Learning Disabilities Research & Practice, *0*(0), 1–15

DOI: 10.1111/ldrp.12259

35 Years on the Road from Research to Practice: A Review of Studies on Four Content Enhancement Routines for Inclusive Subject-Area Classes, Part II

Joseph B. Fisher
Grand Valley State University
Jean Bragg Schumaker
University of Kansas

This article is the second part of a two-part article focusing on research that has been conducted on Content Enhancement Routines, instructional routines developed to be used during inclusive subject-area instruction. Part I of this article (Schumaker and Fisher, 2021) reviews the original validation studies that were conducted on four Content Enhancement Routines. This second part of the article reviews 10 empirical studies that have been conducted comparing the effects of two professional development methods (i.e., a computerized workshop and a live workshop) for instructing teachers to use the same four teaching routines. In every study, teacher knowledge of the routine and teacher preparation for using the routine were measured. In four of the studies, teacher implementation of the routine within inclusive classes as well as student performance were also measured. Results were reported for the whole group of students in all four studies, and for students with LD in three of the studies. In all of the studies, teachers made large and significant gains in performance on all measures after both workshop conditions, representing large effect sizes. All in-service teachers performed the routine at a high level of quality in their classes after 3 hours of instruction. In two studies, the teachers who participated in the computerized instruction earned significantly higher implementation scores than the teachers who participated in the live instruction. Regarding student performance across the studies, the whole group of students and the students with LD earned significantly higher scores on the posttests than on the pretests for both groups of teachers, again representing large effect sizes. Additionally, in two studies, the whole groups of students whose teachers used the software earned significantly higher scores on posttests than the whole groups of students whose teachers participated in live sessions. These studies replicate and extend the studies reviewed in Part I of this article; they show that quality teacher use of four Content Enhancement Routines results in increases in performance for all students, and for students with LD in inclusive classes.

One of the most difficult challenges facing the education field is ensuring that instructional methods that have been shown to yield positive student outcomes are disseminated to teachers across the nation in such a way that teachers can use the methods with fidelity and produce the same types of outcomes (Darling-Hammond et al., 2017; Desimone, 2009; Fisher et al., 2010). In other words, developers need to conduct research studies built on rigorous methodologies that are acceptable to the field. They also need to show that teachers accept and can use the new instructional methods at a high level of fidelity, and that student performance improves when the new methods are used. Preferably, student performance improves in such a substantial way that it can make a socially valid difference in students' lives, thus making the instructional method worth using. Not surprisingly, there are many twists and turns along this research-to-practice road, and researchers rarely complete the whole journey.

This two-part article tells the story of one such journey, undertaken by a group of researchers, that focuses on the Content Enhancement Routines. These routines are sets of teacher and student behaviors that were developed to be used in inclusive classes in which diverse groups of students are enrolled, including students with learning disabilities (LD) and other disabilities. The routines all involve the coconstruction of meaning between the teacher and students working in partnership, and the recording of that meaning on a specially designed graphic organizer. Each routine is designed for a particular part of the instructional process in a course, such as introducing the whole course or introducing a unit, defining a concept, answering a course question, or learning vocabulary. Part I of this article (see Schumaker & Fisher, 2021) describes in detail four Content Enhancement Routines and the original research on these routines that shows that use of the routines produces statistically significant outcomes for subgroups of students enrolled in inclusive classes, including students with LD. Figure 1 shows the "research to practice road" undertaken by the developers

Website: www.edgeenterprisesinc.com

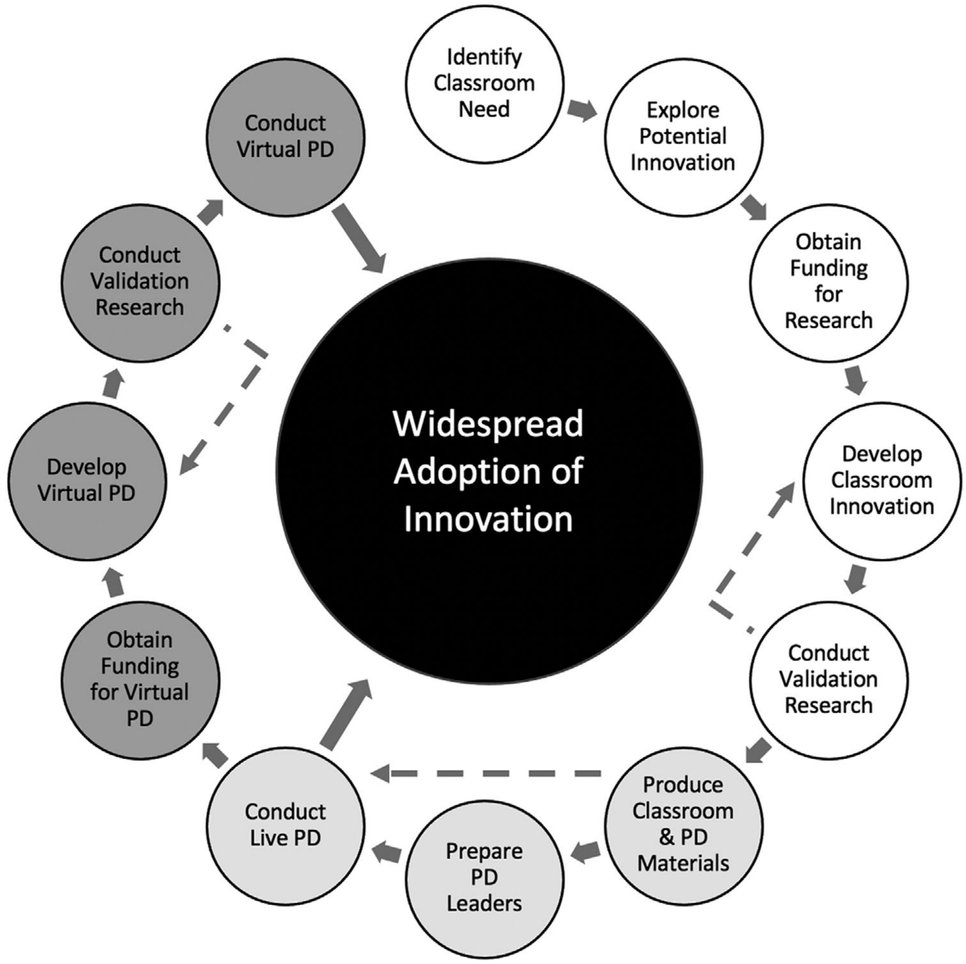

FIGURE 1 Stages of the research-to-practice process.

of the Content Enhancement Routines. The research studies reviewed in Part I of this article (Schumaker & Fisher, 2021) on a selected group of the routines fit within the "Conduct Validation Research" stage of the research-to-practice road for these new instructional methods. This article (Part II) reflects the "practice part" of the "research-to-practice" road. It reviews the research studies that have been conducted after the initial validation studies in Part I. The researchers believed that these additional studies were needed to address issues that arose regarding the maintenance of a professional development effort across the years. (These issues are explained in the discussion section of the Part I article.)

Indeed, this "practice part" of the research-to-practice road resulted from a natural phenomenon in the education field. Once the Content Enhancement developers started publishing research articles about the outcomes that can be produced by using the routines, educators began to ask the developers to provide workshops in schools across the nation for teachers. As a result, the need naturally arose for materials, presentations, and other products associated with the Content Enhancement Routines. The developers' efforts in this regard, to date, have produced a total of 14 instructor's manuals, along with presentations and other training materials associated with 14 Content Enhancement Routines, including the four routines reviewed in these current Part I and Part II articles. In addition, the developers realized quickly that they could not personally provide professional development experiences for the whole nation. As a result, they began to certify educators to provide professional development about the Content Enhancement Routines. That effort has resulted in an International Professional Developer Network composed of educators located in state departments of education, teacher-training institutions, and school districts across the nation and internationally.[1]

Although this network is now composed of hundreds of highly trained professionals, the developers began to worry that the Network was not fulfilling its purpose of putting the Content Enhancement Routines into the hands of large numbers of teachers in a way that enables teachers to use groups of routines in each course. Most of the professional developers in the network are associated with their own organization (e.g., a state department, university, or school district),

[1] For information on the International Network of Professional Developers associated with the Content Enhancement Routines, go to https://sim.ku.edu/birth-sim-network.

and do not have extra time to work outside their jobs. Additionally, school districts often cannot afford the costs associated with engaging a professional developer (e.g., travel expenses, hotel, food, and stipend). As an alternative, school administrators sometimes create professional development formats that consist of one-time, large-group, stand-and-deliver presentations (Borko et al., 2009; Tait-McCutcheon & Drake, 2016). Such an approach rarely improves teacher classroom instruction or student learning, when such improvement should be the central purpose of professional development (Deshler, 2015; Gersten & Dimino, 2001). To improve classroom instruction and educational outcomes, teachers need to participate in high-quality professional development over a period of time; doing so, however, often requires expertise, time, funds, and coordination beyond the means available (State et al., 2019). School personnel may not be willing to fund the sustained types of professional development that are required to put in place even three or four Content Enhancement Routines in inclusive courses across the curriculum for large numbers of teachers.

Indeed, to make high-quality professional development more *widely available* and *easily accessible* to teachers as well as *cost effective* for school districts, members of the educational technology community have advocated the use of e-learning or computerized learning (Bates et al., 2016; Collins & Liang, 2015; Dede et al., 2009; Edinger, 2017; Whitehouse et al., 2006). Computerized learning uses electronic communication for instruction (Wentling et al., 2000). Such programs combine text, audio, and video into interactive software that can be distributed digitally and are often called multimedia programs. The *availability* of these programs is expanding rapidly. Their growth has been especially rapid at institutions of higher education in the form of online courses (Castro, 2019); other organizations, however, are rapidly making e-learning programs available to teachers (Chang et al., 2018). State departments of education, public school districts, foundations (e.g., the George Lucas Foundation), professional organizations (e.g., ASCD), and even broadcasters (e.g., PBS) are distributing such programs. Researchers have indicated that when computerized programs are made readily available, teachers will use them to enrich their pedagogy (Walker et al., 2008).

In addition to being available, these multimedia programs are also potentially *accessible* to teachers. Today, teachers have widespread access to computers and the Internet both at school and home (Kleiman, 2004; Wells et al., 2006). In a recent report, *Education Superhighway* (2019) revealed that 99% of the nation's public schools have access to high-speed Internet. In other words, 2.8 million U.S. teachers have the connectivity needed to access online professional development.

Such multimedia programs have also been found to be *cost-effective* for delivering professional development to teachers. Abbott et al. (2006) documented that e-learning programs assisted in translating a validated instructional practice into classroom practice across a school district at a cost not possible without those programs. In other words, compared to a face-to-face approach, an e-learning approach allowed these researchers to deliver professional development using fewer personnel to conduct in-service workshops, consult, observe, and problem solve with teachers than would have been the case otherwise. In addition to reducing personnel costs, e-learning programs also reduce travel costs when compared to face-to-face approaches (Knight & Skrtic, 2020; Wentling et al., 2000).

These advantages of computerized professional development programs were attractive to the developers of the Content Enhancement Routines. As a result, beginning in 1994, the developers started a new effort, referred to as "Virtual PD" inFigure 1.” They sought and started to receive federal funding to develop and test computerized programs for the four Content Enhancement Routines featured in Part I of this article (Schumaker & Fisher, 2021). For each routine, they developed a software program consisting of text, narration, interactive quizzes and activities, digital assignments for creating a graphic device for the routine, short videoclips of teachers and students working in partnership to implement each step of the routine, and longer video examples of how teachers have used the whole routine at different levels of schooling (elementary, middle-school, and high-school settings) in different subject-areas. They then tested the effects of each software program in at least two studies. In most cases, a pair of studies was conducted: one study with *pre-service* teachers (or teachers enrolled in a college course) and one study with *in-service* teachers and their students. The in-service studies allowed the researchers to gather data on teacher implementation of routines in classrooms and student learning following implementation.

In all the studies, the teachers were randomly placed in an experimental group (called the “Virtual Workshop [VW] Group”) or an alternate treatment group (called the “Actual Workshop [AW] Group”). Additionally, in all the studies, the VW teachers worked through the software program individually on computers. The AW teachers took part in a live workshop with a certified professional development leader. This live workshop involved lecture, discussion, videoclips, and written activities. The content and the time spent (3 hours) across the live workshop and the software program were controlled. The same talking points were covered, the same visual devices and examples were used, the same videoclips were shown, and the same activities were completed as in the software program. In some of the studies, general education teachers teaching subject-area courses to inclusive classes of students participated along with their students. In three of these studies, the performance data collected from the students were disaggregated to display the performance of students with LD along with the performance of the whole group of students. The performance of students with LD was highlighted since this group of students has been shown to exhibit substantial skill deficits and to earn failing or barely passing grades in subject-area courses (Fore et al., 2008; Hughes & Schumaker, 1991; Schwartz et al., 2021a; Warner et al., 1980). These studies and the results associated with each of the four Content Enhancement Routines are described below.

THE CONCEPT MASTERY ROUTINE PROFESSIONAL DEVELOPMENT STUDIES

In the very first study in this line of programmatic research, Fisher developed the first iteration of a computerized professional development program for the Concept Mastery Routine, and he tested it in two studies (Fisher et al., 1994). In the preservice study, 58 teachers-in-training volunteered. They were individuals enrolled in an instructional methods course and were working toward earning their initial teaching certificates. Twenty-nine each were randomly assigned to the VW and AW groups (Fisher et al., 1999). Before and after they participated in their respective workshops, the teachers took a test measuring their knowledge of the Concept Mastery Routine and another test measuring their performance as they planned a lesson by creating a Concept Diagram. Both groups earned significantly higher scores on the posttest than on the pretest for each measure, representing large effect sizes. (See the Within-Group effects for Study 1 in Table 1.) The results showed no difference between the groups on the posttest. (See the Between-Group Effects in Table 1.) Their satisfaction ratings were similar on 11 of the 14 items on a satisfaction questionnaire. The three items for which statistically significant differences were found pertained to how well the teachers thought they would remember, could summarize, and had understood the information about the routine, with the VW teachers rating their understanding higher than the AW teachers.

The second study, also reported by Fisher et al. (1999), included 10 in-service teachers (8 general education teachers and 2 special education teachers), who volunteered to use the routine in their general-education subject-area inclusive classrooms. Five were randomly selected to receive live instruction; the other five worked through the software program on individual computers. A multiple-baseline across-teachers design was utilized for the implementation measure. Results showed that both groups of teachers used the routine in their classes after instruction at high levels of quality. (See Study 2 in Table 1.) Both groups of teachers earned significantly higher implementation scores after instruction than during baseline (representing large effect sizes), with no overlap between the baseline and postinstruction scores in either group. Additionally, the teachers in both groups earned significantly higher scores on the knowledge posttest and the planning posttest compared to the pretests, representing large effect sizes in each case. The satisfaction ratings of the in-service teachers in the two groups were similar.

For a second pair of studies (i.e., two more studies) related to the Concept Mastery Routine, Fisher et al. (2010) revised the software program and again compared its effects to the effects of live instruction. In a manner similar to the previous preservice study, 30 teachers participated in the VW group, and 29 teachers participated in the AW group. All of these were general education teachers of subject-area courses who were enrolled in a graduate-level college course. The teachers took a knowledge test before and after training. They were also asked to complete a Concept Diagram before and after instruction. They completed a satisfaction questionnaire after training. The results showed that the teachers in both groups earned substantially and statistically higher scores on the knowledge and preparation posttests than on the pretests. (See Study 1 in Table 2.) There were no differences between the posttest scores of the VW and AW groups on either outcome measure. The mean satisfaction ratings were similar across the groups on 10 of the 14 items. On four of the items, the teachers who received the live training provided significantly higher satisfaction scores: willingness to implement the routine, the level of enjoyment during the professional development (PD) experience, participant engagement during the experience, and overall satisfaction with the PD.

In the *in-service* part of this same research project, Fisher et al. (2010) found similar results across the groups, but, this time, student results were also gathered. Four in-service teachers participated in each of the two groups, along with their 125 students. The students participated in the teachers' use of the Concept Mastery Routine when the teachers used the routine according to their assigned placement in the multiple-baseline across-teachers design. Data on the teacher knowledge measure and the preparation measure were not gathered. A pretest/posttest experimental design was used to determine the effects of the teachers' implementation of the Concept Mastery Routine on student learning about the targeted concepts.

The results (see Study 2 in Table 2) showed that the teachers in both groups made substantial gains with regard to implementing the routine in their classes. The average post-training implementation score was 75.2% for the AW group, and was 88.5% for the VW group. There was 0% overlap between the baseline and postinstruction scores for either group. Both groups earned significantly higher implementation scores after training than before training, representing large effect sizes. Moreover, the VW group earned significantly higher after-training implementation scores than the AW group, also representing a large effect size.

With regard to student acquisition of knowledge about the concept taught through the use of the Concept Mastery Routine, both groups of students earned significantly higher scores on the posttest than on the pretest. An ANCOVA revealed no significant differences between the posttest scores of the student groups while controlling for the pretest scores. Likewise, there were no significant differences between the satisfaction ratings of the two groups of students.

THE CONCEPT COMPARISON ROUTINE PROFESSIONAL DEVELOPMENT STUDIES

In the third pair of studies in this line of programmatic research, Schumaker et al. (2010) focused on the Concept Comparison Routine. They made refinements in the basic structure of the software program and then inserted unique information about the Concept Comparison Routine into that structure. Their measures were tailored to the Concept Comparison Routine; in other words, the teachers in their studies had to show their knowledge of that routine and create a Concept Comparison Diagram during the written tests. Additionally, two satisfaction questionnaires were used. On the Workshop Satisfaction Questionnaire, the teachers indicated their satisfaction with their respective workshop. On the

TABLE 1
Results of the First Pair of Professional Development Studies on the Concept Mastery Routine

Study & sample size	*Design*	*Measure*	*Group*	*Means*		*Within-Group Effects*			*Between-Group Effects*		
				Pre (SD)	*Post (SD)*	*Statistic*	p *Value*	*Effect size*	*Statistic*	p *Value*	*Effect size*
Study 1 Fisher et al. (1999) $N = 58$ Preservice teachers $n = 29$ experimental $n = 29$ control	PPCG	Teacher knowledge	AW	2.30% (0.90)	53.99% (6.31)	$t(28) = 17.05$	$p < .01^*$	**$d = 8.225$**[a]	$F(1, 55) = 0.44$	$p = .51$	**$d = 0.174$**
			VW	0.90% (0.67)	49.09% (6.46)	$t(28) = 15.35$	$p < .01^*$	**$d = 7.497$**[a]			
		Teacher preparation	AW	8.30% (6.69)	84.14% (11.44)	$t(28) = 38.70$	$p < .01^*$	**$d = 5.99$**[a]	$F(1, 55) = 0.00$	$p = .98$	**$d = 0.000$**
			VW	10.03% (7.47)	84.76% (10.92)	$t(28) = 37.99$	$p < .01^*$	**$d = 5.931$**[a]			
Study 2 Fisher et al. (1999) $N = 10$ In-service teachers $n = 5$ experimental $n = 5$ control	PPCG	Teacher knowledge	AW	6.32% (1.82)	60.00% (3.56)	$z = 2.21$	$p < .04^*$	**$d = 14.006$**[a]	$X^2(1, N = 10) = 0.18$	$p < .67$	**$d = 0.271$**[a]
			VW	5.36 % (1.23)	63.68% (4.15)	$z = 2.19$	$p < .04^*$	**$d = 13.856$**[c]			
		Teacher preparation	AW	39.09% (38.18)	85.45% (7.42)	$z = 1.80$	$p < .04^*$	**$d = 1.215$**[c]	$X^2(1, N = 10) = 2.22$	$p < .14$	**$d = 1.068$**[c]
			VW	11.81% (16.05)	93.63% (10.95)	$z = 2.19$	$p < .04^*$	**$d = 4.422$**[c]			
	MBAT	Teacher implementation	AW	19.03% (11.82)	78.25% (13.27)	**$t(23) = 11.586$**	**$p < .001^*$**	**$d = 4.5575$**[c]	**$F(1, 50) = 0.864$**	**$p = .357$**	**$d = 0.263$**[a]
			VW	12.45% (9.15)	84.68% (5.54)	**$t(21) = 10.505$**	**$p < .001^*$**	**$d = 3.5954$**[a]			

Note. Bolded within-group and between-group effects were calculated after the article was published. Bolded effect sizes were calculated using calculators provided by Lenhard and Lenhard (2016).
PPCG = Pretest-Posttest Control-Group Design; MBAT = Multiple-Baseline Across-Teachers Design; AW = Actual Workshop; VW = Virtual Workshop.
[a] Reflects small effect size.
[b] Reflects medium effect size.
[c] Reflects large effect size.
*Indicates statistically significant p value.

TABLE 2
Results of the Second Pair of Professional Development Studies on the Concept Mastery Routine

				Means		*Within-Group Effects*			*Between-Group Effects*		
Study & sample size	*Design*	*Measure*	*Group*	*Pre (SD)*	*Post (SD)*	*Statistic*	p *Value*	*Effect size*	*Statistic*	p *Value*	*Effect size*
Study 1 Fisher et al. (2010)	PPCG	Teacher knowledge	AW	0.93% (2.03)	65.72% (11.77)	$t(28) = -35.15$	$p < .01$*	**$d = 5.518$**[c]	$F(1, 58) = 1.28$	$p = .263$	**$d = 0.295$**[c]
$N = 59$ Teachers in-training			VW	0.53% (2.17)	68.26% (10.35)	$t(29) = -29.67$	$p < .01$*	**$d = 6.537$**[c]			
$n = 30$ experimental $n = 29$ control		Teacher preparation	AW	6.13% (7.46)	87.51% (8.98)	$t(28) = -35.39$	$p < .01$*	**$d = 7.341$**[c]	$F(1, 58) = 0.60$	$p = .440$	**$d = 0.202$**[a]
			VW	4.50% (5.46)	85.80% (10.86)	$t(28) = -29.67$	$p < .01$*	**$d = 6.974$**[c]			
Study 2 Fisher et al. (2010)	MBAT	Teacher implementation	AW	1.79% (.97)	75.20% (6.16)	**$t(19) = 37.82$**	**$p < .001$***	**$d = 13.29$**[c]	**$F(1, 43) = 35.01$**	**$p < .001$**	**$d = 1.805$**[c]
$N = 8$ In-service teachers			VW	1.85% (1.28)	88.51% (4.36)	**$t(18) = 53.09$**	**$p < .001$***	**$d = 19.86$**[c]			
$n = 4$ experimental $n = 4$ control	PPCG	All student acquisition	AW	10.25% (9.17)	62.00% (16.81)	$t(7) = 8.35$	$p = .004$*	**$d = 2.824$**[c]	$F(1, 7) = 0.30$	$p = .606$	**$d = 0.387$**[c]
$N = 125$ Students $n = 76$ experimental $n = 49$ control			VW	14.75% (11.78)	67.75% (4.25)	$t(7) = 9.93$	$p = .002$*	**$d = 4.374$**[c]			

Note. Bolded within-group and between-group effects were calculated after the article was published. Bolded effect sizes were calculated using calculators provided by Lenhard and Lenhard (2016).
PPCG = Pretest-Posttest Control-Group Design; MBAT = Multiple-Baseline Across-Teachers Design; AW = Actual Workshop; VW = Virtual Workshop.
[a] Reflects small effect size.
[b] Reflects medium effect size.
[c] Reflects large effect size.
*Indicates statistically significant p value.

Software Program Satisfaction Questionnaire, the teachers who received the computerized instruction indicated their satisfaction with the software program.

In the *preservice study*, 28 teachers participated in the computerized instruction, and 32 teachers participated in the live instruction. (Forty-eight were general education teachers, nine were special education teachers, and the rest were earning their certificates.) Both groups of teachers earned significantly higher scores on the knowledge posttest and the diagram-completion posttest than on the respective pretests, representing large effect sizes. (See Study 1 in Table 3.) An ANCOVA revealed that the VW teachers earned significantly higher scores on the knowledge posttest than the AW teachers, representing a large effect size. Such a difference was not found on the teacher preparation test because both groups of teachers earned very high mean scores on the posttest (above 96%). Both groups of teachers were satisfied with their training, and there were no differences found between their satisfaction ratings. The teachers who participated in the computerized instruction indicated that they were satisfied with the specified elements of the software program.

In the *in-service* study (see Study 2 in Table 3), 11 teachers used the software program, and 10 teachers took part in the live workshop about the Concept Comparison Routine. All of the teachers were currently teaching inclusive general-education subject-area courses in 6th through 12th grades. The teachers' students ($N = 292$) participated in the teachers' use of the Concept Comparison Routine when it was implemented by the teachers in their classes, and each student only had one teacher using the routine. The researchers disaggregated the scores of students with LD from the scores of the whole groups of students for the first time in this series of professional development studies. A multiple-baseline design revealed that the postinstruction implementation scores by both groups of teachers in their classrooms were significantly higher than the baseline implementation scores. Furthermore, an ANCOVA revealed that the VW teachers earned significantly higher implementation scores after instruction than the AW group, representing a large effect size. In fact, in 31 of 32 lessons where they implemented the routine, the VW teachers exceeded the mastery criterion (i.e., 80%). In contrast, the AW teachers exceeded the mastery criterion in 24 of 30 lessons. Nevertheless, there was 0% of overlap of scores between the baseline and postinstruction conditions earned by both groups. Interestingly, the AW group, on average, barely exceeded the mastery criterion (with a mean score of 84%), while the VW group substantially exceeded it (with a mean score of 91.18%). With regard to their knowledge scores, both groups of in-service teachers earned significantly higher scores after their instruction than on the pretests. Likewise, their scores on the diagram-completion test were significantly higher after instruction than before instruction. There were no statistical differences between the groups on these measures, or in terms of their satisfaction with the training. The group of teachers who received the computerized instruction were highly satisfied with their instruction, providing an overall rating of 6.42 on a 7-point scale.

Concerning the student results, the data from the groups of students were analyzed using hierarchical linear model (HLM) analyses to account for the nested nature of the data. Both whole groups of students earned significantly higher scores on tests of concept knowledge after their teachers used the routine than before the teachers used the routine. There were no differences in the posttest scores earned by the two groups of students. When the scores of students with LD were analyzed separately, the students of teachers in the VW group ($n = 39$) earned significantly higher scores on the posttest than on the pretest, as did the students of teachers in the AW group ($n = 34$). Large effect sizes were realized by both subgroups of students with LD with regard to within-group differences. Although the students with LD in the VW group earned higher posttest scores than those in the AW group, and a medium effect size resulted, no significant difference was found between the posttest scores of the students with disabilities in the two groups.

THE CONCEPT ANCHORING ROUTINE PROFESSIONAL DEVELOPMENT STUDIES

For the fourth pair of studies in this line of research, Schumaker et al. (2021a) focused on a routine that required even more abstract thinking than the previous routines—the Concept Anchoring Routine. This routine requires teachers (and students) to create an analogy between a new concept and a known concept. Analogical reasoning is a more difficult mental process than naming the characteristics and examples of a concept. The researchers questioned whether computerized instruction could be as effective with a routine that requires a more difficult level of higher order thinking. Furthermore, as a result of potential criticism that their software programs did not include all the elements of professional development recommended in the literature (e.g., Darling-Hammond et al., 2017; Desimone, 2009), the researchers wondered whether adding the element of collaboration among teachers participating in the live workshop would produce a difference between the groups. Thus, teachers in the AW groups in these Concept Anchoring Routine studies participated in discussions and collaborative activities; those in the VW group worked through the software program individually.

In the first study of the pair, 21 teachers served in the VW group, and 22 served in the AW group. They were all general education teachers enrolled in a retraining program to become special education teachers. A teacher knowledge test tailored to the Concept Anchoring Routine and a test requiring the teachers to create an Anchoring Table were used. A different training satisfaction questionnaire was developed for each group, with the questions tailored to the type of training. With regard to the knowledge test, both groups of teachers earned significantly higher scores on the posttest than on the pretest, and there were no statistical differences between the groups' posttest scores. Similar positive results were achieved with the Anchoring Table Test for both groups. Large effect sizes were achieved on both measures. (See Study 1 in Table 4.) Ratings on the questionnaire were positive for both groups, but the ratings could not be compared across the groups because the questionnaires were different.

TABLE 3
Results of the Professional Development Studies on the Concept Comparison Routine

				Means		*Within-Group Effects*			*Between-Group Effects*		
Study & sample size	*Design*	*Measure*	*Group*	*Pre* (SD)	*Post* (SD)	*Statistic*	p *Value*	*Effect size*	*Statistic*	p *Value*	*Effect size*
Study 1 Schumaker et al. (2010) $N = 60$ Teachers in-training $n = 28$ experimental $n = 32$ control	PPCG	Teacher knowledge	AW	8.78% (6.84)	54.56% (21.37)	$F(1, 31) = 151.21$	$p < .0001^*$	**$d = 2.102$**[c]	$F(1, 56) = 29.11$	$p = .0001^*$	$d = 1.396$[c]
			VW	2.44% (4.23)	79.96% (14.47)	$F(1, 26) = 786.03$	$p < .0001^*$	**$d = 5.286$**[c]			
		Teacher preparation	AW	34.63% (18.01)	96.09% (8.46)	$F(1, 31) = 329.13$	$p < .0001^*$	**$d = 3.215$**[c]	$F(1, 57) = 0.145$	$p = .704$	**$d = 0.099$**
			VW	27.75% (18.85)	96.75% (7.30)	$F(1, 27) = 321.14$	$p < .0001^*$	**$d = 3.535$**[c]			
Study 2 Schumaker et al. (2010) $N = 21$ In-service teachers $n = 11$ experimental $n = 10$ control	PPCG	Teacher knowledge	AW	13.00% (11.37)	87.80% (12.99)	$t(9) = -15.87$	$p < .001^*$	**$d = 4.565$**[c]	$F(1, 18) = 0.067$	$p = .798$	**$d = 0.113$**
			VW	12.45% (11.98)	86.45% (10.09)	$t(10) = -13.81$	$p < .001^*$	**$d = 4.976$**[c]			
		Teacher preparation	AW	35.10% (26.91)	96.40% (4.65)	$t(10) = -7.42$	$p < .001^*$	**$d = 2.283$**[c]	$F(1, 18) = 0.113$	$p = .741$	**$d = 0.147$**
			VW	34.27% (26.50)	97.09% (4.18)	$t(10) = -7.40$	$p < .001^*$	**$d = 2.379$**[c]			
$N = 292$ Students $n = 134$ experimental $n = 158$ control	MBAT	Teacher implementation	AW	14.79% (4.00)	84.07% (6.00)	$t(9) = -27.86$	$p < .001^*$	**$d = 9.31$**[c]	$F(1, 18) = 8.78$	$p = .008^*$	$d = 1.295$[c]
			VW	15.18% (2.10)	91.18% (5.65)	$t(10) = -35.67$	$p < .001^*$	**$d = 13.71$**[c]			
$N = 73$ Students w/ LD $n = 39$ experimental $n = 34$ control	PPCG	All student acquisition	AW	17.64% (19.93)	73.99% (27.78)	$F(1, 9) = 545.30$	$p < .0001^*$	**$d = 1.732$**[c]	$F(1, 18.7) = 0.05$	$p = .82$	**$d = 0.098$**
			VW	20.53% (15.62)	74.59% (22.92)	$F(1, 10) = 569.89$	$p < .0001^*$	**$d = 2.047$**[c]			
		Students w/ LD acquisition	AW	15.85% (18.67)	56.88% (33.05)	$F(1, 8) = 49.27$	$p < .0001^*$	**$d = 1.13$**[c]	$F(1, 11.3) = 1.45$	$p = .25$	**$d = 0.526$**
			VW	15.54% (15.31)	69.09% (20.62)	$F(1, 7) = 221.40$	$p < .0001^*$	**$d = 2.193$**[c]			

Note. Bolded effect sizes were calculated after the article was published using calculators provided by Lenhard and Lenhard (2016).
PPCG = Pretest-Posttest Control-Group Design; MBAT = Multiple-Baseline Across-Teachers Design; AW = Actual Workshop; VW = Virtual Workshop; LD = Learning Disabilities.
[a] Reflects small effect size.
[b] Reflects medium effect size.
[c] Reflects large effect size.
*Indicates statistically significant *p* value.

TABLE 4
Results of the Professional Development Studies on the Concept Anchoring Routine

				Means		Within-Group Effects			Between-Group Effects		
Study & sample size	Design	Measure	Group	Pre (SD)	Post (SD)	Statistic	p Value	Effect size	Statistic	p Value	Effect size
Study 1 Schumaker et al.	PPCG	Teacher knowledge	AW	0.00% (0.00)	71.0% (17.20)	$F(1, 21) = 375.52$	$p < .001^{*}$	$\eta^2 = 0.947$[c]	$F(1, 41) = 0.004$	$p = .952$	**$d = 0.019$**
(2021a) $N = 43$ Teachers			VW	1.00% (2.00)	71.10% (20.70)	$F(1, 21) = 253.58$	$p < .001^{*}$	$\eta^2 = 0.924$[c]			
In-training $n = 21$		Teacher preparation	AW	8.40% (13.70)	94.20% (1.40)	$F(1, 21) = 515.80$	$p < .001^{*}$	$\eta^2 = 0.961$[c]	$F(1, 41) = 0.047$	$p = .830$	**$d = 0.066$**
experimental $n = 22$ control			VW	0.05% (1.00)	94.80% (9.40)	$F(1, 21) = 532.14$	$p < .001^{*}$	$\eta^2 = 0.962$[c]			
Study 2 Schumaker et al.	PPCG	Teacher knowledge	AW	0.00% (1.20)	82.80% (12.40)	$F(1, 10) = 192.45$	$p < .001^{*}$	$\eta^2 = 0.951$[c]	$F(1, 21) = 0.002$	$p = .966$	**$d = 0.018$**
(2021a) $N = 24$ In- service			VW	3.70% (8.00)	81.90% (16.20)	$F(1, 11) = 532.14$	$p < .001^{*}$	$\eta^2 = 0.980$[c]			
Teachers $n = 12$		Teacher preparation	AW	0.00% (0.00)	97.70% (5.40)	$F(1, 10) = 1947.3$	$p < .001^{*}$	$\eta^2 = 0.994$[c]	$F(1, 21) = 0.425$	$p = .521$	**$d = 0.266$**[a]
experimental $n = 12$ control			VW	5.80% (7.20)	97.70% (5.40)	$F(1, 11) = 3854.5$	$p < .001^{*}$	$\eta^2 = 0.997$[c]			
$N = 244$ Students $n = 130$	MBAT	Teacher imple-mentation	AW	16.10% (5.10)	84.20% (9.00)	$F(1, 22) = 1.160$	$p < .001^{*}$	$\eta^2 = 0.961$[c]	$F(1, 21) = 1.009$	$p = .327$	**$d = 0.410$[a]**
experimental $n = 114$ control			VW	15.20% (6.90)	87.40% (5.40)	$F(1, 22) = 810.223$	$p < .001^{*}$	$\eta^2 = 0.974$[c]			
$N = 62$ Students w/ LD	PPCG	All student acquisition	AW	17.21% (14.78)	51.15% (34.96)	t(20.5) = 7.13	$p < .0001^{*}$	$d = 0.949$[b]	$F(1, 17.9) = 7.91$	$p = .012^{*}$	$d = 0.38$[a]
$n = 26$ experimental			VW	14.67% (12.66)	70.37% (34.00)	$t(19.8) = 10.95$	$p < .0001^{*}$	$d = 1.49$[c]			
$n = 36$ control		Students w/ LD acquisition	AW	13.00% (12.63)	46.58% (29.67)	$t(35) = 6.001$	$p < .0001^{*}$	$d = 1.0$[c]	$F(1, 61) = 2.139$	$p = .149$	$d = 0.387$[a]
			VW	16.27% (13.48)	59.64% (33.89)	$t(24) = 6.860$	$p < .001^{*}$	$d = 1.3$[c]			

Note. Bolded effect sizes were calculated after the article was published using calculators provided by Lenhard and Lenhard (2016).
PPCG = Pretest-Posttest Control-Group Design; MBAT = Multiple-Baseline Across-Teachers Design; AW = Actual Workshop; VW = Virtual Workshop; LD = Learning Disabilities.
[a] Reflects small effect size.
[b] Reflects medium effect size.
[c] Reflects large effect size.
*Indicates statistically significant *p* value.

In the *in-service* study, 24 teachers participated, with 12 teachers randomly assigned to each group. These general education teachers taught a variety of grade levels (4th, 8th, 9th, 11th, and 12th) and subjects (e.g., English, science, algebra, social studies, and computer science). A new questionnaire was added to determine teacher satisfaction with the routine once the teachers had implemented it in their courses. Results of the multiple-probe design showed that the teachers who participated in both training workshops earned significantly higher implementation scores when using the routine in their classrooms after training than during baseline. Although the group that participated in the VW group earned more scores above the arbitrarily set mastery level (80%) than the other group (i.e., VW: 30/36 versus AW: 25/36), no statistical difference was found between the groups' after-training scores. Similar results were achieved on the Knowledge Test and on the Anchoring Table Test; both groups earned significantly higher scores on the posttest than on the pretest, and their scores represented high-quality performances. No differences were found between the groups on these measures. Likewise, there were no differences between the groups with regard to the median satisfaction ratings of the routine after it was implemented. (See Study 2 in Table 4.)

There were 244 students with parental consent enrolled in the participating in-service teacher's classes. These students participated in the Concept Anchoring Routine each time their teachers implemented it. Each student had only one teacher implementing the routine. They also took a concept acquisition test and filled out a satisfaction questionnaire. When the student test scores were analyzed for the whole groups of students using the multilevel modeling (MLM) approach to allow for the nested nature of the data, the students in both groups earned significantly higher scores on the posttest than on the pretest, representing large effect sizes. Moreover, the whole group of students in the VW teachers' classes earned significantly higher scores than the whole group of students in the AW teachers' classes, representing a small effect size. There were no differences between the median satisfaction ratings of the whole groups of students.

The pretest and posttest scores earned by the subgroups of students with LD were initially compared using paired samples *t* tests because these students were not distributed in an equivalent way across classes. Both subgroups of students with LD earned significantly higher scores on the posttest than on the pretest, representing large effect sizes. An ANCOVA was then used to compare the difference between the scores of the VW students with LD and the AW students with LD. It was not significant, and the effect size representing this difference was small. There were no differences between the median satisfaction ratings of the subgroups of students with LD.

THE QUESTION EXPLORATION ROUTINE PROFESSIONAL DEVELOPMENT STUDIES

In the fifth pair of studies focusing on the professional development of teachers learning a Content Enhancement Routine, Schumaker et al. (2021b) targeted the Question Exploration Routine. This routine was considered to be even more cognitively difficult than the concept routines targeted in the studies described above. The researchers wondered whether computerized instruction would be as effective with an even more cognitively difficult routine. Additionally, they wondered whether the added element of collaboration among participants (Darling-Hammond et al., 2017; Desimone, 2009) would make a differrence in the performance for the live-workshop group. One study was conducted with general education teachers taking a graduate-level college course and seeking certification in special education, and one study was conducted with in-service teachers who implemented the routine in their inclusive general education subject-area classes. Regarding the 20 graduate students in Study 1, 10 were randomly assigned to each of the two training groups. The measures for this study were tailored to the Question Exploration Routine, including a Knowledge Test, a Question Exploration Guide Test, a training satisfaction questionnaire (the same for both training groups), and a software satisfaction questionnaire (administered to the VW group only).

The results of Study 1 showed that on the Knowledge Test and on the Question Exploration Guide Test, the teachers in both groups earned significantly higher scores on the posttest than they earned on the pretest, with the within-group differences representing very large effect sizes. (See Study 1 in Table 5.) No difference was found between the posttest scores of the two groups. Likewise, no differences were found between the groups' training satisfaction ratings. The ratings of the VW group on the software satisfaction questionnaire indicated that they were highly satisfied with the software program; the overall mean rating was 6.27 on a 7-point scale.

With regard to the *in-service* teachers, 21 general education teachers from a variety of subject areas volunteered. Ten were randomly assigned to each of the training groups. Their 262 students participated in the Question Exploration Routine when the teachers used it in their classes. Each student had only one teacher use the routine. The same measures used in the first study were used in this study, with the addition of an implementation measure, a routine satisfaction questionnaire, a student knowledge test, and a student satisfaction questionnaire. On the implementation measure, the teacher knowledge measure, and the teacher preparation measure, both groups of teachers earned significantly higher scores after training than during baseline, and their mean scores after training represented high levels of performance. (See Study 2 in Table 5.) Within-group effect sizes were very large on all measures. There were no differences between the posttest scores of the two groups of teachers. After training, 30 out of 30 performances of the routine in the classrooms by the VW teachers met the mastery criterion of 80% or above; for the AW group, 25 out of 30 performances met the mastery criterion. There was no overlap of scores between baseline and postinstruction conditions for either group. Meanwhile, the ratings of the groups on the satisfaction questionnaires were high, and there were no statistical differences in satisfaction ratings across the two groups of teachers. The VW group teachers

TABLE 5
Results of the Professional Development Studies on the Question Exploration Routine

				Means		Within-Group Effects			Between-Group Effects		
Study & sample size	*Design*	*Measure*	*Group*	*Pre (*SD*)*	*Post (*SD*)*	*Statistic*	p *Value*	*Effect size*	*Statistic*	p *Value*	*Effect size*
Study 1 Schumaker et al.	PPCG	Teacher knowledge	AW	1.20% (2.70)	66.70% (13.38)	$F(1, 9) = 234.80$	$p < .001^{*}$	$d = 9.69^{c}$	$F(1, 17) = 0.856$	$p = .368$	$\mathbf{d = 0.414^{a}}$
(2021b) $N = 20$ Teachers			VW	0.80% (1.69)	73.00% (19.39)	$F(1 ,9) = 149.24$	$p < .001^{*}$	$d = 7.73^{c}$			
in-training $n = 10$ experimental		Teacher preparation	AW	25.5% (19.9)	92.7% (8.6)	$F(1, 18) = 104.89$	$p < .001^{*}$	$d = 4.58^{c}$	$F(1, 17) = 2.369$	$p = .142$	$\mathbf{d = 0.688^{b}}$
$n = 10$ control			VW	25.1% (21.89)	93.9% (8.6)	$F(1, 18) =$ 103.961	$p < .001^{*}$	$d = 4.56^{c}$			
Study 2 Schumaker et al.	PPCG	Teacher knowledge	AW	8.33% (7.63)	80.90% (9.84)	$F(1, 9) =$ -484.21	$p < .0001^{*}$	$d = 13.92^{c}$	$F(1, 16) = 2.17$	$p = .16$	$\mathbf{d = 0.659}^{b}$
(2021b) $N = 20$ In- service			VW	7.90% (8.50)	86.80% (8.47)	$F(1, 9) =$ -520.26	$p < .0001^{*}$	$d = 14.43^{c}$			
teachers $n = 10$ experimental		Teacher preparation	AW	22.50% (19.47)	92.70% (8.72)	$F(1, 9) = 180.0$	$p < .001^{*}$	$d = 8.49^{c}$	$F(1, 17) = 0.03$	$p = .864$	$\mathbf{d = 0.078}$
$n = 10$ control $N = 243$ Students			VW	25.10% (19.47)	93.90% (8.62)	$F(1, 9) = 164.0$	$p < .001^{*}$	$d = 8.10^{c}$			
$n = 127$ experimental	MBAT	Teacher imple- mentation	AW	25.98% (6.30)	84.57% (8.27)	$F(1, 9) = 216.78$	$p < .001^{*}$	$d = 9.31^{c}$	$F(1, 17) = 0.245$	$p = .627$	$\mathbf{d = 0.220}^{b}$
$n = 116$ control $N = 70$ Students w/			VW	25.24% (7.60)	86.20% (4.39)	$F(1, 10) = 517.21$	$p < .001^{*}$	$d = 13.71^{c}$			
LD $n = 32$ experimental	PPCG	All student acquisition	AW	10.52% (14.15)	47.14% (31.25)	$t(32.3) = 4.43$	$p < .0001^{*}$	$d = 1.539^{c}$	$F(1, 20.8) = 3.28$	$p = .085$	$\mathbf{d = 1.26}^{c}$
$n = 38$ control			VW	8.94% (13.09)	63.25% (27.60)	$t(57) = 5.26$	$p < .0001^{*}$	$d = 2.514^{c}$			
		Students w/ LD acquisition	AW	8.20% (14.13)	30.79% (27.80)	$t(37) = 4.678$	$p < .001^{*}$	$d = 0.76^{b}$	$F(1, 31) = 0.528$	$p = .47$	$\mathbf{d = 0.174}$
			VW	3.94% (10.16)	48.52% (33.99)	$t(29) = 7.546$	$p < .001^{*}$	$d = 1.0^{c}$			

Note. Bolded effect sizes were calculated after the article was published using calculators provided by Lenhard and Lenhard (2016).
PPCG = Pretest-Posttest Control-Group Design; MBAT = Multiple-Baseline Across-Teachers Design; AW = Actual Workshop; VW = Virtual Workshop; LD = Learning Disabilities.
[a] Reflects small effect size.
[b] Reflects medium effect size.
[c] Reflects large effect size.
*Indicates statistically significant *p* value.

rated the software program in the "satisfied" range, with an overall rating at the 5.61 level on the 7-point scale.

In relation to the student results, which were obtained using HLM analysis, both whole groups of students earned significantly higher scores on the posttest than on the pretest, with differences representing large effect sizes. No difference was found between the posttest scores of the two groups of students. Because of the small number of students with LD and the uneven nature of their placement in the classes, ANCOVAs were used to analyze their results. The students with LD earned significantly higher scores on the posttest than on the pretest, with differences representing medium or large effect sizes. There were no differences between the two LD subgroups' posttest scores, however. No differences were found between the satisfaction ratings of the two whole groups of students, or between the ratings of the two subgroups of students with LD.

DISCUSSION

To summarize, the series of professional development studies that have focused on four Content Enhancement Routines have produced a compilation of results that replicate and expand on each other. First, the studies indicate that computerized professional development, as well as live instruction conducted by certified professional developers, can be very effective. Both groups of teachers made significant and substantial gains from pretest to posttest on all measures, representing large effect sizes in all studies. These findings applied to teacher knowledge, planning for lessons through the creation of graphic organizers, and implementation of the routine in the classroom. The quality of the implementation of the routines by the teachers was above a level considered to be "mastery" by the developers of the routines. Additionally, except in one instance related to four questionnaire items, the teachers who used computerized instruction were at least equally satisfied with the training mode and with the routine as those who participated in live instruction. The students of teachers who were trained through computerized instruction scored at least as well on tests of knowledge as the students whose teachers were trained through live instruction. Furthermore, the whole groups of students and students with LD made significant gains from pretest to posttest, representing large effect sizes. The whole groups and students with LD indicated high levels of satisfaction with the instruction that they received from their teachers, regardless of the group assignment of the teachers.

These conclusions are only part of the story covered by the 10 studies reviewed in this Part II article. In several instances across the studies, the group of teachers who engaged in the computerized instruction achieved results that were significantly better than the teachers who engaged in the live instruction. For example, in the Fisher et al. (2010) Study 2, the teachers who used the software program earned significantly higher implementation after-training scores than the live-instruction teachers. Likewise, in the Schumaker et al. (2010) Study 2, the VW teachers earned significantly higher implementation after-training scores than the AW teachers. In both studies, the effect sizes were large. Indeed, in three studies (Schumaker et al., 2010; Schumaker et al., 2021a, 2021b), more teachers who experienced the computerized instruction produced implementation performances above the 80% mastery level than the teachers who experienced live instruction. Additionally, Schumaker et al. (2010) and Schumaker et al. (2021a) found that the whole group of students of the VW teachers earned significantly higher scores than the whole group of students of the AW teachers. In other words, across the studies, when a significant between-group difference was found, the computerized instruction produced significantly higher scores on the implementation measures and on the student knowledge measure than did the live instruction. In no study did the AW group of teachers or students earn significantly higher scores than the VW group of teachers or students. In fact, with three exceptions, the mean scores of the computerized instruction group surpassed the mean scores of the live-instruction group. To the knowledge of the current authors, in no other studies to date have differences on different kinds of measures been found between computerized and live instruction.

The student results that have been achieved are also worthy of further discussion. The results of several of the studies are unique in that, not only were the performance scores of the whole groups of students compared, but the scores of the subgroups of students with LD were also compared within inclusive classes. When combined with the results of the original studies that validated each routine (see Schumaker & Fisher, 2021), these results add evidence that shows that each routine can be used to enhance the learning of different subgroups of students in inclusive subject-area classes. In fact, because the individual routines have the same structure and include the same instructional procedures, the preponderance of the evidence across *all* the studies shows that the Content Enhancement Approach as a whole can be used to enhance the learning of different subgroups of students in inclusive subject-area classes.

LIMITATIONS

Some limitations are worthy of note. First, in order to have a uniform measure of student learning across classes of students in the professional development studies, the student tests that were used were different from the types of tests that students typically take. For example, the format required students to write out their answers as opposed to choosing a multiple-choice item. This format likely reduced the potential scores that students could achieve. Additionally, the students took only one posttest in each study, and they only experienced one Content Enhancement Routine in each study. How they might have performed after experiencing several routines in a course across several tests is a question for further research. No studies to date have focused on the use of several routines in the same class.

Second, although the teachers who used the computerized programs indicated that they were satisfied with the training mode and rated aspects of the software program highly, across several studies they indicated that the training time (3 hours) required to complete the software program was too long. In all of the studies, the teachers were required to use the software program in one sitting (one class period for graduate students and one after-school session for in-service teachers) to ensure that each teacher completed the whole program. The teachers who participated in the live professional development sessions did not give similar low levels of ratings for the length of training, even though those sessions also lasted 3 hours. Perhaps sitting at a computer for a 3-hour block of time is perceived differently than taking part in a live workshop session containing discussions and activities.

Third, the 14 studies summarized here and in Part I of this article (Schumaker & Fisher, 2021) have focused on only four Content Enhancement Routines. The results may be limited to these routines. Since, however, all the routines are structured similarly and involve the same types of activities, the results are likely to be replicated in future studies with other routines.

Fourth, the development of the software programs tested in the studies reviewed here was expensive. A 4-year federal grant provided the funding to develop and test the four software programs (Schumaker & Fisher, 2000). Now that the technology has been validated, additional funds would be needed to develop future software programs for other routines and other professional development targets. Currently, there are no sources of funding for similar software programs that can be used by teachers in schools across the nation. Nevertheless, such programs are needed to provide the necessary training for the nation's teachers in a scaled-up fashion.

FUTURE RESEARCH

The studies reviewed here and in the earlier validation studies reviewed in Schumaker and Fisher (2021) show that both live instruction with a certified trainer and computerized software programs produce high levels of teacher knowledge, preparation, and performance in the classroom plus student learning. They also seem to indicate that the added elements of a live instructor plus discussions and collaborative activities do not affect the outcomes in a serious way, as long as certain other elements (e.g., explicit instruction, video models, multiple exemplars in multiple subject areas) are present. Whether other professional development elements such as those recommended by Desimone (2009), like coaching and professional learning community, have a differential impact on results is not known. Future research might explore these additional elements to determine whether value is added when they become part of the professional development paradigm. For example, whether teachers use a single routine longer, or use more routines throughout a full course, if they work with other teachers in the same subject area needs to be determined.

Additionally, how student learning might be affected over time when teachers use several routines within each unit of study and administer their own tests tailored to state or national standards might also be explored. Whether student learning might be enhanced for all subgroups of students across a whole course, and whether more students pass the course, is certainly worthy of future investigation. Since there are several more Content Enhancement Routines that have been developed and validated through research, future studies might expand the research in this area to include packages of routines. Research might focus on helping teachers to build whole courses around the use of routines, and then on measuring the results with various subgroups of students on unit and competency tests.

Future studies might also explore different ways of allowing teachers to complete the software programs. For example, they might be allowed to complete a program in more than one block of time, or they might be allowed to keep the program and use it for review from time to time. Again, doing so might be a way of enhancing the length of time for which teachers will continue to use the routines during a whole course. Once they have taught a whole course with routines one time, future research might also focus on how they make adjustments in their course plans for following years. Do they add more instances of routine use? Do they make adjustments in their future plans based on what they have learned by coconstructing the information with students?

IMPLICATIONS

The preponderance of the evidence shared in this article indicates that teachers can easily learn about Content Enhancement Routines and implement them at a high level of quality in their courses while producing improved levels of learning in their students, including those with LD. In every case, statistically significant gains were realized by teachers and students as a whole group, and by students with LD. In most cases, the gains are substantial and socially significant. In other words, gains were not simply a matter of a few points of difference. Both teachers and students made substantial gains, with most of the teachers implementing the routines above the 80% level. According to the original validation studies for the four Content Enhancement Routines (see Schumaker & Fisher, 2021), the performance of students in all subgroups improved enough so that substantially larger percentages of students were earning passing grades on the tests in all subgroups. This finding is important because an intervention designed for inclusive classes must be shown to address the needs of all subgroups of students in diverse classes. Thus, the results show that the elements of pedagogy included in the routines (e.g., explicit instruction, graphic organizers depicting relationships among information, partnership learning between teachers and students, active student participation) and implemented at a high level of quality by the teachers can yield improved student learning among all subgroups in inclusive classes. The developers of the routines have followed a clear research-to-practice course that began with validation studies of each routine and were followed by professional development studies. Their story can serve as an example for other developers wishing to ensure that their programs are used in the nation's schools.

The implications for practice are clear: if teachers of inclusive subject-area courses wish to improve their student outcomes, one way to do so is to use the Content Enhancement Routines highlighted here. In every validation study reviewed in the Part I article (Schumaker & Fisher, 2021), the effects of one routine were compared to the effects of the lecture/discussion method, which is the most prominently used instructional method in subject-area courses (Putnam et al., 1992; Schumaker & Deshler, 1984). In every case, the use of the routine was superior to the lecture method, but also produced substantially more passing grades on tests than the lecture method among subgroups of students whose performance scores had room to grow.

Obviously, as shown here, the road from research to practice is a long one. The first funding for the first research study on the Content Enhancement Approach was awarded by the federal government in 1985 (Deshler & Schumaker, 1985). A total of six federal grants provided the funding for the following years. More than three decades later, 14 individual routines that cover all aspects of teaching a subject-area course have been validated, instructor's manuals have been published, a national cadre of professional developers has been trained, and professional development materials have been produced. Thousands of teachers have been trained by the International Professional Developer Network to use at least one Content Enhancement Routine, so the training manuals, methods, and materials have been thoroughly field-tested. Additionally, the approach as a whole has been supported through many replications, and four professional development software programs have been validated and are available to schools. The pedagogy and the technology are now available for a scaling-up effort related to Content Enhancement. The final step in the research-to-practice process can now be taken. Developers need to develop and make available more professional development programs that will enable schools to scale up their use of Content Enhancement Routines, and educators need to use these programs in a way that results in improved student performance.

Acknowledgments

The authors wish to thank the other primary researchers who developed and tested the Content Enhancement Routines reviewed in this article, including Drs. Janis A. Bulgren, Donald D. Deshler, and B. Keith Lenz. The authors would also like to thank Dr. Lisa Walsh for her coordination of the research studies reviewed here. These research studies were 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.

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. https://doi.org/10.1080/10573560500203525

Bates, M. S., Phalen, L., & Moran, C. (2016). Online professional development: A primer. *Kappan*, *97*(5), 70–73. http://doi.org/10.1177/0031721716629662

Borko, H., Whitcomb, J. & Liston, D. (2009). Wicked problems and other thoughts on issues of technology and teacher learning. *Journal of Teacher Education*, *60*(3), 3–7. https://doi.org/10.1177/002248710832848

Castro, R. (2019). Blended learning in higher education: Trends and capabilities. *Education and Information Technologies*, *24*, 2523–2546. http://doi.org/10.1007/s10639-019-09886-3

Chang, V., Gütl, C., & Ebner, M. (2018). Trends and opportunities in online learning, MOOCs, and cloud-based tools. In J. Voogt, G. Knezek, R. Christensen, & K. W. Lai (Eds.), *Second handbook of information technology in primary and secondary education* (pp. 1–19). Springer. https://doi.org/10.1007/978-3-319-71054-9_64

Collins, L. J., & Liang, X. (2015). Examining high-quality online teacher professional development: Teachers' voices. *International Journal of Teacher Leadership*, *6*(1), 18–34.

Darling-Hammond, L., Hyler, M. E., & Gardner, M. (2017). Effective teacher professional development. *Learning Policy Institute*. https://learningpolicyinstitute.org/product/teacher-prof-dev

Dede, C., Ketelhut, D. J., Whitehouse, P., Breit, L., & McCloskey, E. (2009). A research agenda for online teacher professional development. *Journal of Teacher Education*, *60*(1), 8–19. http://doi.org/10.1177/0022487108327554

Deshler, D. D. (2015). Moving in the right direction but at what speed, and how smoothly? *Remedial and Special Education*, *36*(2), 72–76. http://doi.org/10.1177/0741932514558093

Deshler, D. D., & Schumaker, J. B. (1985). *The development and validation of a set of teaching routines for enhancing the performance of low-achieving students in secondary mainstream courses*. Office of Special Education Programs, U. S. Department of Education.

Desimone, L. M. (2009). Improving impact studies of teachers' professional development: Toward better conceptualizations and measures. *Educational Researcher*, *38*(3), 181–199. http://doi.org/10.3102/0013189×08331140

Edinger, M. J. (2017). Online teacher professional development for gifted education: Examining the impact of a new pedagogical model. *Gifted Child Quarterly*, *61*(4), 300–312. http://doi.org/10.1177/0016986217722616

Education Superhighway. (2019). *State of the states: The classroom connectivity gap is closed*. https://stateofthestates.educationsuperhighway.org/#national

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. http://doi.org/10.2307/1511271

Fisher, J. B., Schumaker, J. B., Culbertson, J., & Deshler, D. D. (2010). Effects of a computerized professional development program on teacher and student outcomes. *Journal of Teacher Education*, *61*(4), 302–312. http://doi.org/10.1177/0022487110369556

Fisher, J. B., Schumaker, J. B., & Deshler, D. D. (1994). *Development and validation of an interactive hypermedia program for training teachers in the use of a complex instructional practice*. Office of Special Education Programs, U.S. Department of Education.

Fore, C., Hagan-Burke, S., Burke, M. D., Boon, R. T., & Smith, S. (2008). Academic achievement and class placement in high school: Do students with learning disabilities achieve more in one class placement than another? *Education and Treatment of Children*, *31*(1), 55–72. https://doi.org/10.1353/etc.0.0018

Gersten, R., & Dimino, J. (2001). Realities of translating research into classroom practice. *Learning Disabilities Research and Practice*, *16*(2), 120–130. https://doi.org/10.1111/0938-8982.00013

Hughes, C. A., & Schumaker, J. B. (1991). Test-taking strategy instruction for adolescents with learning disabilities. *Exceptionality*, *2*, 205–221. https://doi.org/10.1080/09362839109524784

Kleiman, G. M. (2004). *Meeting the need for high quality teachers: E-learning solutions*. U. S. Department of Education Secretary's Leadership Summit on eLearning. https://archive.wgu.edu/system/files/artifact-files/20040712%20-%20Meeting%20the%20Need%20for%20High%20Quality%20Teachers%20-%20e-Learning%20Solutions.pdf

Knight, D. S., & Skrtic, T. M. (2020). Cost-effectiveness of instructional coaching: Implementing a design-based, continuous improvement model to advance teacher professional development. *Journal of School Leadership*, *31*(4), 318–342.

Lenhard, W., & Lenhard, A. (2016). Calculation of effect sizes. Retrieved from: https://www.psychometrica.de/effect_size.html.

Bibergau (Germany): *Psychometrica*. doi: https://doi.org/10.13140/RG.2.1.3478.4245

Putnam, L. M., Deshler, D. D., & Schumaker, J. B. (1992). The investigation of setting demands: A missing link in learning strategy instruction. In L. Meltzer (Ed.), *Strategy Assessment and Instruction for students with learning disabilities: From theory to practice* (pp. 325–351). Pro-Ed.

Schumaker, J. B. & Deshler, D. D. (1984). Setting demand variables: A major factor in program planning for the LD adolescent. *Topics in Language Disorders*, *4*(2), 22–40. https://doi.org/10.1097/00011363-198403000-00006.

Schumaker, J. B., & Fisher, J. B. (2000). *Preparing teachers for academic diversity*. National Institute of Health Small Business Innovations Grant.

Schumaker, J. B., & Fisher, J. B. (2021). 35 years on the road from research to practice: A review of studies on four Content Enhancement Routines for inclusive subject-area classes, Part I. *Learning Disabilities Research and Practice*.

Schumaker, J. B., Fisher, J. B., & Walsh, L. D. (2010). The effects of a computerized professional development program on teachers and students with and without disabilities in secondary general education classes. *Learning Disability Quarterly*, *33*(2), 111–131. https://doi.org/10.1177/073194871003300204

Schumaker, J. B., Fisher, J. B., Walsh, L. D., & Lancaster, P. E. (2021a). The effects of multimedia vs. live professional development on teachers' and students' performance related to the Question Exploration Routine. *Learning Disabilities Research & Practice*. https://doi.org/10.1177/0022219420988001

Schumaker, J. B., Walsh, L. D., Fisher, J. B., & Graner P. S. (2021b). The effects of computerized vs. face-to-face professional development on teachers' and students' performance related to the Concept Anchoring Routine. *Journal of Learning Disabilities*. https://doi.org/10.1111/ldrp.12232

Schwartz, A. E., Hopkins, B. G., & Steifel, L. (2021). The effects of special education on the academic performance of students with learning disabilities. *Journal of Policy Analysis and Management*. Advanced online publication. https://doi.org/10.1002/pam.22282

State, T. M., Simonsen, B., Hirn, R. G., & Wills, H. (2019). Bridging the research-to-practice gap through effective professional development for teachers working with students with emotional and behavioral disorders. *Behavioral Disorders*, *44*(2), 107–116. http://doi.org/10.1177/0198742918816447

Tait-McCutcheon, S., & Drake, M. (2016). If the jacket fits: A metaphor for teacher professional learning and development. *Teaching and Teacher Education*, *55*, 1–12. http://doi.org/10.1016/j.tate.2015.12.005

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

Warner, M. M., Schumaker, J. B., Alley, G. R., & Deshler, D. D. (1980). Learning disabled adolescents in the public schools: Are they different from other low achievers? *Exceptional Education Quarterly*, *1*(2), 27–35. https://doi.org/10.1177/074193258000100207

Wells, J., Lewis, L., & Greene, B. (2006). Internet access in U.S. public schools and classrooms: 1994–2005 *(NCES 2007–020)*. 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*. 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). Harvard Education Press.

About the Authors

Joseph B. Fisher is a professor of special education in the College of Education at Grand Valley State University. He completed his doctoral training with Jean B. Schumaker and Donald D. Deshler at the University of Kansas Center for Research on Learning. He is a former middle-school special education teacher, and he continues to provide instruction to children with disabilities in K–12 schools alongside his graduate and undergraduate students. He has a particular interest in the research and development of instructional practices to improve the reading and writing abilities of struggling learners.

Jean Bragg Schumaker is Professor Emeritus, University of Kansas (KU) and president of Edge Enterprises, Inc. Her doctoral degree from KU is in developmental and child psychology. Retired from the KU Center for Research on Learning and the Department of Special Education, she devotes her time to developing new instructional programs for students and teachers. She is a developer of the *Content Enhancement Routines*, the *Learning Strategies Curriculum*, the *Cooperative Thinking Strategies Series*, and the *Community Building Series*. Her research interests focus on the use of strategic instruction for students who need intensive and explicit instruction.