29). Retrieved from http://epaa.asu.edu/epaa/v10n29.html

Baer, D. M., Wolf, M. M., & Risley, T. R. (1968). Some current dimensions of applied behavior analysis. *Journal of Applied Behavior Analysis*, *1*, 91-97.

Boudah, D. J., Logan, K. R., & Greenwood, C. R. (2001). The research to practice projects: Lessons learned about changing teacher practice. *Teacher Education and Special Education*, *24*(4), 290-303.

Bulgren, J. A., Schumaker, J. B., & Deshler, D. D. (1988). Effectiveness of a concept teaching routine in enhancing the performance of LD students in secondary-level mainstream classes. *Learning Disability Quarterly*, *11*(1), 3-17.

Bulgren, J. A., Schumaker, J. B., & Deshler, D. D. (1993). *The Concept Mastery Routine.* Lawrence, KS: Edge Enterprises.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research.* Dallas, TX: Houghton Mifflin.

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. M. (2009). A research agenda for online professional development. *Journal of Teacher Education*, *60*(1), 8-19.

De La Paz, S., Hernandez-Ramos, P. F., & Barron, L. (2004). Multimedia environments in mathematics teacher education: Preparing regular and special educators for inclusive classrooms. *Journal of Technology and Teacher Education*, *12*(4), 561-575.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181-199.

Elges, P., Righettini, M., & Combs, M. (2006). Professional development and recursive e-learning. *Computers in the Schools, 23*(1), 45-57.

Fisher, J. B., Deshler, D. D., & Schumaker, J. B. (1999). The effects of an interactive multimedia program on teachers' understanding and implementation of an inclusive practice. *Learning Disability Quarterly, 22*(2), 127-142.

Fisher, J. B., & Schumaker, J. B. (2008). *The professional development program for the Concept Mastery Routine.* Lawrence, KS: Edge Enterprises.

Gersten, R., & Dimino, J. (2001). Realities of translating research into classroom practice. *Learning Disabilities Research and Practice, 16*(2), 120-130.

Ginsburg, A., Gray, T., & Levin, D. (2004). *Online professional development for mathematics teachers: A strategic analysis.* Washington, DC: American Institutes for Research.

Harlen, W., & Doubler, S. (2007). Researching the impact of online professional development for teachers. In R. Andrews (Eds.), *The Sage handbook of e-learning research* (pp. 466-486). London, UK: Sage.

Harris, J., & Grandgenett, N. (2002). Teachers' authentic learning with e-learning. *Learning and Leading With Technology, 30*(3), 54-58.

Kirkpatrick, D. (2006). *Evaluating training programs: The four levels* (3rd ed.). San Francisco, CA: Berrett-Koehler.

Kleiman, G. M. (2004, July). *Meeting the need for high quality teachers: E-learning solutions.* Paper commissioned for the U.S. Department of Education National Technology Plan and the U.S. Department of Education Secretary's Leadership Summit on e-Learning, Orlando, FL.

Knight, J. (2004). Instructional coaches make progress through partnership. *Journal of Staff Development, 25*(2), 32-37.

Knight, J. (2007). *Instructional coaching: A partnership approach to improving instruction.* Thousand Oaks, CA: Corwin Press.

Laferriere, T., Lamon, M., & Chan, C. K. K. (2006). Emerging e-trends and models in teacher education and professional development. *Teaching Education, 17*(1), 75-90.

Laiw, S. S. (2008). Investigating students' perceived satisfaction, behavioral intention, and effectiveness of e-learning: A case study of the Blackboard system. *Computers and Education, 51*(1), 864-873.

Leach, J., Patel, R., Peters, A., Power, T., Ahmed, A., & Makalima, S. (2004). Deep impact: A study of the use of handheld computers for teacher professional development in primary schools in the global south. *European Journal of Teacher Education, 27*(1), 5-28.

Liaupsin, C. J. (2003). The comprehensive evaluation of professional development software: A critique of methodology. *Journal of Special Education Technology, 18*(1), 29-37.

Peterson, C. L., & Bond, N. (2004). Online compared to face-to-face teacher preparation for learning standards-based planning skills. *Journal of Research on Technology in Education, 36*(4), 345-360.

Shih, M., Feng, J., & Tsai, C. C. (2008). Research and trends in the field of e-learning from 2001-2005: A content analysis of cognitive studies in selected journals. *Computers and Education, 51*(2), 955-967.

Snow-Renner, R., & Lauer, P. A. (2005). *McREL insights: Professional development analysis*. Aurora, CO: Mid-continent Research for Education and Learning.

Walker, D. A., Downey, P. M., & Sorensen, C. K. (2008). E-learning modules for teacher development: Project REAL. *Techtrends, 52*(5), 59-62.

Warren, L. L., & Holloman, H. L., Jr. (2005). On-line instruction: Are the outcomes the same? *Journal of Instructional Psychology, 32*(2), 148-151.

Wells, J., Lewis, L., & Greene, B. (2006). *Internet access in U.S. public schools and classrooms: 1994-2005* (NCES 2007-020). Washington, DC: U.S. Department of Education, National Center for Education Statistics.

Wentling, T. L., Waight, C., Gallaher, J., La Fleur, J., Wang, C., & Kanfer, A. (2000). *E-learning: A review of literature.* Champaign: University of Illinois, Knowledge and Learning Systems Group.

Whitehouse, P. L., Breit, L. A., McCloskey, E. M., Ketelhut, D. J., & Dede, C. (2006). An overview of current findings from empirical research on online teacher professional development. In C. Dede (Ed.), *Online professional development for teachers: Emerging models and methods* (pp. 13-29). Cambridge, MA: Harvard Education Press.

About the Authors

Joseph B. Fisher, PhD, is associate professor in the Department of Special Education, Foundations, and Technology at Grand Valley State University. His research interests include the design and validation of inclusive instructional practices to improve academic outcomes in K-12 classrooms.

Jean B. Schumaker, PhD, is professor emeritus at the University of Kansas and president of Edge Enterprises, Inc., a research and publishing company. Her current interests focus on the instruction of writing strategies through live instruction and computerized formats.

John Culbertson, MA, MEd, is director of research at the American Society of Radiologic Technologists. His current research projects include working with the Bureau of Labor Statistics to track the supply and demand of radiologic technologists in the United States.

Donald D. Deshler, PhD, is Williamson Family Distinguished Professor of Special Education and director of the Center for Research on Learning at the University of Kansas. His research interests include design and validation of schoolwide frameworks for improving literacy outcomes.