J Behav Educ (2009) 18:189–210
DOI 10.1007/s10864-009-9088-3

ORIGINAL PAPER

Generalizing Teaching Techniques from University to K-12 Classrooms: Teaching Preservice Teachers to Use What They Learn

Mary Catherine Scheeler · Kristie Bruno · Erin Grubb · Terri Lynn Seavey

Published online: 26 July 2009

Abstract Preservice teachers learn evidence-based practices in university classrooms but often fail to use them later on in their own K-12 classrooms. The problem may be a missing link in teacher preparation, i.e., failure to teach preservice teachers to generalize newly acquired techniques. Two experiments using multiple baseline designs across participants assessed effectiveness of a model to promote generalization and maintenance of a specific teaching skill. In Experiment 1, preservice teachers' maintenance of behavior deteriorated from practicum to student teaching when intervention consisted of training to criteria alone. When a programming for generalization component (program common stimuli) was added to the intervention, teachers in Experiment 2 generalized and maintained behavior across settings (student teaching to own classrooms) at a higher average than occurred during intervention.

Keywords Immediate feedback · Generalization · Preservice teaching · Bug-in-ear technology

Introduction

Teacher educators are fortunate today in that we have a strong knowledge base in evidence-based practices to draw upon in our work with preservice teachers in

M. C. Scheeler (✉) · T. L. Seavey
The Pennsylvania State University, Penn State Great Valley, 30 East Swedesford Road, Malvern, PA 19355, USA
e-mail: mcs13@psu.edu; merci1s@msn.com

K. Bruno
Norristown Area School District, 401 N. Whitehall Road, Norristown, PA 19403, USA

E. Grubb
Unionville-Chadds Ford School District, 740 Unionville Road, Kennett Square, PA 19348, USA

university classrooms (Brophy and Good 1986; Christenson et al. 1989; Ellis et al. 1994; Gersten et al. 1997; Rosenshine and Stevens 1986; Stein et al. 1998; Wittrock 1986). We know what to teach and how to teach it but we may be missing a critical component in teacher preparation, i.e., teaching future teachers to generalize their newly acquired skills to their own classrooms. Newly certified teachers may be considered to be highly qualified by No Child Left Behind (NCLB) standards but if they are not using highly effective teaching techniques with their own students in their own classrooms their students will bear the consequences of ineffective teaching. Ultimately, poor student achievement may not be the result of a poorly prepared teacher but rather a teacher who fails to use effective practices once in her own classroom. Unfortunately, teaching techniques that teachers learn and practice in university classrooms, practicum, and student teaching settings do not always transfer to real world classroom settings.

The need to teach preservice teachers to generalize techniques across time and settings is not a recent phenomena but one that has been well established in the literature (Boudah et al. 2001; Bowles and Nelson 1976; Engelmann 1988; Gersten et al. 1995; Greenwood and Abbot 2001; Han and Weiss 2005; Horton 1975; Leach and Conto 1999; Noell et al. 1997; Robinson and Swanton 1980; Rose and Church 1998; Scruggs and Mastropieri 1994; Vaughn et al. 2000). In a recent review of extant literature on programming for generalization with preservice teacher candidates, no empirical studies were identified to inform the field on specific ways to provide such training (Scheeler 2008). Could the reason that preservice teachers are not maintaining newly acquired techniques and generalizing them to their own classrooms be that they are not taught to do this? Could programming for generalization be the missing link in teacher preparation? Of course it is quite possible that this is occurring in institutions of higher learning across the country but currently there is no empirical support to suggest it is occurring.

Stokes and Baer (1977) are recognized for providing the seminal work in programming for generalization by developing a summary of nine techniques based on 120 articles related to generalization by school-aged students. These are: Train and hope, sequential modification, introduce to natural maintaining contingencies, train sufficient exemplars, train loosely, use indiscriminable contingencies, program common stimuli, mediate generalization, and train to generalize. In a follow up to Stokes and Baer, Rutherford and Nelson (1988) concluded that behavior changes in school-aged students were often not maintained because teacher behavior that produced the original changes did not continue or maintain once the behavior had been changed, a very ineffective approach resulting in deterioration of not just effective teacher behavior, but desired student outcomes. Prater and Sileo (2004) state that for decades, teacher education has been marginalized and under-funded. If we continue to prepare preservice teachers only to see the effects of preparation deteriorate once they enter the teaching profession, the field of teacher education may find itself maintaining an under-valued, marginalized reputation.

In a review of the literature on programming for generalization in preservice teacher preparation programs, Scheeler (2008) identified four factors as highly likely to support sustainability of teaching techniques. These are: (a) Providing immediate feedback for acquisition of new behavior; (b) training to mastery level on

the new behavior; (c) programming for generalization; and (d) providing performance feedback in classroom settings. Using these four factors Scheeler developed a sequential model for promoting preservice teachers' generalization of newly acquired teaching techniques from university to school settings (see Fig. 1).

In the first step of the model, the focus is on efficient and effective teaching of requisite teacher behavior through use of immediate (as opposed to delayed) feedback. Immediate feedback is useful because it emphasizes and reinforces correct

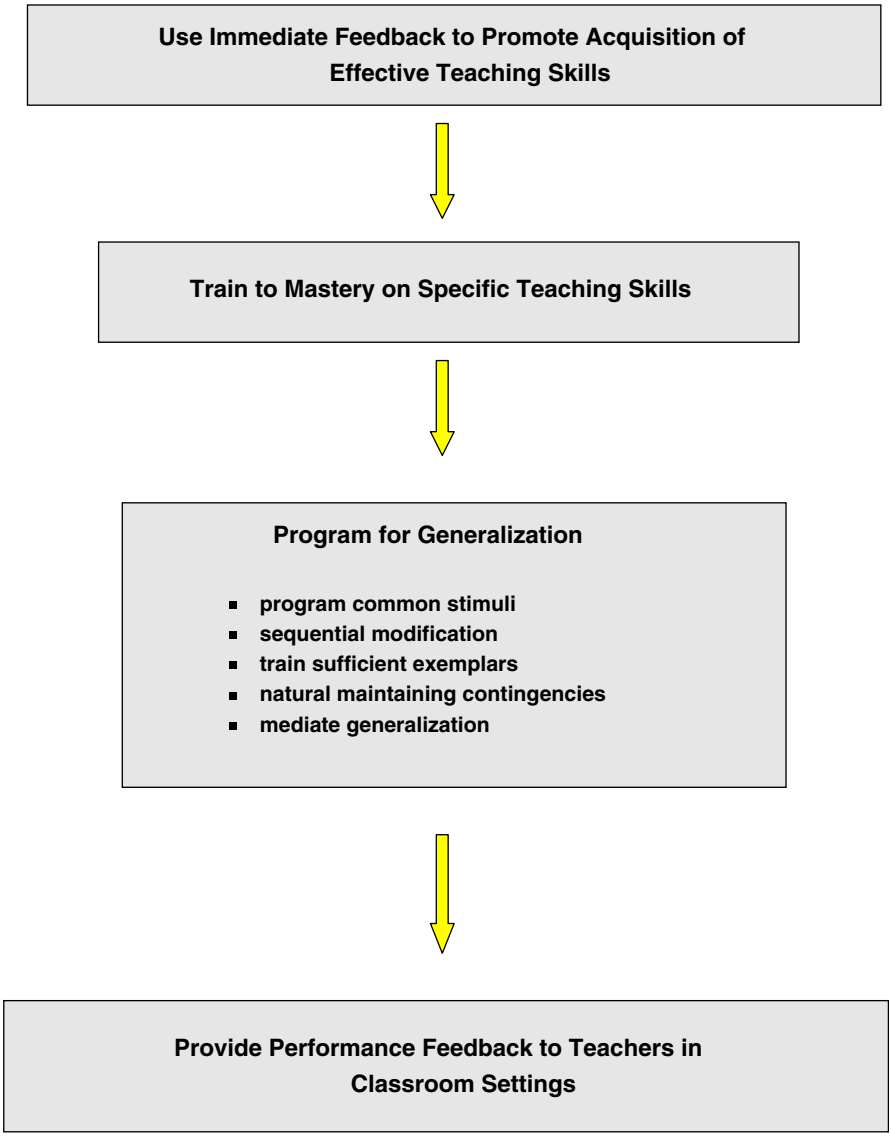

Fig. 1 Model for promoting generalization and maintenance of effective teaching skills

technique and stops the learner from practicing errors so the learner can practice correct technique in the very next learning trial (Coulter and Grossen 1997; O'Reilly et al. 1992, 1994; Scheeler and Lee 2002; Scheeler et al. 2006; Sharpe et al. 1997, Van Houten 1980). An advantage of using immediate feedback is that it provides teachers with an opportunity to change behaviors while practicing a new technique instead of repeating errors until delayed feedback is delivered (Coulter & Grossen). Empirical evidence can be found in the literature to support the use of immediate feedback with preservice teachers to effectively change teaching behaviors such as error correction procedures and point awarding (Coulter & Grossen), appropriate use of prompts and positive consequences (O'Reilly et al. 1994; 1992), and completion of learning trials (Scheeler and Lee 2002; Scheeler et al. 2006). Immediate feedback has also been documented to work with high school students acting as peer coaches to deliver feedback to each other on oral presentation skills (Scheeler et al. 2009). Recent advances in technology such as bug-in-ear technology (BIE) to deliver immediate feedback make it easy to provide feedback in a cost-effective, unobtrusive way. Use of BIE technology allows the supervisor to give feedback to the preservice teacher unobtrusively with minimal disruption to the teaching environment.

The second step in the model focuses on practicing skills until a criterion level of performance is reached. Teachers cannot sustain and generalize what they have not learned well (Engelmann 1988; Rose and Church 1998) and academic learning requires discipline and practice (Lindsley 1992). In their review of 49 studies on effects of preservice and in-service training on maintenance of teaching skills, Rose and Church found that many of the studies provided fixed amounts of training rather than training to criteria; thus, there was no way of determining whether the teachers failed to maintain newly acquired skills or if they ever really acquired the skills in the first place. Training to mastery on specific teaching behaviors would therefore seem to be an important consideration to encourage maintenance of behavior when preparing teachers. Rose and Church concluded that it made little sense to study maintenance except in cases where trainees reached a criterion level of performance of teaching skills.

The third component of the model requires systematic programming for generalization in preservice teacher preparation programs. Of the five techniques cited in the model as possible appropriate choices, two of these, program common stimuli and sequential modification, have actually been used with in-service teachers (see Epps and Lane 1987; Scruggs and Mastropieri 1994 for descriptions). Programming common stimuli seems particularly relevant to address the challenge of programming for generalization from university to K-12 settings. Stokes and Baer (1977) describe it is a technique that involves arranging to have similar stimuli occurring in both treatment and non-treatment settings. This would involve either bringing elements of the classroom setting into the training site (university classroom) or bringing salient elements of the training site into the next setting (field experience or own classroom). By systematically arranging to have similar stimuli in both settings, the preservice teacher is more likely to demonstrate newly acquired teaching techniques because the common stimulus serves as a cue or reminder to perform the behavior. This technique would be easy to use, require no new materials, and take very little time to implement.

The final step in the model focuses on providing in-class support in the form of performance feedback in order to maintain teaching techniques. Feedback (ideally immediate feedback as in step one of the model) could be from a person in authority (university supervisor, school administrator) or a peer mentor or consultant as long as the person giving feedback is well trained and effective (Noell et al. 1997). Three studies focused on maintenance and/or generalization of teaching skills (Bowles and Nelson 1976; Robinson and Swanton 1980; Rose and Church 1998) and one study described a generalization technique taught to special education and general education teachers to aid generalization in their students (Epps and Lane 1987).

Although the components of the model described in Fig. 1 have empirical support, studies have not been done on the entire model nor have studies been done with preservice teachers as the target population. The purpose of this research therefore was threefold. In the first experiment, effects of training one specific teaching technique to mastery in order to promote maintenance and generalization of this technique across settings by preservice teachers was examined. Although the methodology for Experiment 1 is essentially a replication of previous work (see Scheeler and Lee 2002; Scheeler et al. 2006), the focus of Experiment 1 was on maintaining the effects of training rather than assessing use of immediate feedback on a specific teaching behavior. Documentation of an evidence-based practice requires multiple single-subject studies. Horner et al. (2005) proposed that a practice may be considered evidence based when among other things, a minimum of five single-subject studies meeting acceptable criteria and documenting experimental control have been published in peer reviewed journals. Given the dearth of empirical studies on maintenance of behavior by preservice teachers, the purpose of Experiment 1 was to begin to build documentation for an evidence-based practice. In the second experiment, a generalization component was added to assess effects of deliberately programming for generalization on behavior of teachers across settings from university to school-based classrooms. The third purpose was to determine the acceptability of the intervention to participants.

The following research questions were developed based on the purpose: (a) Does training to mastery on one specific teaching behavior increase sustainability of that teaching behavior across settings, i.e., practicum in a university classroom to student teaching setting in a public school classroom? (Experiment 1), and (b) Does use of a generalization training package consisting of training to mastery plus programming for generalization increase sustainability of teaching behavior across settings, i.e., student teaching setting to K-12 classroom setting (Experiment 2)?

Experiment 1

Method

Participants

Three students, Andrew, Sierra, and Kristen (referred to hereafter as preservice teachers) who were graduate students enrolled in the special education teacher-

training program at a large Eastern University participated in the first experiment. These students were selected from an initial pool of six students based on their willingness to participate and because they were able to obtain parental permission for their students to be videotaped during lessons. The preservice teachers were told that they would be participating in a study on improving teacher preparation. Andrew and Sierra held general education teaching certificates and were working towards earning an additional certification in special education. Neither was currently teaching in a special education classroom. The third preservice teacher, Kristen, had no teaching certification but had worked as a behavior analyst in a private school for students with disabilities. Each preservice teacher successfully completed an instructional methods course in direct instruction teaching procedures where they learned how to complete learning trials using either praise for correct answers, or specific error correction procedures, prior to entering the practicum.

Setting

The experiment was conducted during the summer and fall semesters in a practicum site at the university (summer semester), followed by a student teaching placement (fall semester). Both courses, practicum and student teaching, were 14 weeks in duration. The practicum was a three credit graduate course taken at the end of 33 credits in special education coursework and was a prerequisite for student teaching. Classes met one night a week for 14 weeks in a university classroom. During each session in practicum, the student teacher was instructed to teach IEP academic goals to a student using direct instruction procedures over a 90 min period. The IEP goals selected were all in the area of reading, specifically decoding skills. Observation sessions were 20 min long.

Following successful completion of practicum requirements, the preservice teachers entered student teaching at three different locations. Andrew's placement was in a special education classroom in a vocational school, Sierra student taught in an elementary learning support classroom, and Kristen's placement was in an approved private school for students with moderate to severe disabilities.

Materials

During practicum, each session was videotaped for the purpose of data collection. The third and fourth authors provided feedback via bug-in-ear wireless technology (BIE). The BIE consisted of a portable battery-operated transmitter and receiver with a capability of transmitting over 150 feet. The first author collected data from the videotapes and the second author who was naive to the purpose of the experiment collected reliability data. During student teaching the following semester, lesson probes were audio taped by the preservice teachers for the purpose of data collection. A graduate research assistant collected reliability data. Taping was done on 20–30 min segments of regularly scheduled lessons.

Dependent Measures

There was one dependent measure in Experiment 1: Percentage of three-term contingency trials completed by the preservice teacher. Three-term contingency trials (TTC trials) are basic learning units consisting of presentation of an antecedent (first term), student response to the antecedent (second term), and feedback to the student from the teacher (third term). Completing TTC trials was selected as the targeted teaching behavior because they are strong predictors of effective instruction (Albers and Greer 1991). Feedback may take the form of praise for correct responses or error correction for incorrect responses. Praise was defined as a combination of a positive statement linked to the behavior being reinforced such as "Excellent work using a strategy to write your paragraph."

Each time the preservice teacher presented an antecedent (A) to the student it was counted as the start of a TTC trial (opportunity to respond). If the student then responded with an answer (B) and the preservice teacher provided either praise or corrected an error with the student who made the error (C), it was counted as a completed TTC trial. The following are examples of correctly completed TTC trials targeted in this study:

1. Correct student response:
   - (A) Preservice teacher: "What is the answer to 12 × 2?"
   - (B) Student: "24"
   - (C) Preservice teacher: "Correct. 12 × 2 equals 24."
2. Incorrect student response: (consists of two separate TTC trials)
   - (A) Preservice teacher: "What is the answer to 12 × 2?"
   - (B) Student: "20"
   - (C) Preservice teacher: "No. The answer for 12 × 2 is 24." (End of 1st TTC trial)
   - (A) Preservice teacher: "Again. Tell me the answer for 12 × 2." (Start of next TTC trial.)
   - (B) Student: "24"
   - (C) Preservice teacher: "Now you've got it. 24 is the correct answer."

If the preservice teacher presented an antecedent (A) to the student (i.e., opportunity to respond) and either the student responded correctly (B) but the teacher did not provide praise, or, the student responded incorrectly (B) and the teacher failed to correct the answer with the student who made the error, it was not counted as a completed TTC trial. The following are non-examples of TTC trials:

1. Correct student response:
   - (A) Preservice teacher: "What is the answer for 12 × 2?"
   - (B) Student: "24"
   - (C) Preservice teacher: Says nothing in response to student's answer and continues on to the next question, therefore does not complete the TTC trial by giving feedback to the student.

2. Incorrect student response:
   (A) Preservice teacher: "What's the answer to 12 × 2?"
   (B) Student: "20"
   (C) Preservice teacher: "Does anyone here know the answer?" (Non-example of a TTC trial because the teacher responds by repeating the question instead of correcting the error with the student who made it.)

Data Collection

Data were recorded using a pencil and paper on a simple recording sheet consisting of three columns with the headings of A (antecedent delivered by teacher), B (student response), and C (consequence by teacher). Each time the preservice teacher asked a question, the observer made a check or slash mark in the A column. (Number of checks in this column were tallied at the end of each session and counted overall as the number of opportunities to respond.) The observer also scored the number of responses by students and the number of consequences by the teacher. Completed learning trials were those trials that had all three columns in a row checked off by the observer (A, B, and C). Depending on the content of each lesson the exact number of TTC trials varied. In order to account for this variability, a percentage of TTC trials was used, whereby the total number of completed TTC trials was divided by the total number of TTC trials per session, multiplied by 100. Only the direct instruction component of the lesson was videotaped. Data were not collected during independent work time.

Social Validity

A questionnaire was emailed to the preservice teachers by the first author at the conclusion of Experiment 1 and was used to assess social validity. The questionnaire consisted of the following questions: (1) Did you like receiving feedback using wireless technology? Why/why not? (2) How did you feel about wearing the bug-in-ear device? and (3) Would you use wireless technology to give feedback in the future if you find yourself in the role of peer coach, co-operating teacher, etc.? Participants returned their completed questionnaires via email or returned hard copies to the first author's office. The first author compiled a list of comments from the questionnaires.

Design

A multiple-baseline design across participants was used to assess effects of training a specific teaching behavior, completion of TTC trials, to mastery, on maintenance and generalization. Participants, preservice teachers, were originally trained in a practicum site and maintenance and generalization of the teaching behavior were then evaluated in the next setting, student teaching.

Baseline

During baseline preservice teachers received delayed feedback from the research assistants (third and fourth authors) on completion of TTC trials following each instructional session in a post-teaching conference. Feedback included corrective feedback on incomplete learning trials (these were identified and examples of how to correctly complete the trials were described) as well as specific praise for completing trials. A university supervisor also observed the preservice teachers throughout the semester as part of course requirements but the content of the feedback did not include TTC trial completion and consisted of delayed feedback only. Each preservice teacher remained in baseline until performance was stable for the target behavior. This was determined by computing the average baseline and determining that no data point varied more than 50% from the mean (Alberto and Troutman 2002).

Prior to the intervention condition, each preservice teacher practiced receiving feedback through the BIE device by the first author while teaching his or her assigned student. The preservice teacher received short phrases of verbal feedback through an earpiece or "bug" that fit inside his or her ear. Comfort level was determined by asking the preservice teacher questions such as "How does the BIE feel in your ear? Are you distracted when you hear phrases through the BIE? Does the feedback interfere with your teaching?" Feedback on the dependent variable, TTC trials, was not given during this practice time. Each preservice teacher had the option to continue "practicing" until comfortable with the BIE or to drop out of the study if uncomfortable with the procedure. All three preservice teachers adapted quickly (under 10 min) and all opted to continue with the intervention phase of the experiment.

Intervention

Preservice teachers were observed one or two times per week during the 12-week practicum. Observation sessions were 20–30 min long. During observation sessions, preservice teachers were given BIE earpieces to wear attached to receivers which they placed in a pocket or affixed to a belt. The research assistant wore a microphone attached to an FM transmitter. The researcher instructed the preservice teacher to turn the receiver to the on position at the start of the session. Research assistants sat in the back of the room (approximately 4–7 m away), videotaped the session, and provided verbal feedback to the preservice teacher on completion of TTC trials. Feedback consisted of short phrases to either praise completion of learning trials by the preservice teacher (e.g., "Good correction," or "Good job reinforcing answers") or to remind the participants to complete TTC trials when they were not doing so (e.g., "Correct the error," or "Remember to praise"). Frequency of feedback delivered by the research assistants to the preservice teachers was not measured in this study. No other feedback, verbal or written was given and the intervention continued until the preservice teacher reached criteria for mastery, i.e., 90% completion of TTC trials over three consecutive sessions with immediate feedback. Once criterion was achieved, immediate feedback was faded in the following way. First, feedback delivered via BIE was shortened to one or two words.

In the next session, the BIE was turned off but still worn by the participants and no feedback was delivered. In the third session of fading, the BIE was removed but kept in view of the participant. For the last session of fading, the BIE was completely removed from the classroom thus ending the intervention phase of the experiment.

Maintenance

Maintenance probes were conducted the following fall semester during student teaching for each preservice teacher. Each was given a tape recorder and blank cassette tapes and was asked to tape his or herself for 20–30 min while delivering direct instruction lessons over the course of student teaching. The tape recorders were placed on a desktop during instruction and were turned on and off by the participants. Each was asked to do as many tapes as he or she could and to spread the taping out over the course of the semester. At the end of the semester the tapes were returned to the researcher and data were collected using the same procedure as with baseline and intervention. Results of data collection were not shared with the university supervisor and the researcher did not evaluate or assign a grade to any preservice teacher during either phase of Experiment 1.

Interobserver Agreement

A graduate research assistant (second author) who was naïve to the experiment's purpose independently collected agreement data from videotapes (baseline and intervention) and audiotapes (maintenance). The first and second authors watched 15 min timed lesson segments. Using the data collection instrument previously described, each observer simultaneously but independently coded data by counting the total number of completed learning trials (consisting of an antecedent by the teacher, response by the student, and feedback in the form of praise or error correction by the teacher). Total count interobserver agreement was calculated by dividing the smaller count by the larger count and multiplying by 100 to obtain a percentage of agreement.

Prior to collecting interobserver agreement data, training was conducted consisting of instructions and corrective feedback using videotapes of the participants during two different training sessions. Training continued until the first author and research assistant reached 90% total count interobserver agreement on the occurrence of completed TTC trials on three 10 min samples of observations.

Baseline agreement data were collected for 33% of the sessions for Andrew, Sierra, and Kristen. Mean agreement was 100% across preservice teachers for completion of TTC trials during baseline. Intervention agreement data were collected for 29% of the sessions for Andrew, 25% for Sierra, and 22% for Kristen. Mean agreement during intervention was 93% across all preservice teachers for completion of TTC trials. Agreement data were collected for 25% of the maintenance sessions for Andrew, and 33% for Sierra and Kristen with 90% mean agreement for completed three-term contingency trials.

The first author collected treatment integrity data by observing the third and fourth authors during treatment sessions using a checklist of the procedures. The checklist

consisted of such items as: (a) The participant wore the BIE device in the "on" position; (b) The researcher gave immediate feedback (within 3 s) to the participant; (c) Feedback consisted of short phrases and was specific to TTC trial completion; and (d) The session was videotaped. Treatment integrity data were collected for 25% of the sessions with procedures being followed correctly 100% of the time.

Results

The following research question was addressed in Experiment 1: Does training to mastery on one specific teaching behavior (TTC trials) increase sustainability of that teaching behavior across settings, i.e., practicum in a university classroom to student teaching in a public school classroom? Figure 2 represents the percentage of completion of TTC trials by preservice teachers in practicum (baseline and intervention) and student teaching (maintenance). Percentage of TTC trials completed in baseline ranged from 1 to 86 (Andrew, $M = 1$; Sierra, $M = 40$, range = 26–55; Kristen, $M = 69.5$, range = 50–86). All three teachers reached criterion of 90% over three consecutive trials on the dependent variable, completion of TTC trials during intervention (practicum setting). Percentage of TTC trials completed in intervention ranged from 41 to 100 (Andrew, $M = 87.5$, range = 61–100; Sierra, $M = 83.7$, range = 41–100; Kristen, $M = 94$, range = 83–100). During maintenance phase (student teaching setting) percentage of completed TTC trials ranged from 55 to 95 (Andrew, $M = 66.5$, range = 62–78; Sierra, $M = 55$, range = 55–79; Kristen, $M = 76.3$, range = 65–95).

All three teachers responded to a questionnaire at the conclusion of Experiment 1 to assess social validity. The questionnaire consisted of the following questions: (1) Did you like receiving feedback using BIE technology? Why/why not?, (2) How did you feel about wearing the bug-in-ear device?, and (3) Would you use BIE technology to give feedback in the future if you find yourself in the role of peer coach, co-operating teacher, etc.? All three teachers reported that they liked receiving immediate feedback via bug-in-ear technology. They reported that it was not intrusive, easy to adjust to, and helpful with increasing new teaching behavior. One teacher reported she would consider using the bug-in-ear with her own students and two reported that they would consider using it with student teachers if/when they are in the role of co-operating teacher.

Discussion

Results of Experiment 1 suggest that immediate feedback delivered via BIE technology was effective in increasing completion of TTC trials to an established mastery level of 90% over three consecutive trials. This is not surprising since immediate feedback has been shown to produce higher, steadier rates of responding by teachers and is more effective and efficient in changing teacher behavior than delayed feedback (Coulter and Grossen 1997; O'Reilly et al. 1994; Scheeler, et al. 2006). However, the teaching behavior was not maintained at criterion level and in fact deteriorated by approximately 23% overall from just practicum to student teaching suggesting that training to mastery on specific teaching skills is not enough

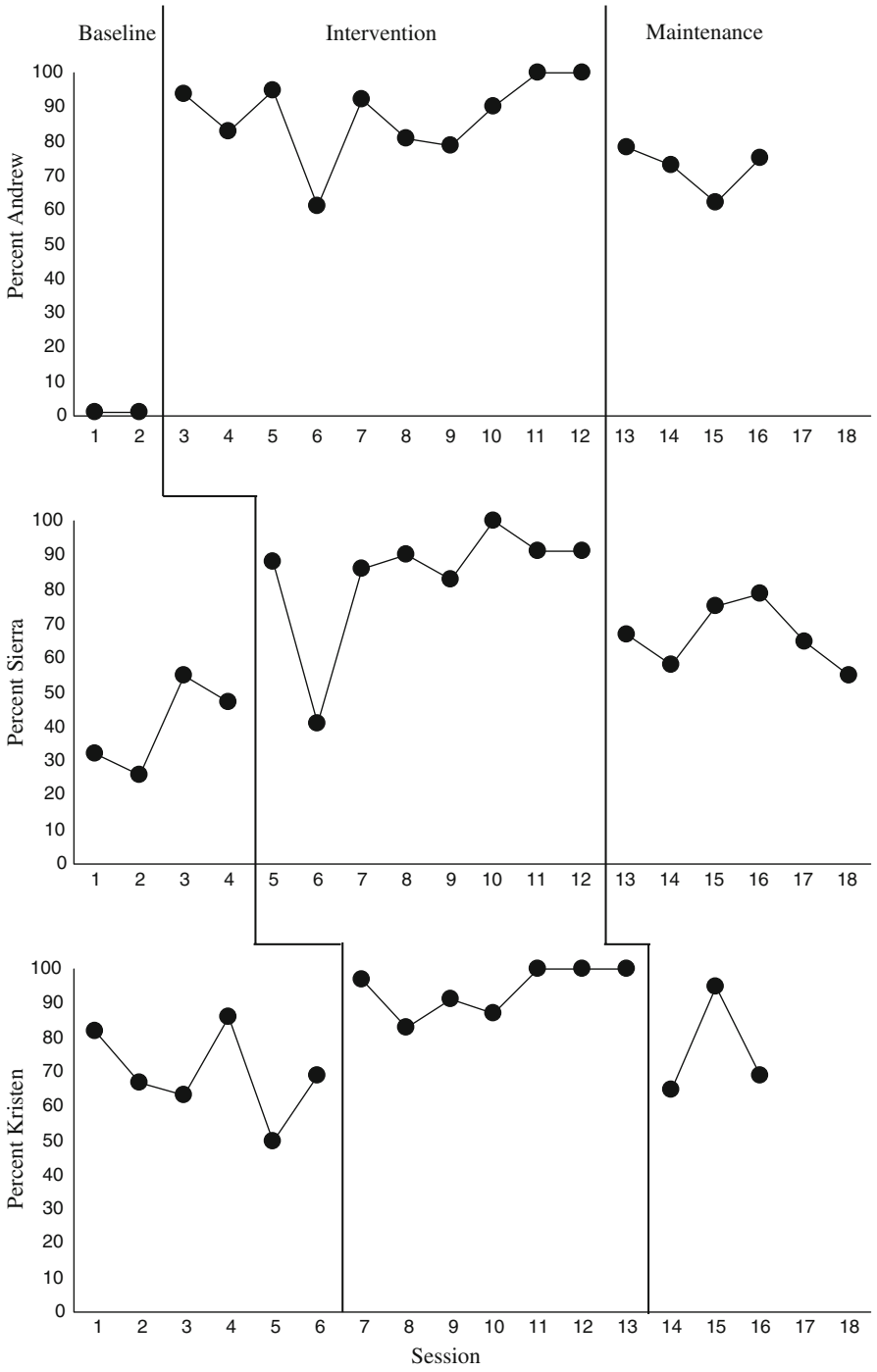

Fig. 2 Percent of completed TTC trials per lesson in Experiment 1

to sustain them in the next setting. Preservice teachers may learn effective teaching techniques but if they do not maintain their newly acquired skills generality of these skills to the next setting is in jeopardy. In this experiment the preservice teachers had not even entered their own classrooms and were already losing skills while still in a supervised setting (student teaching). Baer, Wolf, and Risley (1968) cautioned that generality must be programmed, rather than expected. It is not enough to train and hope that the new learning continues. In her model for promoting generalization and maintenance of effective teaching skills (see Fig. 1; Scheeler 2008) suggests programming for generalization as the third step in the sequence, following using immediate feedback to promote acquisition of teaching skills and training to mastery. As evidenced in Experiment 1, completion of the first two steps alone in the model was insufficient to sustain and generalize newly acquired teaching skills. In Experiment 2, the third step of the model was added to program generalization of teaching skills from student teaching settings to teachers' own classrooms as in-service teachers.

Experiment 2

Method

Participants

Two student teachers, Ali and Emily (referred to hereafter as teachers) who were undergraduate seniors enrolled in the special education teacher-training program at a large Eastern University participated in the second experiment. The first phase was conducted in the spring semester student teaching placement which was 15 weeks long and the second phase was during the following September through December in the teacher's own classroom. Both teachers expressed a willingness to participate and were therefore selected on a volunteer basis. During their junior year each successfully completed a practicum using direct instruction teaching procedures. Ali and Emily were selected from a pool of nine student teachers who originally expressed interest. Out of these, six dropped out prior to the start of the experiment citing that they did not want to be videotaped and/or have anything extra to manage during student teaching. One other student agreed to participate and did so in the first phase of the experiment (student teaching) but dropped out prior to the start of the second phase (own classroom) citing that she was having difficulty managing student behavior in her classroom and did not want the added task of audio taping her lessons.

Setting

Both teachers were completing their student teaching in schools in a large urban school district in southeastern Pennsylvania. Ali's student teaching placement was in a life skills class in a middle school. Emily's placement was in a fifth grade learning support classroom in an elementary school. Specific lessons were academic in nature including direct instruction reading (Emily), and calendar skills (Ali). All

lessons took place in the teacher's own classroom within the regularly scheduled instructional time. Lessons were videotaped for the purpose of data collection by a graduate research assistant. Observation sessions were 20 min long.

Following the end of the school year and summer break, data collection resumed the following September, this time in the teachers' own classrooms. Ali moved to the west coast and was now teaching in a middle school life skills class and Emily was teaching an autistic support class in an elementary school in the same school district in which she student taught. Teachers were asked to periodically record data via audio taping of 20 min segments of regularly scheduled lessons.

Materials

The same materials as described in Experiment 1 (BIE device, video cameras, and tape recorders) were used in Experiment 2.

Dependent Measure

The dependent measure, percentage of TTC trials completed by the student teacher, was the same as used in Experiment 1. Data were collected using the same data recording instrument as described in Experiment 1.

Social Validity

A questionnaire similar to the one used in Experiment 1 was administered to the student teachers to assess social validity. Questionnaires in both studies were written assessments, completed individually by participants in their own classrooms.

Design

In Experiment 2, a multiple-baseline design across participants was used to assess effects of a maintenance and generalization training package which included: a) Training to mastery level on completion of TTC trials during student teaching and b) developing a plan for generalization to the next setting, i.e., the teachers' own classrooms, post graduation from the university.

Baseline

Procedures in baseline were identical to those in Experiment 1 with the exception that this condition took place while the participants were student teaching in special education classrooms in public school settings.

Intervention

Prior to starting the intervention phase the student teachers received training in using the BIE to receive feedback, similar to Experiment 1 but instead of teaching one student while practicing, they taught a "mock" lesson to the graduate assistant

while the researcher provided feedback. As with the preservice teachers in Experiment 1, the student teachers felt comfortable enough to proceed with intervention after approximately 10 min of training. Student teachers were observed one or two times per week during the 14-week field experience. A research assistant followed the same procedure as in Experiment 1, videotaping the lesson and giving immediate feedback via BIE. Once criteria for mastery of 90% over three consecutive sessions was reached the intervention was faded by first turning off the BIE but still wearing the earpiece for one session, next removing the BIE but keeping it in view while teaching a lesson and finally, completely removing the BIE from view during instruction. No feedback was given during fading of the intervention.

Maintenance and Generalization

At the conclusion of the semester the student teachers met for approximately 45 min with the first author to review techniques for generalization that they had previously been taught in a special education course prior to student teaching. The techniques were taken from the Alberto and Troutman (2002) text. They were presented with a handout listing the techniques and the first author described each and asked the student teachers to give examples. The first author then asked the student teachers to focus on the technique, programming common stimuli, which could be done by introducing elements of the training session (student teaching) into the natural environment (teacher's own classroom) to serve as prompts for the appropriate behavior (TTC trial completion). Ways to do this were "brainstormed" and finally, the student teachers were asked to develop a plan to use program common stimuli to help them to generalize TTC trial completion to their own classrooms in the start of the next school year. (Both student teachers had already secured teaching positions for the fall prior to completion of their student teaching.) Ali's plan for programming common stimuli involved bringing materials she used in her lessons to teach calendar skills while student teaching to her own classroom where she would teach the same thing to the same age/level of students. Emily decided she would have the BIE earpiece (each student teacher had their own), in view in her own classroom while teaching to remind her to complete TTC trials. Maintenance probes were done the same way as in Experiment 1 with each teacher receiving a tape recorder and tapes and asked to audio tape as many lessons as she could, over the course of the following school year. One difference between Experiments 1 and 2 was a break of approximately 4 months between the end of intervention and probes in the next setting and probes were spread out throughout the school year, not just one semester.

Interobserver Agreement

Agreement data were collected using the same procedure as in Experiment 1 by the same graduate research assistant. Data were independently collected from videotapes (baseline and intervention) and audiotapes (generalization phase). Baseline agreement data were collected for 40% of the sessions for Ali and 43% for

Emily. Mean agreement was 96% across both teachers for completion of TTC trials. Intervention agreement data were collected for 30% of the sessions for Ali and 23% for Emily. Mean agreement during intervention was 95% for both teachers for completion of TTC trials. Agreement data were collected for 33% of the maintenance sessions for Ali and Emily with 96% mean agreement for completed three-term contingency trials. The first author collected fidelity of treatment data using a checklist of procedures for 20% of sessions. Treatment integrity was 100% for procedures.

Results

The following research question was addressed in Experiment 2: Does use of a generalization training package consisting of training to mastery plus training for generalization increase sustainability of teaching behavior across settings, i.e., student teaching setting to public school classroom setting, post-graduation? Figure 3 represents the percentage of completion of TTC trials by preservice teachers in student teaching (baseline and intervention) and their own classrooms as in-service teachers (generalization). Percentage of TTC trials completed in baseline ranged from 29 to 63 (Ali, $M = 54.2$, range = 44–63; Emily, $M = 37.6$, range = 29–60). Both teachers reached criterion of 90% over three consecutive trials on the dependent variable, completion of TTC trials during intervention. Percentage of TTC trials completed in intervention ranged from 35 to 100 (Ali, $M = 74.5$, range = 35–92; Emily, $M = 85.4$, range = 69–100). During generalization phase (own classroom setting) percentage of completed TTC trials ranged from 86 to 96 (Ali, $M = 87$, range = 86–88; Emily, $M = 93.4$, range = 87–96).

Given the small number of data points for Ali in maintenance, (e.g., 2 overall compared with 12 data points in treatment) and Emily (e.g., 5 data points in maintenance compared with 15 data points in treatment phase), a more appropriate comparison would be with the final five data points in treatment for both participants. Interpreting data based on the last five data points during treatment is possible because the treatment data represent an acquisition process and the last five data points are stable, suggesting acquisition has occurred and in fact, mastery was achieved (90% completion over three consecutive trials.) These data result in ($M = 87$, range = 79–92) for Ali, and ($M = 90.6$, range = 83–100) for Emily. Comparing the latter data points in intervention with maintenance data (Ali, $M = 87$, range = 86–88; Emily, $M = 93.4$, range = 87–96) tempers the overall conclusions of the experiment. In fact, using this comparison of the last five data points, the mean for Ali is the same in both intervention and maintenance but still somewhat higher for Emily (93.4 in maintenance compared with 90.6 in intervention).

Social Validity

A questionnaire similar to the one used in Experiment 1 was administered to the student teachers to assess social validity. Both teachers reported that they liked receiving feedback from the BIE device because they felt that it helped to improve

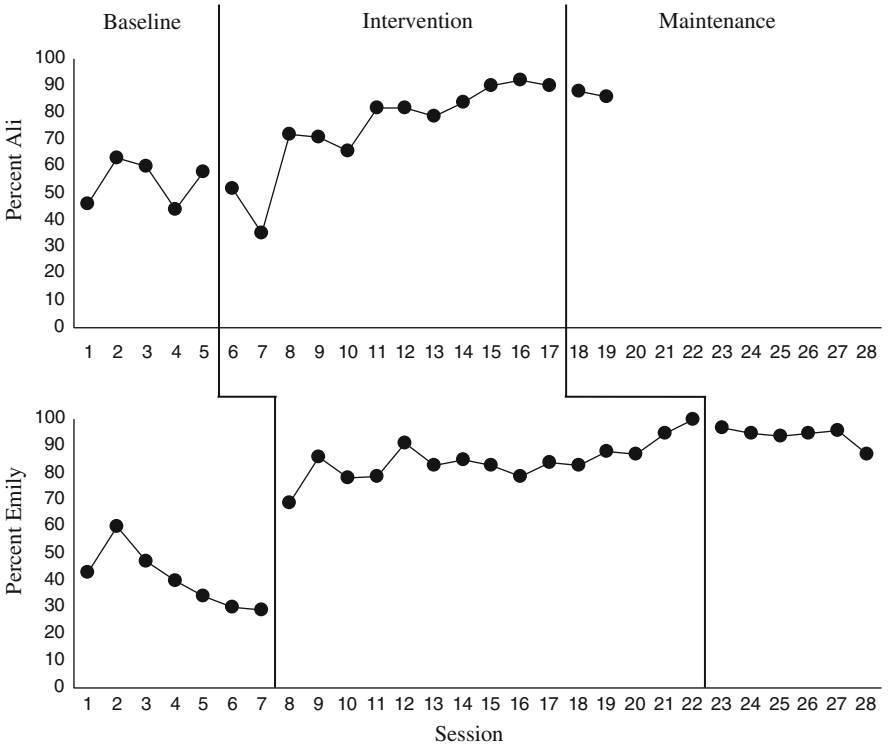

Fig. 3 Percent of completed TTC trials per lesson in Experiment 2

their skills in correcting students and giving feedback. One teacher stated that she didn't mind wearing the device but got "tired of it" as time went on, although she perceived it as helping her to improve her teaching. Both teachers also reported they would use the BIE technology with student teachers. One teacher added that she would also conference with a student teacher afterwards to add more in-depth explanations on areas in which a student teacher could work on.

Discussion

Results of Experiment 2 are preliminary given that there are only two participants and the first participant, Ali had only two data points in maintenance. Using the final five data points to more appropriately compare results from intervention with maintenance yields the same results for Ali in both phases and slightly higher results for Emily. Still, this is an encouraging finding because there is an obvious need to prepare teachers to generalize newly acquired skills across settings—from university classrooms and supervised field experiences to K-12 classroom settings (Boudah et al. 2001; Bowles and Nelson 1976; Engelmann 1988; Gersten et al. 1995, 2000; Gersten and Dimino 2001; Greenwood and Abbot 2001; Han and Weiss 2005; Horton 1975; Leach and Conto 1999; McLeskey and Waldron 2004; Noell

et al. 1997; Robinson and Swanton 1980; Rose and Church 1998; Scruggs and Mastropieri 1994; Vaughn et al. 2000) and the intervention used in Experiment 2 is one way to fill this need. Without some type of plan for this, newly acquired teaching skills will deteriorate when teachers transfer from training settings to actual teaching settings to less than 30% of what was practiced (Englemann). Noell et al. found deterioration of teacher implementation of a reinforcement based strategy to occur after just 4 days of beginning implementation. Han and Weiss found implementation rates of school-wide initiatives dropped off by as much as 60% one year after teachers were trained in program components and concluded that training alone is not enough to sustain new learning. The intervention described in Experiment 2, combining a package of immediate feedback for efficient acquisition of teaching techniques, training to mastery on a specific technique, and following up with a plan to promote generalization, appears to be effective in addressing the problem of maintaining and generalizing new learning to school settings. Our highly trained preservice teachers are then better prepared to continue using the skills and techniques they learn when they leave the university training sites and move on to their own classrooms.

General Discussion

Results of this study suggest that using immediate feedback to promote acquisition of evidence-based teaching skills is an effective and efficient technique for teacher educators to use. Each teacher involved in the study increased completion of TTC trials, the targeted behavior, and reached criterion. However, use of the targeted teaching technique began to decline once the intervention had been faded when no other support was given. It was not enough to just train to mastery level on a teaching technique in order to sustain it, so in Experiment 2 an additional step was added, i.e., programming for generalization. Although the results of Experiment 2 are encouraging the findings are preliminary in that there were only two participants involved and the first participant, Ali, had just two data points in the maintenance phase of the experiment. Nonetheless, results were maintained not just across settings (student teaching site to teachers' classrooms), but across time (4 months—May through August).

Implications for Practice

Vaughn et al. (2000) make a point about the importance of sustaining evidence-based practices when they describe the difference between a 54 year old male heart patient who does not sustain research-based practices his physician has asked him to implement and the fourth-grade teacher who after an extensive professional development program ceased to use the practices in the following year. The distinction they make is that the 54 year old heart patient bears the consequence of failure to sustain the practice. In the case of the teacher, however, her students bear the consequence of failure to sustain evidence-based practices. Vaughn et al. go on to say that change is exceedingly hard especially when the person who has the most

to gain from the change (student) is not the person who is asked to do it (teacher). If teachers experience difficulty changing behavior when they are inservice teachers in school settings, it is essential to make sure that they continue to use the evidence-based techniques they learn in university classrooms in the next setting in order to decrease the need to change. In order to do this, teacher educators may need to make curriculum modifications to ensure that programming for generalization is included in coursework and fieldwork. However, a positive effect of this modification could be that teacher educators will be modeling an evidence-based practice for the preservice teachers to use with their own students who, according to Vaughn et al. have the most to gain from the change.

A second implication of this study is that teacher educators and school district personnel should be encouraged to collaborate more so there can be a smooth transition from one setting to the next. If teacher educators are aware of practices used and valued in schools, they can use this knowledge in their teaching through examples, case studies, etc. If school administrators are aware of specific evidence based practices that are being taught in university settings, they are in a better position to provide in-class performance feedback to sustain new learning. A seamless continuation of evidence-based teaching techniques from one setting to the next should result in positive consequences for students, teachers, and school administrators.

Limitations and Implications for Future Research

There are several limitations to the study that are important to note. First, data were gathered in maintenance phase (Experiment 1) and generalization phase (Experiment 2) by having teachers audio or video tape themselves and subsequently return the tapes to the first author, therefore self-selection and possible editing of the tapes could have occurred. However, teachers in Experiment 1 were naive to the outcome other than receiving feedback on a specific behavior (completion of TTC trials). They were never told what the dependent variable was during the experiment. Teachers in Experiment 2 were asked to continue with the behavior (completion of TTC trials) at the end of the intervention phase and were asked to develop a way to program common stimuli. However, there was a break over the summer months before the school year started because they did student teaching during spring semester. One would expect that this natural break in time (approximately 4 months) would contribute to the deterioration of new learning, so the fact that both teachers continued to complete TTC trials at a high level is impressive especially in light of Noell et al. 1997, who found deterioration in teacher implementation of new learning to occur in just 4 days. Future researchers might want to consider finding another way to unobtrusively yet directly observe teachers in the generalized setting to assess maintenance and generalization. Another option could be to use technology, e.g. webcams to observe teachers from a distance. Use of technology could also help with the second limitation in the experiment, i.e., collecting more data in the generalized setting. In Experiment 2, we were only able to collect 2 data points for Ali and 5 data points for Emily. It would be beneficial to replicate this study with more teachers and more data in the generalization phase in

order to support the intervention as a promising practice. Again, webcam technology could provide a reasonable cost and time efficient way to gather data without having to travel to the actual school settings. In Experiment 2, one of the inservice classrooms was on the opposite coast of the United States making frequent travel for the purpose of data collection improbable if not impossible.

A third limitation of the experiments described herein is that there may be more factors contributing to lack of generalization of teaching techniques that are not accounted for. Some examples might be school climate, administrative mandates, and lack of training in evidence-based practices, to name a few possibilities. Future researchers might want to identify these or other contributing factors and develop ways to address them.

A fourth limitation pertaining to Experiment 2 is that programming common stimuli was the only generalization technique used although five techniques are listed in the model for promoting generalization and maintenance of effective teaching skills (e.g., sequential modification, train sufficient exemplars, introduce to natural maintaining contingencies, and mediate generalization). It would be interesting in future research to determine the effectiveness and social validity of each of these techniques in teacher preparation.

A final limitation is that there were only two participants in Experiment 2. Horner et al. (2005) recommend one demonstration and two replications to be considered a sound experimental design. Experiment 2 had just one demonstration and one replication. Results of Experiment 2 must therefore be considered preliminary and future replication should involve at least three participants in the multiple baseline design.

In addition to the suggestions for future research that have been already stated there are several additional areas that future researchers might address. One is that the last step in the model for promoting generalization and maintenance has not been assessed within the sequence of steps. It would be beneficial to determine the impact adding this step has to an already powerful intervention. It may also be beneficial to assess the frequency of immediate feedback delivered to teachers as this was not measured in the current study. Finally, future researchers might want to look at adding a self-management component to the model, particularly during the last step. In addition to support from others, self-management of teaching techniques might have long-term, long-lasting effects on teacher behavior.

Teacher educators should be commended for their work in preparing the next generation of teachers to use evidence-based practices. The challenges faced by teacher educators are enormous, as are the responsibilities of making sure preservice teachers know what to teach, and how to teach it when they enter the school setting. Countless hours are spent by faculty and students to this end. However, what students learn in university classrooms is often forgotten or just not used once they obtain teaching positions and become in-service teachers. Preservice teachers may begin to lose newly learned techniques before they even leave the university, as seen in Experiment 1, when teachers moved from practicum to student teaching. One can argue that there is a missing link in teacher preparation, i.e., actively programming for generalization to the next setting. If this is done in a deliberate, systematic way we will eliminate the ineffective "train and hope" technique so prevalent in

institutions of higher learning. We have been lamenting the lack of generalization for 40 years (see Baer et al. 1968). The time for doing something about it is now.

References

Albers, A. E., & Greer, R. D. (1991). Is the three-term contingency trial a predictor of effective instruction? *Journal of Behavioral Education, 1*, 337–354. doi:10.1007/BF00947188.

Alberto, A., & Troutman, P. (2002). *Applied behavior analysis for teachers* (5th ed.). New York: Merrill Publishing Co.

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. doi:10.1901/jaba.1968.1-91.

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*, 290–303.

Bowles, P. E., & Nelson, R. O. (1976). Training teachers as mediators: Efficacy of a workshop versus the bug-in-the-ear technique. *Journal of School Psychology, 14*, 15–26. doi:10.1016/0022-4405(76)90058-3.

Brophy, J. E., & Good, T. L. (1986). Teacher behavior and student achievement. In M. L. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed.). Englewood Cliffs, NJ: Prentice-Hall.

Christenson, S. L., Ysseldyke, J. E., & Thurlow, M. L. (1989). Critical instructional factors for students with mild handicaps: An integrative review. *Remedial and Special Education, 10*, 39–48. doi:10.1177/074193258901000505.

Coulter, G. A., & Grossen, B. (1997). The effectiveness of in-class instructive feedback versus after-class instructive feedback for teachers learning direct instruction teaching behaviors. *Effective School Practices, 16*, 21–35.

Ellis, E. S., Worthington, L. A., & Larkin, M. J. (1994). *Executive summary of research synthesis on effective teaching principles and the design of quality tools for educators* (Tech. Rep. No. 6). University of Oregon, National Center to Improve the Tools of Educators.

Engelmann, S. (1988). The logic and facts of effective supervision. *Education & Treatment of Children, 11*, 328–340.

Epps, S., & Lane, M. P. (1987). Assessment and training of teacher interviewing skills to program common stimuli between special and general education environments. *School Psychology Review, 16*, 50–68.

Gersten, R., Chard, D., & Baker, S. (2000). Factors enhancing sustained use of research-based instructional practices. *Journal of Learning Disabilities, 33*, 445–457. doi:10.1177/002221940003300505.

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

Gersten, R., Morvant, M., & Brengleman, S. (1995). Close to the classroom is close to the bone: Coaching as a means to translate research into classroom practice. *Exceptional Children, 62*, 52–66.

Gersten, R., Vaughn, S., Deshler, D., & Schiller, E. (1997). What we know about using research findings: Implications for improving special education practice. *Journal of Learning Disabilities, 30*, 446–476. doi:10.1177/002221949703000501.

Greenwood, C. R., & Abbot, M. (2001). The research to practice gap in special education. *Teacher Education and Special Education, 24*, 276–289.

Han, S. S., & Weiss, B. (2005). Sustainability of teacher implementation of school-based mental health programs. *Journal of Abnormal Child Psychology, 33*, 665–679. doi:10.1007/s10802-005-7646-2.

Horner, R. H., Carr, E. G., Halle, J., McGee, G., Odom, S., & Wolery, M. (2005). The use of single subject research to identify evidence-based practice in special education. *Exceptional Children, 71*, 165–179.

Horton, G. O. (1975). Generalization of teacher behavior as a function of subject matter specific discrimination training. *Journal of Applied Behavior Analysis, 8*, 311–319. doi:10.1901/jaba.1975.8-311.

Leach, D. J., & Conto, H. (1999). The additional effects of process and outcome feedback following brief inservice teacher training. *Educational Psychology, 19*, 441–462. doi:10.1080/0144341990190405.

Lindsley, O. R. (1992). Why aren't effective teaching tools widely adopted? *Journal of Applied Behavior Analysis, 25*, 21–26. doi:10.1901/jaba.1992.25-21.

McLeskey, J., & Waldron, N. C. (2004). Three conceptions of teacher learning: Exploring the relationship between knowledge and the practice of teaching. *Teacher Education and Special Education, 27*, 3–14.

Noell, G. H., Witt, J. C., Gilbertson, D. N., Ranier, D. D., & Freeland, J. T. (1997). Increasing teacher intervention implementation in general education settings through consultation and performance feedback. *School Psychology Quarterly, 12*, 77–88. doi:10.1037/h0088949.

O'Reilly, M. F., Renzaglia, A., Hutchins, M., Koterba-Bass, L., Clayton, M., Halle, J. W., et al. (1992). Teaching systematic instruction competencies to special education student teachers: An applied behavioral supervision model. *Journal of the Association for Persons with Severe Handicaps, 17*, 104–111.

O'Reilly, M. F., Renzaglia, A., & Lee, S. (1994). An analysis of acquisition, generalization and maintenance of systematic instruction competencies by preservice teachers using behavioral supervision techniques. *Education and Training in Mental Retardation and Developmental Disabilities, 29*, 22–33.

Prater, M. A., & Sileo, T. W. (2004). Fieldwork requirements in special education preparation: A national study. *Teacher Education and Special Education, 27*, 251–263.

Robinson, V., & Swanton, C. (1980). The generalization of behavioral teacher training. *Review of Educational Research, 3*, 486–498.

Rose, D. J., & Church, R. J. (1998). Learning to teach: The acquisition and maintenance of teaching skills. *Journal of Behavioral Education, 8*, 5–35. doi:10.1023/A:1022860606825.

Rosenshine, B., & Stevens, R. (1986). Teaching functions. In M. C. Wittrock (Ed.), *Handbook of research on teaching* (3rd ed., pp. 376–391). New York: Macmillan.

Rutherford, R. B., & Nelson, C. M. (1988). Generalization and maintenance of treatment effects. In J. Witt, S. N. Flintt, & F. M. Gresham (Eds.), *Handbook of behavior therapy in education*. New York: Plenum.

Scheeler, M. C. (2008). Generalizing effective teaching skills: The missing link in teacher preparation. *Journal of Behavioral Education, 17*, 145–159. doi:10.1007/s10864-007-9051-0.

Scheeler, M. C., & Lee, D. L. (2002). Using technology to deliver immediate corrective feedback to preservice teachers. *Journal of Behavioral Education, 11*, 231–241. doi:10.1023/A:1021158805714.

Scheeler, M. C., Macluckie, M., & Albright, K. E. (2008). Effects of immediate feedback delivered by peer tutors on oral presentation skills of adolescents with learning disabilities. *Remedial and Special Education*, 0741932508327458v1.

Scheeler, M. C., McAfee, J. K., Ruhl, K. L., & Lee, D. L. (2006). Effects of corrective feedback delivered via wireless technology on preservice teacher performance and student behavior. *Teacher Education and Special Education, 29*, 12–25.

Scruggs, T. E., & Mastropieri, M. A. (1994). The effects of generalization training: A quantitative synthesis of single subject research. *Advances in Learning and Behavioral Disabilities, 8*, 259–280.

Sharpe, T., Lounsbery, M., & Bahls, V. (1997). Description and effects of sequential behavior practice in teacher education. *Research Quarterly for Exercise and Sport, 68*, 222–232.

Stein, M., Carnine, D., & Dixon, R. (1998). Direct instruction: Integrating curriculum design and effective teaching practice. *Intervention in School and Clinic, 33*, 227–234. doi:10.1177/105345129803300405.

Stokes, T. F., & Baer, D. M. (1977). An implicit technology of generalization. *Journal of Applied Behavior Analysis, 10*, 349–367. doi:10.1901/jaba.1977.10-349.

Van Houten, R. (1980). *Learning through feedback*. New York: Human Sciences Press, Inc.

Vaughn, S., Klingner, J., & Hughes, M. (2000). Sustainability of research-based practices. *Exceptional Children, 66*, 163–171.

Wittrock, M. C. (Ed.). (1986). *Handbook of research on teaching* (3rd ed.). New York: Macmillan.