Training via Practitioner Journal Articles:

A Pathway to Increasing Teachers’ Procedural Integrity

DISSERTATION

Presented in Partial Fulfillment of the Requirements for the Degree Doctor of Philosophy

in the Graduate School of The Ohio State University

By

Ziwei Xu

Graduate Program in Educational Studies

The Ohio State University

2016

Dissertation Committee:

Dr. Nancy Neef, Advisor

Dr. Moira Konrad

Dr. Laurice Joseph
Abstract

This dissertation consists of an introduction to the dissertation and a statement of problem (Chapter 1), a comprehensive literature review exploring strategies for training teachers to conduct behavior analytic interventions in school settings (Chapter 2), two single-case studies evaluating the effectiveness of using published practitioner journal articles for training special education teachers on preference assessments (Chapter 3) and the system of least prompts (Chapter 4), and a general discussion summarizing the findings and describing future directions for research (Chapter 5).

In Study 1 (Chapter 3), the results of a multiple probe design across participants showed that performance of all 6 participants during simulated (role-played) sessions increased to mastery criterion levels with both preference assessment procedures (paired-stimulus and multiple-stimulus without replacement) following the introduction of a published article. Acquired assessment skills generalized to actual assessment situations involving students with and without disabilities. In Study 2 (Chapter 4), I conducted a replication of Study 1 on a different behavior analytic procedure, the system of least prompts. The results showed that a published article improved performance for all 6 participants, although 2 of 6 needed self-monitoring to reach criterion. Findings are discussed with respect to the strategic use of published practitioner journal articles in staff training.
Acknowledgments

This project would not be possible without the help and support of many talented individuals. First, I would like to thank the teachers and teaching assistants who participated in the research project; they welcomed me into their classrooms, shared their teaching experiences, and inspired my research. I am humbled by their authentic and tenacious pursuit of evidence-based practices for students with special needs. I hope my work can, in some small way, support them in that pursuit.

I would also like to express my special thanks to my advisor Dr. Nancy Neef. You have been a tremendous mentor for me. Thank you for supporting my personal and professional dreams along the way and for allowing me to grow as a teacher, a researcher, a scientist, and a person.

To my dissertation committee members Dr. Nancy Neef, Dr. Moira Konrad, and Dr. Laurice Joseph: thank you for your commitment and support during the dissertation process.

A special thanks to my friends and family. To Marnie and Gleides: thank you for your friendship, mentorship, guidance, and wisdom during my graduate studies. You have encouraged my growth and challenged me to go beyond my comfort zone. To Elle: thank you for collecting data for this research project and allowing me to grow as a supervisor. To my family: words cannot express how grateful I am for all of the
sacrifices that you have made on my behalf, and your love and support was what made me go this far. At the end, a big thank you to Beichen “Danny,” who was always my support in the moments when there was no one to answer my queries.
Vita

2011.................................B.A. English, Shanghai Ocean University, Shanghai, China

2012.................................M.Ed. Special Education, The University of Texas, Austin, Texas

2012 to 2013.........................Behavior Therapist, The Magnolia School, Austin, Texas

2013 to 2014.........................Graduate Fellow, Department of Educational Studies, The Ohio State University, Columbus, Ohio

2014 to present......................Graduate Teaching Associate, Department of Educational Studies, The Ohio State University, Columbus, Ohio
Publications


Field of Study

Major Field: Educational Studies

Area of Emphasis: Special Education & Applied Behavior Analysis
Table of Contents

Abstract ........................................................................................................................................ ii
Acknowledgments .................................................................................................................. iii
Vita ................................................................................................................................................ v
List of Tables .......................................................................................................................... x
List of Figures ........................................................................................................................... xi
Chapter 1: Introduction .......................................................................................................... 1
Chapter 2: Literature Review ................................................................................................. 9
    Method ........................................................................................................................................ 9
    Results ......................................................................................................................................... 15
    Discussion .................................................................................................................................... 27
Chapter 3: Single-Case Study One ......................................................................................... 34
    Method ......................................................................................................................................... 40
    Results and Discussion .......................................................................................................... 49
Chapter 4: Single-Case Study Two ......................................................................................... 57
    Method ......................................................................................................................................... 60
Appendix P: A Cover Page with Reading Log—System of Least Prompts .................. 133
Appendix Q: System of Least Prompts Data Sheet for Participant Use ...................... 135
Appendix R: A Cover Page with Reading Log—Self-Monitoring............................. 137
Appendix S: Integrity Checklists for Training System of Least Prompts ...................... 139
Appendix T: Social Validity Questionnaire for Participants—System of Least Prompts .................................................................................................................. 142
Appendix U: Social Validity Questionnaire for the Director—System of Least Prompts .................................................................................................................. 146
List of Tables

Table 1. Literature review participant demographics...........................................16
Table 2. Descriptions of professional development packages................................20
Table 3. Social validity results from the participants—preference assessments........54
Table 4. Social validity results from the director—preference assessments..............55
Table 5. Social validity results from the participants—system of least prompts.........73
Table 6. Social validity results from the director—system of least prompts.............74
List of Figures

Figure 1. PS assessment implementation graphs ................................................................. 51
Figure 2. MSWO assessment implementation graphs .......................................................... 52
Figure 3. System of least prompts implementation graphs ................................................. 70
Chapter 1: Introduction

Procedural integrity, often used interchangeably with intervention adherence and treatment fidelity, refers to “the extent to which essential intervention components are delivered in a comprehensive and consistent manner by an interventionist trained to deliver the intervention” (Sanetti & Kratochwill, 2009a, p. 448). In the discipline of applied behavior analysis, assessing and maximizing procedural integrity is critical for researchers and practitioners to make data-based decisions on the functional relation between the behavioral interventions and the changes in individuals’ behavioral outcomes (Barnett et al., 2014; Gresham, 2009).

Procedural Integrity as a Methodological Concern

In behavioral analytic literature, procedural integrity is a quality indicator for single-case research studies (Horner et al., 2005; Kratochwill et al., 2010); failing to consider the accuracy of intervention implementation may threaten the internal validity of experimental designs and external validity of research findings. The goal of behavior analytic research is to determine if changes in the dependent variable (individuals’ behavioral outcomes) are due to the implementation of the independent variable (behavioral interventions). Only when single-case researchers demonstrate that the intervention has been implemented as planned without deviations can the impact of the
intervention on the behavioral outcomes be determined (Gresham, MacMillan, Beebe-Frankenberger, & Bocian, 2000). Without accurate implementation, single-case researchers may run the risk of making false-positive or false-negative conclusions about the effect of the intervention, that is, improperly indicating presence of intervention effectiveness or no presence of intervention effectiveness (Peterson, Horner, & Wonderlich, 1982).

One dimension of applied behavior analysis and the quality standards of single-case design directly address the need to conduct formative assessment of procedural integrity (Baer, Wolf, & Risley, 1968; Horner et al., 2005; Kratochwill et al., 2010). Baer et al. (1968) described seven critical dimensions that collectively serve as the primary criteria for defining and judging the value of applied behavior analysis, among which the “technological” dimension is particularly relevant to procedural integrity. As Baer and his colleagues suggested, the behavioral intervention should be written clearly, explicitly, and with replicable precision. In addition, Horner et al. (2005) incorporated this tenet into the quality standards of single-case design (the experimental design used by most behavior analytic researchers). They recommended that researchers describe the independent variable (the intervention) with “replicable precision” and use “overt measurement of the integrity of implementation for the IV” (p. 174). In a behavior analytic single-case experiment, the intervention is applied repeatedly over time, and therefore, a well-described intervention protocol and valid and reliable assessment of procedural integrity would hold promise for systematic implementation and early identification of “intervention drift” (e.g., poor integrity).
Several reviews on the “current” status of procedural integrity have been conducted in the past three decades. They included the reviews of procedural integrity assessment in (a) experimental articles published in flagship behavior analytic journals, including *Journal of Applied Behavior Analysis* 1968–2005 (Gresham, Gansle, & Noell, 1993; McIntyre, Gresham, DiGennaro, & Reed, 2007; Peterson et al., 1982), school psychology literature 1995–2008 (i.e., *Journal of School Psychology, Psychology in the Schools, School Psychology Quarterly, School Psychology Review;* Sanetti, Gritter, & Dobey, 2011), and *Journal of Positive Behavior Interventions* 1999–2009 (Sanetti, Dobey, & Gritter, 2012); (b) intervention research for individuals with different disabilities such as learning disabilities (Gresham et al., 2000), emotional and behavioral disorders (Griffith, Hurley, & Hagaman, 2009), and autism (Wheeler, Baggett, Fox, & Blevins, 2006); (c) experimental research for different interventions such as early interventions (Ledford & Wolery, 2013) and communication interventions (Snell et al., 2010); and (d) intervention research conducted in natural environments such as school settings (Barnett et al., 2014; Fiske, 2008; Gresham, Gansle, Noell, & Gohen, 1993). Although the results of each review cannot be directly compared to one another due to differing samples and areas of focus, the reviews above consistently suggested that (a) procedural integrity has been addressed by an increasing number of researchers over time, (b) operational definitions of behavior analytic interventions are now found pervasively in research (e.g., McIntyre et al., 2007; Sanetti et al., 2012), and (c) it is more common for researchers to operationally define interventions than quantitatively measure implementation (e.g., Barnett et al., 2014; Sanetti & Kratochwill, 2008).
Procedural Integrity as a Professional Concern

For practitioners providing educational or behavioral services (e.g., behavioral consultants, school psychologists, teachers, and teacher assistants), procedural integrity is “the key to wide-scale application of evidence-based practices (EBPs) and to reaching a discipline of greater effectiveness” (Greenwood, 2009, p. 548). EBPs are instructional or intervention approaches that have proven to be effective through multiple high-quality research studies. Taking them to scale in the practices of service-providers is the only way to fully reap these benefits and meaningfully improve the performance of all learners (Torres, Farley, & Cook, 2012). In the process of transferring an EBP from research to practice, procedural integrity serves as the essential quality control and allows practitioners to properly interpret intervention success or failure (Greenwood, 2009; Gresham, 2009; Torres et al., 2012). For example, student outcomes may indicate that an evidence-based behavior analytic intervention produced no change on the target behavior. With accurate measurement of integrity, practitioners would at least be able to identify possible procedural factors, such as inconsistent implementation (low adherence) and low treatment strength (low dosage or poor quality), and then adapt the practice or implementation, if necessary, to meet students’ needs.

A consistent finding in the behavior analytic literature to date is that higher levels of integrity are usually associated with greater changes in behavior whereas treatment-integrity errors can be detrimental to intervention outcomes. This is essentially true for multi-component interventions that have a series of intervention steps. To demonstrate how procedural integrity moderates intervention outcomes, behavior analytic researchers
systematically manipulated different levels of procedural integrity and examined the resultant effects on student outcomes during a variety of skill-acquisition and behavior-reduction procedures, such as discrete-trial training (Carroll, Kodak, & Fisher, 2013; DiGennaro-Reed, Reed, Baez, & Maguire, 2011), prompting and prompt fading (Grow et al., 2009; Holcombe, Wolery, & Snyder, 1994; Noell, Gresham, & Gansle, 2002; Wilder, Atwell, & Wine, 2006), and differential reinforcement of alternative behavior (Northup et al., 1997; St. Peter Pipkin, Vollmer, & Sloman, 2010; Vollmer, Roane, Ringdahl, & Marcus, 1999; Wood, Umbreit, Liaupsin, & Gresham, 2007). Their findings consistently suggested that although the correlation between the level of procedural integrity and treatment outcomes was not always linear, high levels of procedural integrity were strongly linked to better behavioral outcomes and implementation errors of critical intervention components usually resulted in no or slow skill acquisition and/or the perpetuation of problematic behaviors.

Recent Research on Teachers’ Procedural Integrity

There has been a recent push for evidence-based behavior analytic interventions in schools as well as documentation of procedural integrity of teacher-delivered interventions (e.g., McIntyre et al., 2007). Investigating the degree to which interventions are carried out with fidelity by teachers in schools is valuable for several reasons. First, teachers are front-line staff to implement interventions with students who have different academic and behavioral needs. High levels of procedural integrity are likely to produce better outcomes for all students. As a result, teachers who are encouraged by desirable student outcomes would be likely to maintain high integrity levels over time; their
teaching behavior may be positively reinforced by positive student outcomes and/or social reinforcers from peers and school administrators as well as negatively reinforced by the reduction of problem behavior in students. Second, the extent to which teachers implement interventions with accuracy influences a behavior analyst’s ability to properly conduct formative evaluations on teacher and student performance (Gresham, 2009). For example, with accurate measurement of teachers’ procedural integrity, a behavior analyst would be able to determine if a student’s resistance to treatment is a result of an ineffective intervention or a lack of intervention implementation. Third, current legislation, such as the No Child Left Behind Act (U.S. Department of Education, 2002) and Individuals with Disabilities Improvement Act (2004), mandates that teachers be accountable for their educational practices. In other words, teachers are responsible for systematic implementation of evidence-based practices that have the potential to improve student outcomes.

Therefore, a number of review studies have been published in recent years exploring strategies to effectively improve procedural integrity in school settings (Cornelius & Nagro, 2014; Fallon, Collier-Meek, Maggin, Sanetti, & Johnson, 2015; Fiske, 2008; Kretlow & Bartholomew, 2010; McCahill, Healy, Lydon, & Ramey, 2014; Noell et al., 2014; Scheeler, Ruhl, & McAfee, 2004; Solomon, Klein, & Politylo, 2012). Although the behavior analytic intervention implemented by teachers may be different from one another, the studies above consistently suggested that it is the provision of professional development training that enables high levels of procedural integrity in school settings. Furthermore, effective and efficient training practices have included a
combination of antecedent-based and consequence-based approaches (e.g., instruction, modeling, role-play, and feedback) facilitated by a professional with expertise (see Chapter 2).

Despite advances in the efficacy of expert-facilitated training, the limited availability of qualified professionals who can provide comprehensive training is one key barrier to disseminating research findings in applied settings (e.g., public schools). Recent data suggest that there are 430,000 special education teachers working in public school classrooms (National Center for Education Statistics, 2014), but there are fewer than 22,000 certified behavior analysts in the United States (Behavior Analyst Certification Board, 2016). Although it is not clear how many behavior analysts work in public schools, there is no doubt that the number is much lower than the total number of special education teachers and paraprofessionals who may benefit from competency-based training in implementing behavioral analytic procedures. Given the (limited) trainer resources available, trainers may benefit from self-instructional resources that they can use to reduce reliance on expert-facilitated staff training.

**Purpose Statement**

In an effort to further explore this research line, this dissertation consists of one literature review and two single-case studies. Chapter 2 is a comprehensive literature review exploring strategies for training teachers to conduct behavior analytic interventions in school settings. Chapters 3 and 4 are two single-case studies evaluating the use of antecedent-only training strategies for training special education teachers on preference assessments and the system of least prompts, respectively, in the absence of an
expert trainer. Finally, Chapter 5 provides a general discussion summarizing the findings and describing future directions for research.
Chapter 2: Literature Review

The current literature review, which included 29 single-case studies with a focus on increasing teachers’ procedural integrity in school settings, aimed to identify (a) characteristics of students, teachers, and interventionists (i.e., trainers); (b) features of teacher-delivered behavior analytic interventions; (c) components and efficacy of training and follow-up support procedures; and (d) effects of teacher-delivered interventions on student outcomes, if included.

Method

Inclusion Criteria

The review included studies meeting four criteria. First, the study used a single-case experimental design to evaluate the efficacy of a teacher training procedure. Second, the training procedure aimed to increase in-service teachers’ treatment fidelity in implementing a multi-component behavior analytic intervention that directly addressed outcomes for students with or without disabilities in pre-kindergarten to Grade 12 classrooms. Third, the dependent variables included a repeated measure of teachers’ treatment fidelity. Fourth, the study took place in the United States and was published in English between 2000 to July, 2015. Studies were excluded when (a) participants were direct-care staff (e.g., Catania, Almeida, Liu-Constant, & DiGennaro Reed, 2009;
Sarokoff & Sturmey, 2004), paraeducators (e.g., Maggin, Fallon, Sanetti, & Ruberto, 2012), or pre-service teachers (e.g., Lylo & Lee, 2013); (b) an operational definition of the training procedure was not provided (e.g., Dart, Cook, Collins, Gresham, & Chenier, 2012); (c) experiments occurred in a non-school setting, such as agencies that were community-based (e.g., Bolton & Mayer, 2008), visual reality classrooms (e.g., Garland, Vasquez III, & Pearl, 2012), agencies providing educational and residential services to individual with special needs (e.g., DiGennaro, Martens, & Kleinmann, 2007), or early intervention programs for 0–3 years old (e.g., Plavnick, Ferreri, & Maupin, 2010); (d) target behavior analytic interventions did not consist of multiple intervention steps (e.g., specific praise; Duncan, Dufrene, Sterling, & Tingstrom, 2013); (e) target behavior analytic interventions were assessment-oriented, non-instructional procedures (e.g., functional analysis; Pence, St. Peter, & Giles, 2014); and (f) results of studies did not include a graphic display of teacher behavior (e.g., Koegel, Kim, & Koegel, 2014).

**Search Strategies**

Studies were identified using a three-step process. First, I systematically searched nine databases—Academic Search Complete, Academic Search Premier, Education Full Text (H.W. Wilson), Education Research Complete, ERIC, Professional Development Collection, Psychology and Behavioral Sciences Collection, PsycINFO, and Social Sciences Abstracts (H.W. Wilson)—using the following search string: [teacher* OR educator* OR school staff] AND [treat* OR interven* OR implement* OR procedur*] AND [fidelity OR integrity OR adherence]. To have a quality control over the experimental studies, I limited the search to studies that were published in peer-reviewed
journals. This database search yielded 1885 articles for title, abstract, and/or in-text review. After an initial screening, 21 articles met the inclusion criteria. Second, I applied backward and forward search techniques to articles identified in the first step. Backward search involved examining the references cited in the 21 articles, and forward search involved identifying articles that cited the 21 articles. Six studies were included during this process. Third, I conducted a manual search of nine journals (from 2000 to June, 2015). The journals hand-searched included *Journal of Applied Behavior Analysis, Journal of Behavioral Education, Journal of Educational and Psychological Consultation, Journal of School Psychology, Journal of Special Education, Psychology in the Schools, School Psychology Quarterly, Behavior Modification,* and *School Psychology Review.* These journals were selected because they typically publish empirical research on behavior analytic interventions, teacher training, and/or professional development, and the studies identified in the aforementioned two steps were published in these journals. One article was identified during this process.

**Study Coding**

The three-step search yielded 29 single-case studies (28 articles; one article had two studies eligible for inclusion). For each study, I coded key features of the participants, settings, dependent variables (i.e., teacher-implemented behavior analytic interventions), professional development packages, and participant outcomes. In addition, I also recorded the types of single-case design being used.

**Participants and settings.** For participating teachers, I coded gender, ethnicity, grade level, primary occupation category (i.e., special or general education teacher),
highest level of education, years of teaching experience, and past experience related to behavior analysis. For participating students, if student information was reported, I coded gender, ethnicity, age, and primary special education category. Additionally, I also recorded the role of the interventionists who provided professional development support to participating teachers. The interventionists were coded as school-based personnel (i.e., whose occupations in the school were specified), university researchers (i.e., who were university-employed research individuals such as faculty members and graduate students), and unknown (i.e., whose occupations or employment were not explicitly reported). For settings, I coded whether participating teachers were implementing the target behavior analytic procedures in a general or special education classroom. If the intervention was provided in general education setting, I noted whether students with disabilities were integrated in the classroom. If it was implemented in a special education classroom, I recorded whether the students have access to peers who are typical developing (e.g., resource classrooms) or not (e.g., self-contained).

**Dependent variables.** To describe the dependent variables, I recorded the name of target behavior analytic interventions, complexity of procedures (i.e., number of steps or components), and length or frequency of implementation.

**Professional development packages.** First, an exhaustive list of the training and support components that appeared in the professional packages was created. I read author descriptions for the presence of the following components: (a) a description of the target behavior analytic intervention; (b) prompting, defined as antecedent cues that increased the likelihood of accurate or adequate implementation; (c) modeling, defined as the in-
person or video model of accurate implementation of the intervention; (d) role play, defined as the teacher practicing the target intervention in a simulated setting; (e) performance feedback, defined as providing the teacher with information about his or her implementation and guidance on how to improve future performance; and (f) self-monitoring, defined as the teacher monitoring his or her implementation using a checklist or rating scale.

Second, I noted whether initial training and ongoing support were provided based on teachers’ responses to intervention. Specifically, I coded whether teachers’ fidelity improved to a high level by initial training but became unstable or exhibited a downward trend in the absence of follow-up support on the job. Third, I recorded the length and frequency of initial and follow-up training. For studies that described a fading procedure for follow-up training on the job, I also coded whether they gradually thinned the schedule of support delivery (e.g., reducing the frequency of performance feedback from daily to every other day) or withdrew a specified training component (e.g., removing written feedback from daily self-monitoring sheets). Finally, I noted whether procedural integrity of teacher training was monitored over time and whether quantitative results were reported.

**Participant outcomes.** Teacher outcomes with initial training and/or ongoing support were reported using the number of teachers reaching the established mastery criteria or standard quantitative measures (i.e., mean, standard deviation, or range) if the mastery criterion was not specified. To identify the extent to which improved teacher fidelity lasted over time or generalized across untrained contexts, I also described teacher
outcomes in maintenance and generalization condition. Additionally, I recorded whether researchers used social validity measures to evaluate the extent to which teachers found the target behavior analytic interventions and the teacher training procedures socially acceptable. Finally, to evaluate outcomes of teacher-delivered interventions, student performance was visually analyzed and reported in the form of a success estimate. This metric indicates the consistency of treatment effects and informs the extent to which introduction or withdrawal of the intervention coincided with a clear therapeutic change in the pattern of the dependent variable (Brock & Carter, 2013). A success estimate is the ratio of demonstration of an experimental effect (i.e., numerator; number of treatment conditions during which student performance improved) to opportunities of demonstration (i.e., denominator; number of treatment conditions under which teachers received initial training or on-going support and demonstrated higher levels of treatment fidelity). To make a definite decision over the presence of an experimental effect, the coders (the first author and a graduate student) followed the rules for visual analysis developed by Kratochwill et al. (2011) (e.g., examining level, trend, variability, immediacy of the effect, overlap, and consistency of data patterns across similar conditions). The method of success estimate was based upon the metric used by Brock and Carter (2013) but slightly modified to indicate the consistency of treatment effects when teachers demonstrated an increased level of implementation fidelity.
Intercoder Agreement

After the three-step searches, a second coder (a trained graduate student) independently read each abstract or complete text of the 28 articles (29 studies) to make a more definitive decision regarding inclusion. Agreement was 100% (i.e., the two authors agreed on whether the article should be included or excluded for all 28 articles).

I read each study and coded key features of the participants, settings, dependent variables (i.e., teacher-implemented behavior analytic interventions), professional development packages, participant outcomes, and types of single-case design (as discussed above). The second coder independently examined 15 studies and coded all variables for each study. An agreement was defined as matching description of a given component reviewed (e.g., teachers’ gender). Mean intercoder agreement across components was 98% (range: 93–100%).

Results

Twenty-nine single-case studies were reviewed and analyzed. Although multiple-baseline-across-participants was the most common design (n = 25), multiple treatment design was also used (n = 4).

Characteristics of Participants and Teacher Trainers

Across studies, 108 in-service teachers received training on accurate implementation of behavior-analytic interventions in their classrooms. General characteristics of participating teachers are summarized in Table 1.
<table>
<thead>
<tr>
<th>Participating Teachers (N = 108)</th>
<th>Gender</th>
<th>Grade Level</th>
<th>Highest Level of Education</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>66</td>
<td>Pre-K–Kindergarten</td>
<td>24 28 High School Diploma</td>
</tr>
<tr>
<td>Male</td>
<td>8</td>
<td>Grade 1–5</td>
<td>51 59 Associate Degree</td>
</tr>
<tr>
<td>Unknown</td>
<td>34</td>
<td>Grade 6–8</td>
<td>10 12 Bachelor’s Degree</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>Grade 9–12</td>
<td>1 1 Master’s Degree</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td>Unknown</td>
<td>22 56 Unknown</td>
</tr>
<tr>
<td>African-American</td>
<td>6</td>
<td>General Education</td>
<td>71 66 0–5</td>
</tr>
<tr>
<td>Asian-American</td>
<td>1</td>
<td>Special Education</td>
<td>37 34 6–10</td>
</tr>
<tr>
<td>Euro-American/Caucasian</td>
<td>40</td>
<td>Unknown</td>
<td>0 56 &gt;10</td>
</tr>
<tr>
<td>Unknown</td>
<td>61</td>
<td>Unknown</td>
<td>56 45 Unknown</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Target Students (N = 102)</th>
<th>Gender</th>
<th>Age</th>
<th>No Disability</th>
<th>Primary Disability Diagnosis&lt;sup&gt;b&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Female</td>
<td>18</td>
<td>3–5</td>
<td>12</td>
<td>53 ADHD</td>
</tr>
<tr>
<td>Male</td>
<td>60</td>
<td>6–10</td>
<td>50</td>
<td>52 EBD</td>
</tr>
<tr>
<td>Unknown</td>
<td>24</td>
<td>11–13</td>
<td>14</td>
<td>1 LD</td>
</tr>
<tr>
<td>Unknown</td>
<td></td>
<td>14–17</td>
<td>1</td>
<td>1 ASD</td>
</tr>
<tr>
<td>Ethnicity</td>
<td></td>
<td>19–22</td>
<td>3</td>
<td>14 Down Syndrome</td>
</tr>
<tr>
<td>African-American</td>
<td>11</td>
<td>Unknown</td>
<td>22</td>
<td>1 Multiple Disabilities</td>
</tr>
<tr>
<td>Euro-American/Caucasian</td>
<td>26</td>
<td>Unknown</td>
<td>22</td>
<td>32 Multiple Disabilities</td>
</tr>
<tr>
<td>Native American</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Unknown</td>
<td>63</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

<sup>a</sup> Participants whose characteristics were unreported (i.e., unknown) were excluded from calculation.

<sup>b</sup> ADHD = attention-deficit/hyperactivity disorder, EBD = emotional and behavioral disorder, LD = learning disability, ASD = autism spectrum disorder.

Table 1. Literature review participant demographics
Teachers’ demographics and backgrounds were described in varying levels of details. Among the studies that reported teachers’ gender and ethnicity, female teachers ($n = 66; 89\%$) were recruited more often than male teachers and teachers who were European-American ($n = 40; 85\%$) outnumbered teachers from other backgrounds. Participating teachers were working in pre-kindergarten, elementary, middle, and high school classrooms, and nearly half of them taught at the primary level (i.e., Grade 1 to 5; $n = 51$).

When information regarding professional backgrounds was available, participating teachers’ highest level of education ranged from high school diploma (Dufrene et al., 2012) to master’s degree (e.g., Codding, Livanis, Pace, & Vaca, 2008), and their teaching experience ranged from half a month (Codding, Feinberg, Dunn, & Pace, 2005) to as many as 30 years (DiGennaro, Martens, & McIntyre, 2005). Eleven studies reported teachers’ past experience related to behavior analysis. The participating teachers accrued behavior analytic knowledge and experience from a variety of sources, including attending in-service workshops (Bethune & Wood, 2013; Codding et al., 2005; Dufrene et al., 2012; McKenney & Bristol, 2014; Minor, DuBard, & Luiselli, 2014; Zoder-Martell et al., 2013), attending college courses (Mouzakitis, Codding, & Tryon, 2015), and implementing behavior analytic interventions (Kretlow, Cooke, & Wood, 2012; Kretlow, Wood, & Cooke, 2011; Miller, Carlson, & Sigurdsson, 2014; Minor et al., 2014). One study noted that their participants received no formal training on the use of specific behavioral interventions (Sterling-Turner, Watson, & Moore, 2002).
All studies noted the primary occupation category of their participating teachers. The number of general education teachers ($n = 71; 66\%$) exceeded that of special education teachers ($n = 37; 34\%$). When information regarding classroom category was available, eight general education teachers and eight special education teachers implemented instruction in classrooms where students with special needs were integrated, and 18 special education teachers implemented intervention in segregated classrooms for students with disabilities.

Across the studies that specified student demographics and backgrounds ($n = 21$ out of 29 studies), 102 students received teacher-implemented behavior analytic interventions. Student characteristics are reported in Table 1. Nearly half of participating students aged six to ten ($n = 50$). All studies reported primary disability diagnosis of participating students, if there was any. Students who received special education services made up about half of the group ($n = 53; 52\%$).

In most studies, university researchers served as the primary interventionists who implemented initial training and on-going support for in-service teachers ($n = 27$ out of 29 studies). The remaining two studies used ad hoc professionals as primary trainers, including instructional supervisors (Miller et al., 2014), social workers (Sanetti, Fallon, & Collier-Meek, 2013), and special education teachers (Sanetti et al., 2013).

**Features of Teacher-Implemented Interventions**

Teachers received training on correctly implementing a variety of behavior analytic interventions. Most of the interventions aimed to increase desirable behavior and reduce problematic behavior in students ($n = 17$ out of 29 studies). These educational
practices included function-based behavior intervention plans \((n = 11)\), classwide behavior management plans \((n = 2)\), good behavior game \((n = 2)\), check-in-check-out \((n = 1)\), and self-monitoring \((n = 1)\). Teachers in the remaining 12 studies were asked to implement instruction purposed for skill acquisition, including reading instruction \((n = 4)\), math intervention \((n = 4)\), discrete trial training \((n = 2)\), effective instructional delivery \((n = 1)\), and prompting \((n = 1)\).

Depending on the nature of behavior analytic interventions, the procedural complexity ranged from three steps only (e.g., self-monitoring; Noell et al., 2002) to as many as 27 steps (function-based behavior intervention plans; Sanetti et al., 2007). In most studies, teachers were required to implement the target intervention on a daily basis \((n = 23\) out of 29 studies).

**Efficacy of Professional Development Packages**

**Training components.** Across studies, nearly all in-service teachers received a multi-component professional development package that improved treatment fidelity of teacher-delivered behavior-analytic interventions. Twenty-two of 29 studies monitored fidelity of package (i.e., training) delivery over time and reported quantitative results. Descriptions of professional development packages are summarized in Table 2.
<table>
<thead>
<tr>
<th>Studya</th>
<th>Training Components</th>
<th>Data-Based Tiered Trainingc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Description</td>
<td>Prompting</td>
</tr>
<tr>
<td>Andersen and Daly, III (2013)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Bethune and Wood (2013)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Codding et al. (2005)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Codding et al. (2008)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>DiGennaro et al. (2005)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Dufrene et al. (2012)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Duhon et al. (2009) Study 1</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Duhon et al. (2009) Study 2</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Ely et al. (2015)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Table 2. Descriptions of professional development packages continued
<table>
<thead>
<tr>
<th>Study</th>
<th>Description</th>
<th>Prompting</th>
<th>Modeling</th>
<th>Role Play</th>
<th>Feedback&lt;sup&gt;b&lt;/sup&gt;</th>
<th>Self-Monitoring</th>
<th>Data-Based Tiered Training&lt;sup&gt;c&lt;/sup&gt;</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gilbertson et al. (2007)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Gross et al. (2014)</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Kaufman et al. (2013)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Kretlow et al. (2012)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Kretlow et al. (2011)</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Lerman et al. (2008)</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>McKenney and Bristol (2014)</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Miller et al. (2014)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X&lt;sup&gt;b&lt;/sup&gt;</td>
</tr>
<tr>
<td>Minor et al. (2014)</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Mouzakitis et al. (2015)</td>
<td>X</td>
<td></td>
<td></td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
</tr>
<tr>
<td>Noell et al. (2002)</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>X</td>
</tr>
</tbody>
</table>

continued
Table 2 continued

<table>
<thead>
<tr>
<th>Studya</th>
<th>Training Components</th>
<th>Data-Based Tiered Trainingc</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Description</td>
<td>Prompting</td>
</tr>
<tr>
<td>Oliver et al. (2015)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Rodriguez et al. (2009)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Sanetti et al. (2014)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Sanetti et al. (2013)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Sanetti and Kratochwill (2009b)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Sanetti and Kratochwill (2011)</td>
<td>X</td>
<td></td>
</tr>
<tr>
<td>Sanetti et al. (2007)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Sterling-Turner et al. (2002)</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>Zoder-Martell et al. (2013)</td>
<td>X</td>
<td>X</td>
</tr>
</tbody>
</table>

Notes. a. Studies in bold monitored fidelity of independent variable (i.e., teacher training) over time and reported quantitative results.
   b. Teachers’ treatment fidelity (e.g., percentage of correct components) was reviewed as part of performance feedback.
   c. Training components were introduced in sequence based on teacher performance (i.e., a response-to-intervention model).
Nearly all training packages included a description of educational practice \((n = 27; 93\%)\) and performance feedback associated with teachers’ actual implementation \((n = 26; 90\%)\). In most studies, interventionists provided a live or video model and allowed teachers to role-play the strategy in a simulated setting \((n = 19 \text{ and } 18, \text{ respectively})\). Fewer descriptions of training included prompting in natural environment \((n = 8)\) and self-monitoring \((n = 4)\). In 22 studies, training components were introduced in sequence based upon teacher performance (i.e., their responses to intervention).

There are some interesting variations in how descriptions and performance feedback were implemented. Descriptions were provided to teachers with different modalities, including vocal instructions (e.g., Bethune & Wood, 2013), written descriptions (Rodriguez et al., 2009), a combination of vocal and written descriptions (e.g., Andersen & Daly, III, 2013), or instructions delivered in a video clip (Ely et al., 2015). Additionally, some descriptions included a rationale for why the practice was important (Andersen & Daly, III, 2013; Bethune & Wood, 2013; Dufrene et al., 2012; Ely et al., 2015; Gilbertson et al., 2007; Kretlow et al., 2012; Kretlow et al., 2011; Lerman et al., 2008; Sanetti et al., 2013; Sterling-Turner et al., 2002).

Performance feedback was delivered (a) at varying frequencies, ranging from daily (e.g., DiGennaro et al., 2005) to every other week (e.g., Codding et al., 2005); (b) with different communication modalities, including in-person meetings (e.g., Andersen & Daly, III, 2013), written notes (e.g., Kaufman et al., 2013), and emails (e.g., McKenney & Bristol, 2014); (c) within different supervisory contexts, for example, individual consultation (e.g., Bethune & Wood, 2013) or small-group meetings (e.g., Sanetti et al.,
2013); and (d) at a temporal locus of either following (e.g., Rodriguez et al., 2009) each observation or prior to subsequent observations (e.g., Noell et al., 2002). As part of performance feedback, 15 studies provided up-to-date reviews of the teacher’s progress. Except for Bethune and Wood (2013), 14 studies provided teachers with a reader-friendly, graphic display of their treatment fidelity. Among these studies, eight also updated teachers on the progress made by their students as professional development packages were implemented (DiGennaro et al., 2005; Duhon et al., 2009; Gibertson et al., 2007; Gross et al., 2014; Noell et al., 2002; Sterling-Turner et al., 2002; Zoder-Martell et al., 2013).

Twenty-seven of 29 studies described some form of follow-up support after the initial training session. Among those studies, eight indicated that teachers’ fidelity improved to a high level by initial training but became unstable or exhibited a downward trend in the absence of follow-up support on the job. The length of the initial training session was reported across 13 of 29 studies, and duration of the initial training ranged from 5 min to 24 hours. Nineteen of 27 studies indicated the length and/or frequency of follow-up support. The length of follow-up support was less variable than that of the initial training session, ranging from 3 to 85 min, and one follow-up session was no more than 10 min in most studies ($n = 11$ out of 19 studies). Follow-up support was provided at varying frequencies, ranging from one time only to as often as daily. Across these studies, provision of performance feedback was the most commonly used strategy during follow-up support ($n = 22$ out of 27 studies). In addition, some researchers used alternative follow-up strategies, including modeling prompts delivered over an in-ear
earbud (Dufrene et al., 2012), asking teachers to self-monitor their implementation (Mouzakitis et al., 2015; Sanetti & Kratochwill, 2009b; Sanetti & Kratochwill, 2011), and assisting teachers in planning implementation and problem-solving (Sanetti et al., 2014). Finally, only six studies described how they systematically removed follow-up support (i.e., support fading). Four of 6 studies gradually thinned the schedules that follow-up support was delivered (DiGennaro et al., 2005; Gross et al., 2014; Kaufman et al., 2013; Noell et al., 2002), and two remaining studies removed a specific training component contingent upon desired teacher performance (i.e., performance feedback; McKenney & Bristol, 2014; Mouzakitis et al., 2015).

**Training outcomes.** All studies tested whether professional development packages resulted in a higher level of treatment fidelity in teachers. Eleven studies evaluated the extent to which high fidelity lasted over time in the absence of support or with reduced support (38%). And fewer studies investigated the degree that teachers generalized accurate implementation to new students or untrained settings (n = 4; 14%).

Across studies, all professional development packages improved teachers’ treatment fidelity to a higher level and/or greater consistency. In the studies that provided initial training and follow-up support (n = 27), nearly all teachers increased fidelity to a desired level after follow-up support on the job. Fewer teachers were able to implement the interventions with fidelity in the absence of support (i.e., antecedent-based training only; Bethune & Wood, 2013; Gilbertson et al., 2007). When it comes to teachers’ maintenance of accurate implementation, the data were mixed. Among the studies that collected maintenance data, five studies reported high fidelity across all participating
teachers (Bethune & Wood, 2013; Coddington et al., 2005; Dufrene et al., 2012; Gross et al., 2014; McKenney & Bristol, 2014). Four presented mixed results among teachers (Gilbertson et al., 2007; Kaufman et al., 2013; Lerman et al., 2008; Sanetti et al., 2014). And the remaining two studies showed that participating teachers’ behavior unfortunately reversed to baseline performance (Duohon et al., 2009) as support on the job discontinued. In addition, there are some variations as to when maintenance probes were conducted; the length of time that elapsed after discontinuing professional development ranged from less than 1 week (e.g., Dufrene et al., 2012) to as long as 3 months (Lerman et al., 2008).

When information regarding teachers’ skill generalization was available (n = 4 studies), 17 of 20 teachers implemented the interventions with fidelity across untrained settings and/or students. Specifically, skill generalization was demonstrated when the teachers were leading a different activity (Bethune & Wood, 2013; Kretlow et al., 2012) and working with new students who would benefit from the target interventions (Lerman et al., 2008; Mouzakitis et al., 2015).

Twenty of 29 studies evaluated and reported the extent to which participating teachers found the behavior analytic interventions and/or professional development packages socially acceptable. A rating scale was the most common tool for measuring social validity (n = 17 out of 20 studies). Fewer studies used formal or informal interviews (n = 3) or a combination of open-ended questions and a rating scale (n = 4). Results of social validity measures showed that all participating teachers responded positively to the behavior analytic interventions they practiced and/or the training process.
Efficacy of Teacher-Implemented Interventions

Nineteen studies experimentally investigated changes in student outcomes as a result of changes in teachers’ treatment fidelity; however, only 15 studies provided a graphic display of student behavior. They generated 104 demonstrations of an experimental effect in 112 planned opportunities. As a result of high-fidelity interventions, students displayed more on-task behavior (e.g., Gross et al., 2014), improved academic performance (e.g., Duhon et al., 2009), and reduced number of problem behaviors (e.g., Andersen & Daly, III, 2013).

Discussion

The primary purpose of this literature review was to synthesize the available single-case studies in the past 15 years and identify effective professional development practices for monitoring and improving treatment fidelity of teacher-delivered behavior analytic interventions. A systematic review of 29 studies indicated that given appropriate initial training and on-going support on the job, in-service teachers were capable of implementing behavior analytic interventions at a high level of fidelity and with great consistency. Systematic implementation by teachers had the potential to improve educational outcomes for all students. Specifically, the findings are summarized in three key themes.

First, the provision of professional development training enabled accurate implementation of behavior analytic interventions for all in-service teachers, although teachers had fairly diverse levels of education and professional experience. One of the interesting findings was that in the absence of training, neither postsecondary education
(e.g., bachelor’s or master’s degrees) nor extensive in-service training experience was a guarantee for high-level treatment fidelity on the job. Although it remained unknown whether advanced educational or professional backgrounds were significantly correlated with desirable training outcomes, this review emphasized that in-service teachers in school settings would need extensive training and follow-up support in order to improve their implementation capacity (Stahmer et al., 2015). Additionally, a vast majority of participating teachers (66%) were found to be general education teachers working with students without disabilities. The significance of this finding is two-fold. First, behavior analytic interventions and their principles seemed to be applied in general education classrooms to a large extent, which had the potential for improving academic or behavioral outcomes for all students. Second, effective training enabled high-quality implementation of behavior analytic strategies for all participating teachers, regardless of the fact that some general education teachers might adhere to a philosophy of learning other than behavior analysis (e.g., constructivism).

Second, nearly all studies (n = 28; 97%) used a multi-component professional development package to improve teachers’ accurate implementation of behavior analytic interventions. A vast majority of training packages included a description of the educational practice (93%), modeling (66%), role-play (62%), and performance feedback (90%). The findings documented some degree of correspondence with the best training practice identified in behavior analytic literature, which is known as behavioral skills training (BST). BST is an evidence-based, effective, and efficient protocol for training human service staff to implement intervention plans for challenging behavior or teaching
strategies (Parsons, Rollyson, & Reid, 2012; Sarokoff & Sturmey, 2004). BST also involves instruction, modeling, role-play, and performance feedback. However, each step in BST is an integral part of the training protocol, which is in contrast with nearly half of the reviewed training practices ($n = 13; 45\%$). In these training practices, although descriptions of the educational practice and performance feedback were the common training components, modeling and role-play were not consistently employed by researchers. A possible explanation for this may be rooted in the scope and context of training. BST is designed for training novice staff using a group format (Parsons et al., 2012), but the training packages under review primarily supported an individual in-service teacher on the job and oftentimes the training protocol had to be tailored depending on the teacher’s needs as well as the contextual features.

Third, in order to improve teachers’ treatment fidelity, nearly all studies ($n = 27; 93\%$) described an initial training followed by some form of follow-up support on the job. During the initial training phase (e.g., description), in-service teachers were trained to perform work duties that they previously performed with low fidelity or could not perform prior to training. Then continuing training and support (e.g., performance feedback, self-monitoring) was provided until teachers competently demonstrated the skill on the job. This two-part model consists of performance- and competency-based strategies and thus meets the gold standards of evidence-based training as identified by behavioral researchers (e.g., Parsons et al., 2012; Reid et al., 2003). Additionally, some of the reviewed studies reported that teachers’ fidelity improved to a high level by initial training but declined in the absence of follow-up support. The findings were aligned with
previous literature on the (lack of) effectiveness of one-time training events such as 1-day workshops (Brock & Carter, 2013; Cavanaugh, 2013; Noell et al., 2014; Stahmer et al., 2015). Finally, across the reviewed studies, performance feedback is the most popular method for implementation support on the job, which echoed previous literature on the viability of performance feedback as a follow-up support strategy in improving teachers’ treatment fidelity (Cavanaugh, 2013; Fallon et al, 2015; Noell et al., 2014; Scheeler et al., 2004; Solomon et al., 2012; Stormont & Reinke, 2014).

**Implications for Future Practice**

The current review provided potential guidance related to in-service training and on-the-job support for increasing treatment fidelity of teacher-delivered behavior analytic interventions. First, effective and evidence-based professional development packages shall consist of initial training and follow-up support. The former improves teachers’ performance skills, and the latter trains teachers to meet established mastery criteria on the job (Parsons et al., 2012). In addition, expectations for teacher performance should be clearly communicated throughout the whole training process (e.g., providing a clear description of intervention steps, delivering specific and constructive feedback based upon teacher performance), and it is suggested that clear expectations be supplemented by other evidence-based training procedures, such as modeling, role-play, prompting, and self-monitoring. Besides, teacher coaches or supervisors may want to directly monitor teacher performance over time and promptly modify the protocol depending on the teacher’s response to training as well as the student’s response to intervention. When teachers are demonstrating a high level of treatment fidelity that has resulted in positive
student outcomes, coaches and supervisors are then encouraged to build structures to support sustained use across classroom contexts. Empirically validated strategies to support skill generality include but are not limited to self-monitoring (e.g., Mouzakitis et al., 2015), thinning the schedule of support delivery (e.g., Kaufman et al., 2013), arranging indiscriminable contingencies (e.g., supervisory meetings contingent upon treatment fidelity being lower than 100% on a randomly selected day; Gross et al., 2014), and setting avoidance contingencies (e.g., teachers avoiding supervisory meetings contingent upon high-level treatment fidelity; McKenney & Bristol, 2014). Finally, it is suggested that coaches and supervisors demonstrate competencies as they implement the training, and procedural integrity of the training procedures shall be directly monitored over time.

**Limitations**

This review has some limitations. First, the review did not critically evaluate the methodological quality of the included studies. For example, Zoder-Martell et al. (2013) documented only two demonstrations of the treatment effect in their single-case design, which did not meet the single-case design standards proposed by the What Works Clearinghouse (Kratochwill et al., 2010). Second, more than half of the reviewed studies ($n = 17; 59\%$) sampled teachers’ fidelity on some occasions; however, the current review did not investigate whether the methods used to sample treatment fidelity were explicated and defensible. Third, this review examined whether professional development packages were consistently effective, and it is unclear which training components or combination of components were functionally related to the improvement of teachers’ fidelity. Finally,
effect size of professional development packages was not systematically evaluated. A quantitative synthesis of the reviewed studies is warranted to carefully assess the efficacy of professional development packages for in-service teachers and the effects of teacher-delivered behavior analytic interventions on student outcomes.

**Recommendations for Future Research**

Several inquiries about improving teachers’ fidelity need to be addressed in future studies. First, although a vast majority of the studies used performance feedback as the primary competence-based training strategy on the job, there are some emerging alternative methods that need further investigation. Future research, in particular, should examine strategies that not only train teachers to established mastery criteria but also support their sustained use across classroom contexts (e.g., self-monitoring). Additionally, given the mixed maintenance results across the reviewed studies, researchers may also want to identify effective strategies to systematically fade on-the-job support (e.g., gradually thinning the schedule of support delivery) and to train teachers’ teaching behavior to contact naturally existing reinforcement.

Second, future researchers may also want to identify environmental variables that contribute to poor fidelity (e.g., Performance Diagnostic Checklist; Austin, 2000) and then develop function-based training practices. Specifically, it is suggested that researchers experimentally test how contextual variables may have moderated teachers’ treatment fidelity; the contextual variables to be examined include but are not limited to teacher competence, complexity of the intervention, time required to implement the intervention, access to materials and resources necessary to implement the intervention,
actual effectiveness of the training (i.e., teacher outcomes), actual effectiveness of intervention (i.e., student outcomes), and teacher acceptability of training activities.

Third, 27 out of 29 professional development packages involved the administration of non-school personnel (e.g., graduate students, faculty members), and future research should investigate whether school-based personnel (e.g., coaches, district-based consultants) can be trained to appropriately implement critical components of direct training, such as descriptions, modeling, and performance feedback.

Finally, in public and private schools where university-based or school-based behavioral experts are not widely available, in-service teachers may benefit from a set of low-cost, highly portable training resources they can use to reduce reliance on expert-facilitated training. One strategy to minimize the need for expert staff trainer is to use self-instruction packages (e.g., enhanced written instructions) as antecedent-based training procedures. Unfortunately, those procedures were primarily used to train university students (Miljkovic, Kaminski, Yu, & Wishnowski, 2015; Ramon, Yu, Martin, & Martin, 2015) and behavioral staff in human service agencies (Deliperi, et al., 2015; Graff & Karsten, 2012; Lipschultz, Vladescu, Reeve, Reeve, & Dipsey, 2015; Weldy, Rapp, & Capocasa, 2014). Further investigations are warranted to evaluate the extent to which antecedent-based training is applicable for in-service teachers, especially when expert-facilitated training is not available.
Chapter 3: Single-Case Study One

This chapter includes a stand-alone research report of a single-case study. This section includes a brief introduction, a description of the method, a summary of results, and a brief discussion of the findings and its implications.

Staff Training on Stimulus Preference Assessments

Direct teaching procedures that support skill acquisition for individuals with disabilities often involve identifying stimuli that may function as potent reinforcers (Cooper et al., 2007; Paramore & Higbee, 2005). In order to identify potential reinforcers to be used in consequence-based procedures, researchers have developed multiple formats of stimulus preference assessments (Carr, Nicolson, & Higbee, 2000; DeLeon & Iwata, 1996; Fisher et al., 1992; Fleming et al., 2010; Pace, Ivancic, Edwards, Iwata, & Page, 1985). The paired-stimulus (PS) and multiple-stimulus without replacement (MSWO) preference assessments involve multiple presentations of two or more stimuli and require the student to select only one stimulus at a time. The assessor obtains a preference hierarchy based on the student’s selection responses. Numerous studies have documented the predictive validity of the two assessment procedures in that the stimuli selected in the PS and MSWO procedures did function as reinforcers in the subsequent
tests of reinforcement effects and could be used to teach new skills (DeLeon & Iwata, 1996; Fisher et al., 1992; Kodak, Fisher, Kelley, & Kisamore, 2009).

Given the size of the literature on assessing preferences of individuals with special needs (see Cannella, O’Reilly, & Lancioni, 2005, for a review), implementation of stimulus preference assessment procedures are important skills in which to train special education staff who work with individuals with disabilities. Consistent with the findings of the literature review in Chapter 2, effective training practices on preference assessments have included a combination of antecedent-based and consequence-based approaches (e.g., instruction, modeling, role-play, and feedback) facilitated by a professional with expertise (Lerman, Tetreault, Hovanetz, Strobel, & Garro, 2008; Lerman, Vorndran, Addison, & Kuhn, 2004; Roscoe, Fisher, Glover, & Volkert, 2006; Roscoe & Fisher, 2008).

For example, Roscoe and Fisher (2008) evaluated the efficacy of a brief training procedure that incorporated role-play and feedback to train newly hired behavior technicians to conduct PS and MSWO preference assessments. Trainees were first given data sheets and brief summaries of the PS and MSWO assessments from the method sections of the published research articles. Then they were exposed to two 15- to 20- min training sessions (one for PS assessment and one for MSWO assessment) in a counter-balanced order. Each training session involved role-play practices (i.e., the trainer demonstrating all possible client responses) and feedback (i.e., the trainer noting whether or not each target behavior was performed correctly by the trainee). After training, 16 cases (two assessments for eight trainees) demonstrated over 80% correct responses and
14 cases increased to over 90% during the simulated sessions (i.e., implementing the preference assessments with a role-played client).

Compared to Roscoe and Fisher (2008), Lerman et al. (2008) used a similar but more intensive expert-facilitated approach to teach three formats of preference assessments (including PS and MSWO preference assessments) to public school special education teachers who taught children with developmental disabilities. The 5-day training program involved 18 hours of group instructions (i.e., lectures) and 12 hours of individual training on the job (i.e., modeling and feedback). The authors reported that all teachers met the mastery criterion (i.e., two consecutive sessions with 100% accuracy) for all of the skills during the training, and with brief post-observation feedback 6 of 9 teachers’ performance maintained and generalized to their classrooms from 2 to 6 months following the conclusion of training.

**Training via written instructions alone.** Despite advances in the efficacy of expert-facilitated training, the limited availability of qualified professionals who can provide comprehensive training is one key barrier to disseminating research findings in applied settings (e.g., public schools). Therefore, researchers have been investigating training methods that they can use to reduce reliance on expert-facilitated staff training.

One approach that has been used to teach staff to implement preference assessments in the absence of an expert trainer is written instructions. Researchers report that written instructions obtained from different sources (e.g., published articles, other professionals) may lead to different training outcomes (e.g., Graff & Karsten, 2012; Roscoe et al., 2006; Roscoe & Fisher, 2008). Specifically, written instructions obtained
from published research articles (i.e., procedural description from the method sections) have been shown to be insufficient in training novice staff to implement behavior analytic procedures (e.g., stimulus preference assessment) (e.g., Graff & Karsten, 2012; Lavie & Sturmey, 2002; Ramon, Yu, Martin, & Martin, 2015; Rosales, Gongola, & Homlitas, 2015; Roscoe et al., 2006; Roscoe & Fisher, 2008). For example, Rosales et al. (2015) provided inexperienced teachers with excerpts from previously published research articles that outlined correct implementation of three types of stimulus preference assessments (including PS and MSWO procedures), and those written instructions were supplemented with a list of correct responses for each assessment (i.e., task analysis). Following instruction-only training, teachers’ percentage of correct responses did not reach the mastery criterion, averaging only 41%, 54%, and 30% for each assessment. It appeared from the results that such written instructions (i.e., procedural description from the method sections) did not provide teachers with the multiple stimulus and response exemplars necessary for teacher’s correct performance of the procedures (e.g., a full range of learner responses and assessor behavior).

Some expert trainers and researchers therefore have attempted to enhance the content of written instructions, and a few demonstrations of training efficacy have been published (e.g., Graff & Karsten, 2012; Ramon et al., 2015). For example, Graff and Karsten (2012) developed enhanced written instructions for PS and MSWO preference assessments and evaluated the efficacy of those instructions in preparing inexperienced special education teachers to conduct preference assessments and interpret assessment results. Enhanced written instructions, with a large number of relevant discriminative
stimuli embedded in the diagrams and step-by-step examples, were found to be more effective than written instructions extracted from the method section of previously published research articles. After the 11 participating teachers reviewed enhanced written instructions, their performance increased to nearly 100% correct responses for both assessments and demonstrated generalization to actual students.

In a similar approach to training via written instructions, Ramon and her colleagues (2015) used a 32-page self-instructional manual on MSWO preference assessment. This enhanced manual was written in nontechnical language and consisted of an introduction on preference assessment, its importance for individuals with developmental disabilities, step-by-step instructions for the procedures, text boxes that highlighted the target assessor behavior, a full procedural checklist, eight review exercises, and an answer key for each exercise. Consistent with the results of Graff and Karsten (2012), the experimenters found that the manual was more effective in guiding correct implementation than the procedural description adapted from the method sections of published research articles. Eleven of the 18 participants achieved mastery on MSWO preference assessment implementation after reviewing the manual, whereas only one participant achieved mastery after studying the method sections. The results also suggested that enhanced written instruction might not be sufficient to produce mastery criterion performance for some participants. Of the six participants who did not meet the mastery criterion with written instructions, all met mastery after observing a live demonstration of the procedure (i.e., modeling).
In summary, enhanced written instructions may serve as low-cost, highly portable training resource with a potential to reduce reliance on expert-facilitated training. As demonstrated in Ramon et al. (2015), trainers may minimize the amount of trainer-delivered resources needed with the preparation afforded by the instructions, and the savings in trainer time and cost accrued could be used to provide remedial training to a subset of trainees in individualized areas of need.

**Training via Practitioner Journal Articles**

Despite evidence for the efficacy of enhanced written instructions, one key barrier to disseminating research findings among school-based trainers surrounds the limited access to such well-developed instructions, which encompass a large number of relevant information necessary for trainee’s accurate implementation. In search of reliable resources for enhanced written instructions, school-based trainers may benefit from articles previously published in practitioner-oriented journals (i.e., practitioner journal articles).

Unlike research articles whose main purpose is to disseminate research findings, practitioner journal articles seek to translate research findings for use in applied settings by providing procedural descriptions that are specific and detailed enough to allow direct and immediate use by readers (Ludlow & Dieker, 2013; Ludlow, Dieker, & Powell, 2014). For example, according to Ludlow and Dieker, practitioner journal articles published in *Teaching Exceptional Children* (a flagship practitioner-oriented journal) usually (a) provide specific examples and pictures that illustrate how to apply the practice (e.g., step-by-step examples), (b) include tables and figures that assist readers in
implementing the practice (e.g., sample data sheets, diagrams), and (c) are written in nontechnical language. It seems that some practitioner journal articles share common features with existing enhanced written instructions (e.g., Graff & Karsten, 2010). However, little research has systematically examined the effectiveness of published articles in guiding correct implementation of the procedures.

Purpose Statement

The main objective in Study 1 was to extend the current literature on enhanced written instructions (e.g., Graff and Karsten, 2012; Ramon et al., 2015) by evaluating the effectiveness of a published practitioner journal article specific to preference assessments (Cannella-Malone, Sanielny, Jimenez, & Miller, 2013) on special education staff’s implementation of the PS and MSWO procedures.

Method

Participants

I recruited six special education staff members (all females, three teachers and three teaching assistants) from a university-affiliated public preschool program that served young children (aged 12 months to 6 years old) with developmental disabilities and peers without disability. Participants were selected based on the following criteria: (a) granting of informed consent, (b) availability, and (c) having no or limited experience in preference assessments. The selected participants ranged in age from 23 to 43 years old (\(M = 29\)) and had one half to six years (\(M = 1.3\)) teaching experience with individuals with disabilities. All participants spoke and read English as their first language. The highest level of education obtained for each participant respectively was a high school
diploma ($n=1$), a bachelor degree ($n=1$), hours beyond a bachelor’s degree ($n=3$), and a master’s degree ($n=1$). The teacher participants, Reed, Alicia, and Pepper, all held state-issued early intervention specialist licenses, and their former training in applied behavior analysis consisted of college courses and in-service professional development events. Two of three teachers (i.e., Reed and Pepper) indicated that they had former experience reading practitioner journal articles related to early childhood education and special education. The teaching assistant participants, Moon, Elsbeth, and Robyn, received no former training in applied behavior analysis and had no prior experience reading practitioner journal articles.

I also recruited four children from the child-care or special education programs at the school. Participants were selected based on the following criteria: (a) parents or legal guardians signed permission forms for the child to participate in the study, (b) the child’s teacher indicated potential benefits for assessing the child’s food or toy preferences, and (c) the teacher reported that the child was able to scan an array of six items and make a selection response upon vocal prompts (e.g., “pick one”). The selected participants ranged from 2 to 6 years old ($M=3.5$). Two participants had no disability, and two participants had a diagnosis of developmental delay and autism spectrum disorder respectively and were receiving special education services at the time of the study.

**Setting and Materials**

One-to-one assessment sessions were conducted individually in a conference room or a classroom at the school. The rooms were furnished with a 1.5 m x 0.7 m x 0.7 m (width x depth x height) desk and chairs. Prior to each session, the experimenter
supplied all the necessary materials to conduct PS and MSWO preference assessments, including paper, writing utensils, a calculator, a timer, paper plates, paper towel, and a designated pool of six food or activity items. During simulated assessments, the experimenter provided the same food items (banana chips, pretzel sticks, gold fish crackers, raisins, popcorn, and peanuts) or activity items (book, Connect Four, MP3, puzzles, crayons, and smartphone) for both assessments. During generalization probes, four participants used a similar pool of six food items (banana chips, pretzel sticks, gold fish crackers, raisins, popcorn, and veggie straws) and one participant used a pool of six toys that the actual student had access to in the student’s own classroom (iPad, Play-dough, popper car, bubbles, markers, and book).

Simulated Assessments

Participants were evaluated on their preference assessment implementation via simulated assessments, during which the first author role-played the part of a student. For each type of preference assessment, a simulated assessment session consisted of 10 trials. A trial was initiated when the participant presented the item(s) on the table and was concluded after the role-played student emitted a specific prescribed response for that trial (i.e., a typical or atypical response). Session length varied for each method of preference assessment, ranging from approximately 3 to 6 min, and for each participant, two to five assessment sessions were conducted per week.

Case scenarios. For each type of preference assessment, I prepared two one-paragraph scenarios that were associated with assessing the role-played child’s food and activity preferences respectively, and I randomly used one of the two scenarios in each
session. The scenarios described the characteristics of the role-played child (i.e., the age, diagnosis, and listener responding level), listed a pool of food or activity items to be used in the assessment, and indicated which assessment was to be conducted. Appendix A shows two scenarios specific to the PS assessment, and Appendix B shows two scenarios specific to the MSWO assessment.

**Scripts for role-played student.** During simulated assessments, the role-played student used scripts that prescribed typical and atypical responses to ensure that all participants experienced a controlled number of assessment trials with atypical responses (i.e., response other than selecting one stimulus). For each type of preference assessment, I prepared four different assessment scripts and randomly assigned one script to each session. During each PS session, I programmed atypical responses on four of 10 trials. The atypical responses included the student engaging in a behavior other than selecting any stimuli within the allotted time (two trials) and selecting more than one stimulus (two trials). On the remaining six trials, the role-played student simply selected one stimulus within 5 s of the participant placing the stimuli on the table and instructing the student to select one stimulus (i.e., typical response). Appendix C shows an example script used in the PS assessment. During each MSWO session, I programmed atypical responses on two of 10 trials. The atypical responses involved the student selecting more than one stimulus (two trials). On the remaining eight trials, the role-played student displayed typical responses as described above. Appendix D shows an example script used in the MSWO assessment.
Response Measurement

On each of the 10 trials per session, observers scored five (the MSWO procedures) or six (the PS procedures) individual target responses as correct, incorrect, or not applicable. Consistent with Graff and Karsten (2012), observers scored a trial as correct if the participant accurately implemented all the applicable target responses. I used percentage of correct trials to indicate participants’ procedural integrity in preference assessment implementation. Operational definitions for individual target responses, as summarized below, were derived from the published practitioner journal article (Cannella-Malone et al., 2013).

**PS preference assessment.** Specifically, observers recorded whether or not participants emitted the following target responses during each trial: (a) stimulus presentation (applicable in all trials), defined as placing two different items on the table in front of the student; (b) stimulus position (applicable in all trials), defined as placing the items approximately 1 ft. (i.e., 6 in. to 1 ft. ½ in.) in front of the student, approximately 1 ft. from one another, and in the positions prescribed by the data sheet (i.e., left or right); (c) vocal direction (applicable in all trials), defined as delivering a vocal prompt to select an item (e.g., “pick one”); (d) postselection response (applicable in eight trials), defined as moving their hands forward towards the student’s hands to block the selection of more than one stimulus and removing the unselected stimulus before recording the student’s selection; (e) trial termination (applicable in two trials), defined as removing all items from the table when the student did not make a selection response within 4-7 s after the delivery of the vocal direction; and (f) preference hierarchy
(applicable in the one trial), defined as ranking the items based on the number of times each item was selected (larger numbers indicating higher rankings of preference), and if there was an equal number of times for several item, determining the ranking based on the trial(s) in which those items were paired together. Appendix E shows an example data sheet for the PS assessment.

**MSWO preference assessment.** For the MSWO procedures, observers specifically recorded whether or not participants emitted the following target responses during each trial: (a) stimulus presentation (applicable in all trials), defined as placing all six different items (Trial 1 and Trial 6) or the unselected items (the remaining trials) in an array in front of the student; (b) stimulus position (applicable in eight trials), defined as placing the unselected items in a rearranged order that was different from the previous trial; (c) vocal direction (applicable in all trials), defined as delivering a vocal prompt to select an item (e.g., “pick one”); (d) postselection response (applicable in all trials), defined as moving their hands forward towards the student’s hands to block the selection of more than one stimulus and removing the unselected stimuli before recording the student’s selection; and (e) preference hierarchy (applicable in the one trial), defined as ranking the items by summing the orders by which each item was selected (smaller totals indicating higher rankings of preference). Appendix F shows an example data sheet for the MSWO assessment.

**Experimental Design and Procedures**

I used a multiple probe design across participants to evaluate the effects of a practitioner journal article on participants’ implementation accuracy with each type of
preference assessment. The mastery criterion for each preference assessment was 100% correct trials across two consecutive sessions.

**Baseline.** At the beginning of each session, the experimenter arranged the furniture such that the participant always sat on one side of the table and the role-played client sat across from the participant. Then the experimenter gave the participant a hard copy of the case scenario and supplied the items necessary to conduct the preference assessment (e.g., writing utensils, a pool of food or activity items). The participant was instructed to read the case scenario and implement the assessment as the scenario required. The experimenter also informed the participant that she would not answer any questions or provide any feedback related to preference assessment procedures until after the completion of the study. Each session was initiated as the participant presented the item(s) on the tabletop and was concluded under one of the following two conditions: (a) the role-played student emitted specific prescribed responses across 10 trials or (b) the participant terminated the assessment by giving a vocal cue (e.g., “I don’t know.” “I am finished.”).

**Practitioner journal article.** At the conclusion of baseline, the experimenter gave a hard copy of the published practitioner journal article specific to preference assessment (Cannella-Malone et al., 2013) to the participant. The peer-reviewed journal article was 8 pages and contained research evidence pertaining to preference assessments, a detailed description of how to implement different types of assessments and interpret assessment results (including the PS and MSWO procedures), sample data sheets for each assessment type, and a troubleshooting table with suggested solutions to atypical
responses during the assessment. I created a reading log and attached it to the journal article as a cover page. The participant was instructed to read the article during the week and to record the date(s) and time(s) spent reading. Appendix G shows the cover page containing the written instruction and the reading log.

The simulated assessment procedures were similar to baseline except that prior to each session, the experimenter provided the participant with a preference assessment data sheet that was the same as the sample in the practitioner journal article (see Appendix H and I). Consistent with Graff and Karsten (2012), the experimenter allowed the participant to refer to the article when conducting the sessions.

**Generalization probes.** The experimenter conducted generalization probes with actual students after the participant achieved the mastery criterion for a specific method of preference assessment. One participant (Robyn) was unable to implement generalization assessments due to limited availability. For the remaining five participants, generalization probes occurred from one to three weeks after the participant achieved mastery with the role-played student. All actual students had previous exposure to the stimuli assessed during the generalization probes. Each generalization probe consisted of 10 trials. At the beginning of the probe session, the experimenter provided the participant with a pool of stimuli to be assessed and a data sheet that was the same as the sample in the practitioner journal article.

**Interobserver Agreement and Procedural Integrity**

Two observers independently scored participant performance (live and by viewing video-recordings) for at least 33% of sessions in all experimental conditions for
both assessment types. An agreement was defined as both observers identically scoring the participant’s implementation of a target response in a trial. Interobserver agreement for each session was calculated as the number of agreements divided by agreements plus disagreements and converted to a percentage. Across participants, mean agreement was 96% (range: 91–100%) for the PS assessment and 94% (range: 93–100%) for the MSWO assessment.

The primary observer also collected data on the integrity of the independent variable and the role-played student’s implementation of the assessment scripts for at least 33% of sessions in all experimental conditions for both assessment types. The observer used a fidelity checklist of the procedural steps to evaluate the integrity of the independent variable (see Appendix J). Integrity was calculated as the number of steps implemented correctly divided by the total number of steps and converted to a percentage. Mean procedural integrity was 100% for both PS and MSWO assessments. The integrity of the role-played student’s performance was evaluated on a trial-by-trial basis. A trial was scored as correct if the role-played student performed the action prescribed on the script, and the integrity for each session was calculated as the number of correct trials divided by the total number of trials (i.e., 10) and converted to a percentage. Mean integrity of the role-played student’s behavior was 99% (range: 90–100%) for the PS assessment and 98% (range: 90–100%) for the MSWO assessment.

Social Validity Assessment

Data on social validity were collected from six participants and the preschool director. The six participants were instructed to rate anonymously their satisfaction with
the training on a rating scale (1 to 5). The questions included ratings of the importance of conducting preference assessments, the clarity of the practitioner journal article, the adequacy of the time allowed for reading the article, the effectiveness of training, and whether the respondents would use the article if they were responsible for training others to conduct preference assessments. For the director of the preschool program, I explained the procedures and outcomes of the study, by providing the practitioner journal article used for training and showing her 1- to 2-min video clips of simulated assessment sessions at the beginning and end of the study. Then the director was asked to complete a questionnaire with eight rating scale questions (5-point rating scale) and two open-ended questions. The rating-scale questions asked the director to rate her agreement with statements about the importance of staff training in preference assessment implementation, the feasibility of using the published article as a training method, and the outcomes of the training. The open-ended questions asked about the director’s opinions on strengths and weaknesses of the training. Appendix K shows the social validity questionnaire for the participants, and Appendix L shows the social validity questionnaire for the preschool director.

**Results and Discussion**

Figures 1 and 2 depict the percentage of trials with accurate implementation of each assessment method for the six participants who experienced baseline (no training) condition followed by the practitioner journal article condition (training) and generalization probes. Performance for both preference assessments remained low ($M = 0\%$) during baseline. Two of 6 participants (Reed and Robyn) gave a vocal cue (i.e., “I
don’t know”) and terminated the PS assessment at the beginning of the simulated sessions, and 5 of 6 participants (Reed, Alycia, Pepper, Moon, and Robyn) terminated the MSWO assessment during baseline in the same manner. Therefore, their performance was scored as 0% across baseline sessions. For the participants who completed 10 trials in simulated assessments, the percentage of correct trials remained low given recurring incorrect responses on specific target steps of the PS and MSWO assessments (i.e., stimulus position, postselection response, and preference hierarchy).

After training, performance on the PS and MSWO assessments improved immediately across all participants. Most participants scored 70% correct trials or above on the first simulated assessment session after training. The exceptions were Alycia on the PS assessment and Pepper on the MSWO assessment. It is possible that the participants reviewed the article several days before actual implementation and therefore it may have been difficult to remember all steps given the time elapsed. This seemed likely because their performance improved in the subsequent sessions as they reviewed the article immediately prior to implementation. At the conclusion of the simulated assessments, mean percentage of correct trials ranged from 90% to 98% across participants for the PS assessment and from 85% to 100% for the MSWO assessment.

All participants met mastery criterion after three to five simulated assessment sessions for the PS assessment and two to six simulated assessment sessions for the MSWO assessment. The results of generalization probes indicated that accuracy on all target responses of assessment implementation remained high ($M = 100\%$) when actual students participated (except for Robyn, for whom generalization probes were not
Figure 1. PS assessment implementation graphs. The solid circles indicate the percentage of trials implemented correctly during role-play sessions, and the open circles indicate participant’s performance during generalization probes.
Figure 2. MSWO assessment implementation graphs. The solid circles indicate the percentage of trials implemented correctly during role-play sessions, and the open circles indicate participant’s performance during generalization probes.
conducted due to teacher availability). According to the participants’ report, the total duration of time spent reading the article before the participant reached mastery criterion averaged 41 min (range: 20–85 min) for the PS assessment and 43 min (range: 20–90 min) for the MSWO assessment.

The results of the social validity assessment indicated that the respondents were quite satisfied with the training. The social validity scores obtained from the participants and the preschool director are summarized in Table 3 and Table 4, respectively. On a scale of 1 to 5 (with 5 being most favorable), the participants rated the importance of conducting preference assessments a mean of 4.8 (Questions 1 and 2), the acceptability of the training a mean of 4.5 (Questions 4, 6, and 7), the outcomes of training a mean of 4.2 (Question 5). In addition, the participants rated the feasibility of implementing preference assessments a mean of 4.0 and 3.3 for the PS and MSWO assessments respectively; a frequently noted concern was the transferability of the preference assessment procedures from a one-on-one setting to a classroom situation. When asked whether they would use the practitioner journal article to train others who had no prior experience with preference assessments, the participants indicated that they would be highly likely to use the article for training staff ($M = 4.4$) and moderately likely to use it for training parents ($M = 3.9$).

In regards to the director of the preschool program, the mean rating on the social validity questionnaire was 4.9 (range: 4–5). She noted that she would use the article for training new staff on preference assessment implementation and would consider the use of other published practitioner journal articles to teach different assessment and instructional procedures. When asked about the strengths and weaknesses of the training
Table 3. Social validity results from the participants—preference assessments

<table>
<thead>
<tr>
<th>Questions</th>
<th>Mean rating (1 = strongly disagree, 5 = strongly agree)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. The assessment focuses on an important student behavior.</td>
<td>4.6 (range: 3–5)</td>
</tr>
<tr>
<td>2. The assessment will produce useful information for planning intervention or instructional programs.</td>
<td>4.8 (range: 4–5)</td>
</tr>
<tr>
<td>3. The assessment can be easily incorporated into my classroom.</td>
<td>4.0 (range: 2–5)</td>
</tr>
<tr>
<td>4. The procedural description in the article is clear and explicit, and I understand the assessment steps.</td>
<td>4.2 (range: 3–5)</td>
</tr>
<tr>
<td>5. After reading the article, I believe I can accurately implement the assessment with my students.</td>
<td>4.4 (range: 4–5)</td>
</tr>
<tr>
<td>6. I have the necessary materials to implement the assessment accurately.</td>
<td>4.6 (range: 3–5)</td>
</tr>
<tr>
<td>7. The time requirement of reading the article is reasonable.</td>
<td>4.8 (range: 4–5)</td>
</tr>
<tr>
<td>8. If I were going to provide training materials to my colleagues who had no previous exposure to this topic, I would recommend using this article.</td>
<td>4.4 (range: 3–5)</td>
</tr>
<tr>
<td>9. If I were going to provide training materials to parents who had no previous exposure to this topic, I would recommend using this article.</td>
<td>3.8 (range: 2–5)</td>
</tr>
<tr>
<td>Questions</td>
<td>Rating</td>
</tr>
<tr>
<td>-----------</td>
<td>--------</td>
</tr>
<tr>
<td>1. Conducting stimulus preference assessments is an acceptable way to identify things that will motivate students with special needs.</td>
<td>5</td>
</tr>
<tr>
<td>2. Other staff members should be trained to implement the procedure.</td>
<td>4</td>
</tr>
<tr>
<td>3. After participating this study, the staff learned how to implement the procedure.</td>
<td>5</td>
</tr>
<tr>
<td>4. Training via practitioner journal articles was practical and feasible.</td>
<td>5</td>
</tr>
<tr>
<td>5. The practitioner journal article served as an effective method for training staff members on the target procedure, in the absence of an expert trainer.</td>
<td>5</td>
</tr>
<tr>
<td>6. Based on the outcomes, I would use the practitioner journal article if I were going to provide training to staff members who had no previous exposure to the target procedure.</td>
<td>5</td>
</tr>
<tr>
<td>7. Based on the outcomes, I would recommend other colleagues to use the practitioner journal article if they were going to provide training to staff members who had no previous exposure to the target procedure.</td>
<td>5</td>
</tr>
<tr>
<td>8. Practitioner journal articles would be helpful in training staff to implement other assessment or instructional procedures.</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 4. Social validity results from the program director—preference assessments
program, the director commented that it was efficient and accessible, and would save time for trainers compared to other training methods, however she was concerned that without the repetitive role-play assessments the training effects might be obscured or discounted.

Overall, the results of the current study replicate previous studies that demonstrated the effectiveness of enhanced written instruction on inexperienced staff’s preference assessment implementation (Graff & Karsten, 2012; Ramon et al., 2015). The current study also provides direct evidence that the practitioner journal article (i.e., Cannella-Malone et al., 2013) may serve as an economical and socially acceptable training resource used to teach special education staff how to correctly perform full-length PS and MSWO preference assessments, including implementing assessment procedures and interpreting assessment results (i.e., selection of high- and low-preference stimuli). With access to the article, all participants’ performance increased from a baseline average of 0% to the mastery criterion (i.e., 100% across two consecutive sessions). None of the participants needed additional training to meet criterion. In addition, probes were conducted with actual students of the participants and the results demonstrated generalization of assessment skills. Consistent with existing enhanced written instructions (Graff & Karsten, 2012; Ramon et al., 2015), the published practitioner article appeared to encompass a full range of discriminative stimuli pertaining to the assessment procedures, which may have controlled correct responding of the assessor during role-play and generalization sessions.
Chapter 4: Single-Case Study Two

Results from Study 1 raised the question of whether practitioner journal articles could be used to train inexperienced staff on behavior-analytic technologies other than preference assessments. Therefore, this chapter presents a replication of Study 1 as applied to a more complex behavior analytic procedure—the system of least prompts. This section includes a brief introduction, a description of the method, a summary of results, and a brief discussion of the findings and its implications.

Staff Training on the System of Least Prompts

Direct teaching procedures that support skill acquisition for individuals with disabilities often involve using prompts to develop stimulus control and transferring stimulus control from prompts to natural stimuli (Cooper et al., 2007; West & Billingsley, 2005). Wolery and Gast (1984) described four procedures for establishing and transferring stimulus control (i.e., constant time delay, progressive time delay, most-to-least prompting, and least-to-most prompting), all of which involve providing response prompts that result in desired behavior and then fading those prompts until the student performs the behavior independently. One such instructional procedure, the system of least prompts (the least-to-most prompting procedure), requires the instructor to define a hierarchy of response prompts and provide the least intrusive prompt necessary to ensure
a correct response (Ault & Griffen, 2013). The system of least prompts usually produces more rapid skill acquisition, as compared to other prompting procedures (e.g., most-to-least prompting) (Libby, Weiss, Bancroft, & Ahearn, 2008), and has been used to teach a variety of discrete and chained skills (e.g., reading, pretend play, daily living activities, and physical activities) to numerous populations (e.g., individuals with autism, intellectual disabilities, and multiple disabilities) (Barton & Wolery, 2010; Browder, Lee, & Mims, 2011; Manley, Collins, Stenhoff, & Kleinert, 2008; Yanarda et al., 2011). If instructors are able to implement the system of least prompts with high integrity, they are likely to be in a better position to produce substantial increases in students’ behavioral outcomes (Wilder, Atwell, & Wine, 2006).

Given the vast literature on using response prompts to teach new skills (see Demchak, 1990, for a review), implementation of systematic prompting procedures are important skills in which to train special education staff who work with individuals with disabilities. Consistent with the findings of the literature review (Chapter 2), previous researchers have predominantly used a combination of antecedent- and consequence-based approaches facilitated by an expert trainer to support correct implementation of the system least prompts (Lerman et al., 2004; Lerman et al., 2008; Schepis, Reid, Ownbey, & Parsons, 2001).

For example, Schepis et al. (2001) provided preschool staff members with a training package consisting of 60- to 90-min classroom-based training (written and vocal instructions, role-play activities, and feedback) and 20-min on-the-job training (performance monitoring and feedback). After receiving the expert-facilitated training,
each of the four staff members improved her use of the system of least prompts and increased the percentage of teaching opportunities with correct implementation in the preschool classrooms.

In a similar approach, Lerman et al. (2008) conducted a 5-day intensive training program to teach three systematic prompting techniques (including the system of least prompts) to special education teachers in public schools. Similar to Schepis et al. (2001), the training program consisted of two parts: classroom-based instructions (i.e., written and vocal instructions) and individual training on the job (i.e., modeling and feedback). Following the conclusion of training, all of the nine teachers reached the mastery criterion and demonstrated skill generality in their own classrooms.

Despite advances in the efficacy of expert-facilitated training, direct training from an expert trainer may not be available in some applied settings (e.g., public schools). In pursuit of effective training resources that may minimize the need of an expert trainer, the current investigation examined whether the system of least prompts could be established and maintained using written instructions from practitioner journal articles.

Given that written instructions alone might not be sufficient to produce mastery criterion performance for all trainees and remedial training would be needed in some cases (e.g., Ramon et al., 2015), the current study employed self-monitoring as a remedial training procedure to meet individual areas of needs. Self-monitoring is a training procedure that has been shown to increase the accuracy with which teachers implement a variety of protocols (e.g., token economies, the Good Behavior Game) in their classrooms when ongoing expert-facilitated support (e.g., performance feedback) may not be feasible.
(Oliver, Wehby, & Nelson, 2015; Plavnick, Ferreri, & Maupin, 2010). Use of a self-monitoring checklist to increase and maintain implementation accuracy on the system of least prompts has a clear benefit. The system of least prompts is manualized (in the practitioner journal article) with clear components to implement. Therefore, the procedure can be articulated into discrete components, making a checklist easy to develop and use in the classroom by teachers to self-evaluate their implementation over time.

**Purpose Statement**

The purposes of Study 2 were to (a) replicate examination of the effects of a published practitioner journal article (Ault & Griffen, 2013) on implementation as applied to a more complex procedure—the system of least prompts and (b) examine the effects of a remedial training procedure that requires minimal expert-facilitated support (i.e., self-monitoring) in increasing post-training procedural integrity levels. During the investigation, if the participant’s performance did not reach the mastery criterion after reading the practitioner journal article specific to the target procedure, I provided a supplemental practitioner journal article which aimed to guide the participant in the use of self-monitoring checklists to maintain implementation accuracy (McKenna, Flower, & Ciullo, 2014).

**Method**

**Participants**

The six special education staff members who participated in Study 1 agreed to participate in the current experiment. The participants’ characteristics and background
were described in Study 1. I also recruited five students from the child-care or special education programs at the preschool. The students were selected based on the following two criteria: (a) parents or legal guardians signed permission forms for their child to participate in the study, and (b) the student’s teacher indicated potential benefits for teaching the student a chained skill and the target chained skill was either recommended by the teacher or selected from the student’s current IFSP or IEP goals. The selected students ranged from 2 to 6 years ($M = 3.4$). Three students had no disability, and two students were diagnosed with developmental delay and autism spectrum disorder respectively and were receiving special education services at the time of study. The target chained skills were different across students, including zipping the jacket (five steps), putting on socks and shoes (five steps), cutting a line with scissors (five steps), sequencing numbers 1 to 9 (nine steps), and tracing letter E using a dry erase marker (five steps).

**Setting and Materials**

The study was conducted in settings similar to those described in Study 1. Prior to each session, the experimenter provided teachers with the materials necessary to instruct a chained skill using the system of least prompts, including paper, writing utensils, a calculator, a timer, and items specific to the chained skills (e.g., the experimenter supplied a zip-up jacket when the chained skill to be taught was zipping the jacket).

**Simulated Instruction**

Similar to Study 1, I evaluated participants’ implementation of the system of least prompts by observing their interactions with a role-played student. Each simulated
instruction session consisted of 10 trials, and during each session, the participant was required to teach the chained skill “zipping the jacket” on two occasions. A trial involved the instruction of one step of the chained skill using the least-to-most prompting procedure. The trial was initiated when the participant delivered a vocal direction for performing the skill (e.g., “It’s time to zip your jacket!”) (Trial 1 and Trial 6) or provided a prompt specific to a step (e.g., a vocal prompt such as “latching the two parts of the zipper together”) (the remaining trials), and the trial was concluded after the participant delivered the reinforcer contingent upon the student’s independent or prompted completion of that step. The length of a simulated instruction session ranged from 2 to 5 min, and for each participant, four to six instruction sessions were conducted per week.

**Case scenario.** I prepared a one-paragraph scenario that described the characteristics of the role-played student (i.e., the age, diagnosis, listener responding level, preferred form of social attention, and the most intrusive prompt that usually resulted in correct responding) and the task analysis of the chained skill “zipping the jacket” (i.e., a list of five teachable steps). The scenario also indicated that the participant was required to list three types of prompts to be used in a least-to-most sequence, teach the skill on two occasions using the system of least prompts, collect data on role-played student’s responses, and graph data using a weighted point system. Appendix M shows the case scenario provided to the participant in role-play sessions.

**Scripts for role-played student.** In order to ensure that all participants experienced a controlled number of trials with independent (i.e., the student completed a step correctly without any prompt) and prompted responses (i.e., the student completed a
step correctly following a certain prompt), I prepared three different scripts and randomly assigned one script to each session. Across all the scripts, I programmed independent responses on two of 10 trials. On the remaining eight trials, the role-played student engaged in a correct response following the least intrusive prompt (two trials), the intermediate prompt (three trials), or the most intrusive prompt (three trials). Appendix N shows an example script used in role-play sessions. Prior to each instruction session, the experimenter asked the participant to state the prompt sequence to be used, so that the role-played student could emit prescribed responses following designated prompts.

**Response Measurement**

I used percentage of correct trials to determine participants’ procedural integrity, in the same manner as described in Study 1. On each trial, observers scored seven individual target responses as correct, incorrect, or not applicable. Operational definitions for individual responses were derived from the published practitioner journal article (Ault & Griffen, 2013): (a) vocal direction (applicable in two trials), defined as deliver a vocal direction for performing the chained task (e.g., “It is time to zip your jacket”); (b) delayed least intrusive prompt (applicable in eight trials), defined as giving the student 3 to 5 s to respond before giving a least intrusive prompt for the target step; (c) delayed intermediate prompt (applicable in six trials), defined as giving the student 3 to 5 s to respond before giving a prompt that is more intrusive than the previous prompt and less intrusive than the next prompt; (d) delayed most intrusive prompt (applicable in three trials), defined as giving the student 3 to 5 s to respond before giving the most intrusive prompt that enables the student to complete the step; (e) reinforcer delivery (applicable in all trials), defined
as giving the student the preferred form of social attention (e.g., praise, high-fives) contingent upon independent or prompted completion of the step; (f) data collection (applicable in all trials), defined as correctly marking the student’s response as independent or prompted, if prompted, correctly recording the prompt that ultimately leads to correct response for the step; and (g) graphing (applicable in two trials), defined as correctly calculating and graphing the student’s performance using a weighted point system. Appendix O shows an example data sheet used to score the procedural integrity of the participant’s implementation.

In order to decide whether the participant’s prompts were in a correct least-to-most sequence, the observer referred to a prompt hierarchy identified by the published articles (Ault & Griffen, 2013; van Vonderen, de Swart, & Didden, 2010). The prompt hierarchy consisted of (a) indirect verbal prompt, defined as a vocal prompt without specifying what behavior should be performed (e.g., “What is next?”); (b) visual prompt, defined as a gestural prompt or a pictorial prompt (e.g., the instructor pointed to the zipper or presented a picture of a person grasping the zipper); (c) direct verbal prompt, defined as a vocal prompt that specifically tells the student the step to do (e.g., “Grasp the zipper”); (d) model prompt, defined as the instructor modeling the step alive (e.g., the instructor grasped the zipper); (e) direct verbal plus model prompt, defined as a direct vocal prompt and an in-vivo model prompt (e.g., the instructor said “Grasp the zipper” while grasping the zipper on the jacket); and (f) direct verbal plus physical prompt, defined as a direct vocal prompt and a hand-on-hand physical prompt that enables the student to complete the step (e.g., the instructor said “Grasp the zipper” while taking the
student’ hand to hold the zipper). A least-to-most prompt sequence was considered as correct if the participant selected any three types of prompt and delivered them in an order as described above (e.g., “indirect verbal prompt” being the least intrusive prompt, “model prompt” being the intermediate prompt, and “direct verbal plus model prompt” being the most intrusive prompt).

**Experimental Design and Procedures**

Consistent with Study 1, I used a multiple probe design across participants to evaluate the effects of practitioner journal articles on participants’ implementation of the system of least prompts. The mastery criterion was 100% correct trials across two consecutive sessions.

**Baseline.** The procedures were similar to Study 1, except that (a) prior to each simulated instruction session, the experimenter supplied the items necessary to implement the system of least prompts (e.g., writing utensil, items specific to the chained skill), and (b) each session was concluded after the participant taught the chained skill on two occasions.

**Practitioner journal article (the system of least prompts).** Similar to Study 1, at the conclusion of baseline, the experimenter gave the participant a hard copy of the published practitioner journal specific to the system of least prompts (Ault & Griffen, 2013) and a reading log for participant to track the date(s) and time(s) spent reading during the week. Appendix P shows a cover page that is attached to the published article and contains a reading log. The 8-page article contained research evidence pertaining to the practice, two case examples on how to implement the practice and graph data using a
weighted point system, sample data sheets, and a table with suggested practice modifications to different student response patterns. During the simulated instruction session, the experimenter supplied a blank data sheet that was similar to the sample in the article (see Appendix Q) and allowed the participant to access the article during implementation.

Practitioner journal article (self-monitoring). If the participant’s performance averaged no more than 60% or displayed a flat or declining trend on the first four sessions after training, the experimenter provided another published practitioner journal article pertaining to measuring treatment fidelity (McKenna, Flower, & Ciullo, 2014). The 6-page article contained rationale for monitoring and maintaining treatment fidelity, sample fidelity checklists, and a detailed description of how to measure fidelity (using self-monitoring). The participant was instructed to refer to the two articles (i.e., Ault & Griffen, 2013; McKenna et al., 2014) and develop a fidelity checklist on the system of least prompts within a week. In addition, the participant was required to record the date(s) and time(s) spent reading each article and developing a checklist. Appendix R shows a cover page containing the instruction and a reading log.

During the simulated instruction session, the participant implemented the system of least prompts with a role-played student, in the same manner as described in the previous condition, and was instructed to complete the fidelity checklist (developed by the participant) following each session. The completed checklists were collected by the observer following observations.
**Generalization probes.** Generalization probes with actual students were conducted during baseline and after the participant achieved the mastery criterion with a role-played student. Each participant was instructed to work with an actual student and teach a new chained skill: “zipping the jacket” for Robyn, “putting on socks and shoes” for Pepper, “cutting a line with scissors” for Alycia, “sequencing numbers 1 to 9” for Moon and Reed, and “tracing letter E using a dry erase marker” for Elsbeth. The experimenter asked the student’s teacher to develop a task analysis of the target skill and ensured that the participant had access to the task analysis prior to generalization probes. Each probe session consisted of five to nine trials, depending on the number of steps the skill consisted of. At the beginning of each probe session, a blank data sheet that was similar to the sample provided in Ault and Griffen (2013) was supplied, and the experimenter asked the participant to state the prompt sequence to be used. For participants whose performance achieved mastery with self-monitoring (Elsbeth and Pepper), a fidelity checklist developed by the participant was also supplied. During the session, the experimenter allowed the participant to refer to Ault and Griffen (2013) as needed.

**Interobserver Agreement and Procedural Integrity**

Interobserver agreement, procedural integrity of the independent variable, and procedural integrity of the role-played student’s performance were assessed in the same manner as described in Study 1. Mean interobserver agreement ranged from 96% to 99% across participants, with an average score of 98%. Mean procedural integrity was 100% for the independent variable and 98% (range: 90–100%) for the role-played student’s
performance. Appendix S shows the procedural integrity checklists used during baseline and training.

**Social Validity Assessment**

Prior to training, six participants were instructed to complete an anonymous questionnaire with five questions using a 5-point rating scale. The questions were designed to evaluate the extent to which they found the strategy (i.e., the system of least prompts) useful and how confident they were in implementing the strategy accurately prior to training. Appendix T shows the social validity questionnaire provided to the participants prior to training. As the participants achieved the mastery criterion after reading the practitioner journal article pertaining to the practice (i.e., Ault & Griffen, 2013), I collected social validity data among the participants and the preschool director in the same manner as described in Study 1. The participants completed the questionnaire anonymously and the director completed the questionnaire after reviewing the published articles used for training and the video clips of role-play sessions. Appendix T and U show the social validity questionnaires provided to the participants and the director, respectively, after training. For the participants who met the criterion after self-monitoring, I asked them, in open-ended questions, about their satisfaction with the self-monitoring procedure. The questions included the clarity of the article on self-monitoring (i.e., McKenna et al., 2014) and the usefulness of self-monitoring on their implementation of the system of least prompts. Appendix T shows open-ended questions presented to the participants who achieved mastery after self-monitoring.
Results and Discussion

Figure 3 depicts the results for participants’ performance during role-plays with adults and teaching sessions with actual students. All participants demonstrated low procedural integrity during baseline. Mean accuracy across participants ranged from 0% to 7.5% during role-play sessions and remained 0% during generalization probes. For each participant, a few reoccurring implementation errors were observed, including (a) the participant used a prompt sequence consisting of two prompts only (Elsbeth, Robyn, and Reed), (b) the participant skipped the step(s) in task analysis and delivered incorrect step-specific prompts (Pepper, Alycia, Moon, and Robyn), and (c) the participant did not collect data on the student’s responses (all participants).

An immediate increase was observed in each participant’s performance following the introduction of the practitioner journal article (i.e., Ault & Griffen, 2013). Mean accuracy of performance across participants increased to 76.5% (range: 38–96%). The procedural integrity of Pepper and Moon immediately increased to 90% and 80% on the first session following training and met the criterion (i.e., 100% correct trials across two consecutive sessions) in three and four role-play sessions, respectively. Performance of Alicia and Robyn increased moderately on the first session following training (i.e., increased to 50% and 30%, respectively), and further increases were observed in subsequent sessions (M = 95% and 93%, respectively). Compared to Pepper and Moon, it took Alicia and Robyn more role-played sessions before they achieved the criterion. It is possible that the chained skill to be taught during role-plays was different from that described in the practitioner journal article and therefore it may have been taking more
Figure 3. The system of least prompts implementation graphs. The solid circles indicate the percentage of trials implemented correctly during role-play sessions, and the open circles indicate participant’s performance during generalization probes.
time for some participants to get familiar with the steps in the task analysis. The results of
generalization probes indicated that after the four participants achieved the mastery
criterion on the system of least prompts, their teaching skills generalized to an actual
student and a new chained behavior \((M = 100\%)\). Based on the participants’ report, the
total duration of time spent reading the article before the participant reached mastery
criterion averaged 58 min (range: 38–82 min).

Two of 6 participants (Elsbeth and Reed), whose performance increased slightly
after reading the article but did not reach mastery, met the mastery criterion in three to
four sessions after self-monitoring. As self-monitoring was introduced, Elsbeth’s
performance in role-play sessions increased from an average of 38\% (range: 30–50\%) to
93\% (range: 60–100\%) and Reed’s performance increased from an average of 58\%
(range: 50–80\%) to 92\% (range: 70–100\%). The results of generalization probes
indicated that with self-monitoring both of the participants accurately implemented the
procedure with an actual student and on a new chained behavior \((M = 100\%)\). The two
participants, respectively, reported that they spent a total duration of 55 min and 50 min
reading the article pertaining to the system of least prompt, 20 min and 40 min reading
the article pertaining to self-monitoring, and 20 min developing the procedural checklist
used for self-monitoring.

The results of the social validity assessment suggested that the respondents found
the training program a socially acceptable method for teaching staff how to implement
the system of least prompts. Table 5 summarizes the social validity scores obtained from
the participants during baseline \((n = 6)\) and after demonstrating mastery as the
practitioner journal article was introduced \((n = 4)\). Before the training, the participants rated the importance of the target skills (i.e., implementing the system of least prompts) a mean of 4.4 (Questions 1 and 2) and the confidence in accurate implementation a mean of 2.8 (Question 5). After the training, the participants’ ratings on the skill importance remained high \((M = 4.9)\), and the ratings on confidence in implementation increased substantially \((M = 4.2)\). In addition, all the participants agreed or strongly agreed with the statements pertaining to the training procedures, with an average rating of 4.7 (Questions 4, 6, and 7). Similar to the results of Study 1, the participants indicated that they would be highly likely to use the article for training staff and moderately likely to use it for training parents. The two participants who achieved mastery after self-monitoring commented favorably on the content of the supplemental article (i.e., McKenna et al., 2015) and the usefulness of self-monitoring during implementation. They noted that they would like to use self-monitoring to support the implementation of other interventions. Furthermore, the school director strongly agreed with all the statements pertaining to the training, including its goals, procedures, and outcomes (see Table 6). The comments regarding the most- and least-liked aspects of the training, respectively, centered upon the efficiency of the training and the repetitive role-play assessments.

Overall, these results extend previous studies (Graff & Karsten, 2012; Ramon et al., 2015) by providing empirical support for the use of enhanced written instructions to teach implementation of behavior analytic procedures other than preference assessments. After reviewing the published practitioner journal article (i.e., Ault & Griffen, 2013), 4 of 6 participants achieved mastery on implementing the system of least prompts and
<table>
<thead>
<tr>
<th>Questions</th>
<th>Mean rating (1 = strongly disagree, 5 = strongly agree)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Before Training</td>
</tr>
<tr>
<td>1. The intervention focuses on an important student behavior.</td>
<td>4.2 (range: 1–5)</td>
</tr>
<tr>
<td>2. The intervention will improve the student’s behavioral outcomes.</td>
<td>4.6 (range: 4–5)</td>
</tr>
<tr>
<td>3. The intervention can be easily incorporated into my classroom.</td>
<td>5.0 (range: 5–5)</td>
</tr>
<tr>
<td>4. The procedural description in the article is clear and explicit, and I understand the intervention steps.</td>
<td>N/A</td>
</tr>
<tr>
<td>5. After reading the article, I believe I can accurately implement the intervention with my students.</td>
<td>2.8 (range: 1–5)</td>
</tr>
<tr>
<td>6. I have the necessary materials to implement the intervention accurately.</td>
<td>4.2 (range: 3–5)</td>
</tr>
<tr>
<td>7. The time requirement of reading the article is reasonable.</td>
<td>N/A</td>
</tr>
<tr>
<td>8. If I were going to provide training materials to my colleagues who had no previous exposure to this topic, I would recommend using this article.</td>
<td>N/A</td>
</tr>
<tr>
<td>9. If I were going to provide training materials to parents who had no previous exposure to this topic, I would recommend using this article.</td>
<td>N/A</td>
</tr>
</tbody>
</table>

Table 5. Social validity results from the participants—the system of least prompts
<table>
<thead>
<tr>
<th>Questions</th>
<th>Rating</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Preference Assessments</strong></td>
<td></td>
</tr>
<tr>
<td>1. Using the system of least prompts is an acceptable way to promote skill acquisition and reduce prompt dependency for students with special needs.</td>
<td>5</td>
</tr>
<tr>
<td>2. Other staff members should be trained to implement the procedure.</td>
<td>5</td>
</tr>
<tr>
<td>3. After participating this study, the staff learned how to implement the procedure.</td>
<td>5</td>
</tr>
<tr>
<td>4. Training via practitioner journal articles was practical and feasible.</td>
<td>5</td>
</tr>
<tr>
<td>5. The practitioner journal articles served as an effective method for training staff members on the target procedure, in the absence of an expert trainer.</td>
<td>5</td>
</tr>
<tr>
<td>6. Based on the outcomes, I would use the practitioner journal articles if I were going to provide training to staff members who had no previous exposure to the target procedure.</td>
<td>5</td>
</tr>
<tr>
<td>7. Based on the outcomes, I would recommend other colleagues to use the practitioner journal articles if they were going to provide training to staff members who had no previous exposure to the target procedure.</td>
<td>5</td>
</tr>
<tr>
<td>8. Practitioner journal articles would be helpful in training staff to implement other assessment or instructional procedures.</td>
<td>5</td>
</tr>
</tbody>
</table>

Table 6. Social validity results from the program director—the system of least prompts
demonstrated generalization to actual students and untrained target skills. The practitioner journal article seemed to offer several training advantages, including a detailed description of desired behavior in relevant contexts, use of multiple stimulus and response exemplars, and standardized presentation of training that permits consistency. However, 2 of 6 participants required remedial training (i.e., using a self-monitoring checklist) to demonstrate mastery and generalization of their teaching skills. The positive effects of self-monitoring found in the current study may be interpreted in relation to developing and using a checklist. First, it is possible that developing a self-monitoring checklist allowed the participants to closely attend to the critical stimuli (in the published article) that pertained to the instructional procedures. Second, the fact that the participants’ procedural integrity increased during self-monitoring suggests that their responding may have been under the joint stimulus control of the students’ behavior as well as the integrity checklist (Oliver et al., 2015; Plavnick et al., 2010).
Chapter 5: General Discussion

The two studies reported here systematically extend the findings of Graff and Karsten (2012) and Ramon et al. (2015). The results demonstrated that the published practitioner journal articles (i.e., Ault & Griffen, 2013; Cannella-Malone et al., 2013) may serve as effective written instructions in the acquisition of preference assessment and the system of least prompts. The ratings obtained with the social validity assessment were uniformly positive, which supports the appeal of the practitioner journal articles to school-based consumers. The use of practitioner journal articles may offer several training advantages. First, it may serve as a vehicle for training special education staff and minimizing the need for an expert trainer, and the savings in trainer time and cost that accrue through the use of the practitioner articles could be used to obtain additional practitioner resources (e.g., other practitioner journal articles) and provide remedial training in individualized areas of need. Second, the articles have standardized presentations and can be reused, permitting consistent staff training across a range of instructional settings. Third, compared to existing enhanced written instructions, published practitioner journal articles seem to be more typically available in some applied settings such as schools.
Results from the investigation also suggested that continuous access to the written instructions might contribute to high levels of procedural integrity over time. This finding was consistent with the results of Graff and Karsten (2012). In each of the single-case studies, participants were given the practitioner journal articles one week before role-play assessments and had access to the articles while they conducted sessions. Anecdotally, quite a few participants referred to the articles immediately before and during a session. Therefore, it is possible that participants used the articles not only to acquire information about how the procedure was implemented but also as a prompt to guide the assessment or instructional steps. Furthermore, the participants in the studies worked in a preschool in which the system of least prompts was performed on a regular basis while preference assessments were not. The results indicated that ongoing access to the instructions would enable accurate implementation over time, even if the trainees implement the procedure infrequently at the workplace. Future researchers may want to systematically manipulate the level of access to the materials (e.g., single-opportunity access vs. continuous access) and examine the resultant effects on trainee performance.

The necessity for remedial training (i.e., self-monitoring) for a subset of participants substantiates the importance of repeated performance-based assessment and data-based decisions as components of effective staff training programs (e.g., Petscher & Bailey, 2006; Ramon et al., 2015). The instructional materials (i.e., the published practitioner journal articles) were provided to the participants in varying intensities based on participant needs, mirroring a tiered support model as applied to professional development for school-based practitioners (e.g., Myers, Simonsen, & Sugai, 2011;
Thompson, Marchant, Anderson, Prater, & Gibb, 2012). After reading one specific practitioner journal article, all six participants reached the mastery criterion on preference assessments, however, a significant portion of participants did not meet the criterion on the system of least prompts and required self-monitoring as remedial implementation support. It appeared from the results that the effects of the published practitioner journal articles on procedural implementation may differ, which could have been a result of the complexity of the procedures and the clarity with which the procedures are described. Therefore, school-based trainers may want to strategically employ the published practitioner articles and adapt their training methods as needed by different trainees and different procedures.

Furthermore, results of the current investigation demonstrated that effective remedial training was not necessarily facilitated by an expert trainer and, alternatively, that trainees may benefit from remedial training in a self-directed instructional context. In previous studies, remedial training such as modeling (e.g., Ramon et al., 2015), prompting (e.g., Petscher & Bailey, 2006), and corrective feedback (e.g., DiGennaro-Reed, Codding, Catania, & Maguire, 2010) was primarily trainer-directed. There have been a few successful demonstrations of remedial training that minimizes the need of a trainer, such as self-monitoring (e.g., Plavnick et al., 2010), however, trainer involvement was still necessary in those studies. Specifically, for self-monitoring, previous researchers needed a trainer to create the self-monitoring checklist, review the checklist with the trainee, and explain or demonstrate the procedure of self-monitoring (Oliver et al., 2015; Petscher & Bailey, 2006; Plavnick et al., 2010). During the current investigation, I used
the practitioner journal article (i.e., McKenna et al., 2015) as a replacement to trainer-directed training on self-monitoring. With the preparation afforded by the article, the participants created the procedural checklists by themselves and performed self-monitoring that eventually improved the implementation of the procedure. Future research is needed to examine the actual effects of the article on trainees’ skills in developing and using self-monitoring materials, given that trainees’ skills may differ before the introduction of the article.

Results of the current studies also raise the question of whether the published practitioner journal articles could be used to train practitioners other than in-service special education staff on preference assessments and the system of least prompts. The ratings obtained with the social validity assessments suggested that the participants were more likely to recommend the articles to their colleagues than to a student’s parents. It is possible that most of the participants (except for Reed) were not required to conduct parent training as part of their jobs, and therefore the motivational variables may have affected their responding to the questionnaire. Future research should examine whether inexperienced practitioners such as parents, behavior technicians, and pre-service teachers could acquire assessment or teaching skills using the published articles. Also, it may be informative to ask in-service special education staff which specific features of the articles they found most- and least-liked if they were going to provide the articles, as instructional materials, to parents of their students.

The findings of this study are preliminary and should be interpreted with caution given some limitations. First, some participants’ performance demonstrated variability
before reaching the mastery criterion. It was unclear what factors accounted for the variability of responding. This may have been due to competing demands (e.g., the participant was required to attend closely to student responding as well as written information in the data sheet) or selective stimulus control (e.g., the participant’s responding was guided by a subset of, rather than all of, the critical stimuli in the article). Their performance eventually reached mastery and demonstrated stability, however, there could be uncontrolled variables (combined with the articles) that were responsible for their performance increases. Although explicit performance feedback was not provided to participants, it is possible that being allowed to access the articles for a longer period of time and being asked to perform the procedure again may have served as a subtle form of feedback. Future researchers may want to isolate this variable to examine whether this form of feedback would affect participants’ performance.

Second, although results of the two studies showed that the practitioner journal articles may be used to guide correct implementation of the procedures, it remains unknown what critical components of the articles were functionally related to practitioners’ performance increases. Graff and Karsten (2012) conducted a component analysis of the data sheet taken from their enhanced written instructions. Following the introduction of the data sheet alone, increased accuracy was observed in 7 of 10 cases, suggesting that the data sheet may be an important training component. Future research could extend this line of inquiry and determine quality indicators of published practitioner journal articles that are essential to training efficacy. It may also be

80
informative to ask practitioners which specific features of the articles they would find most helpful and assess social validity of each critical training component separately.

Third, as compared to existing methods of integrity analysis (e.g., the percentage of correct steps, a weighted point system), a more stringent method was used in the current studies. Specifically, participants’ procedural integrity was analyzed as the percentage of trials in which all target responses in the trial were implemented correctly. In other words, if any target response on a given trial was implemented incorrectly, the entire trial was scored as incorrect. This method of analysis is consistent with the precedent (i.e., Graff & Karsten, 2012) and has applied relevance in that correct implementation of the entire contingencies (e.g., correct antecedent setup and delivery of differential consequences) may be important to successfully affecting student behavior. However, it may have run the risks of underestimating procedural integrity and misrepresenting the gains participants made. For example, this “all-or-nothing” method may not be sensitive enough to detect performance changes on specific responses that could be essential to student outcomes. Future researchers may want to (a) identify the components of a given procedure that are functionally related to student outcomes, (b) compare the validity, accuracy, and sensitivity of different methods of integrity analysis, and (c) refine existing methods so that they detect important, meaningful changes during implementation.

Despite these limitations, the two studies extend previous research by demonstrating the efficacy of the published practitioner journal articles for guiding and maintaining accurate implementation of preference assessments and the system of least
prompts in a public school setting. In addition, the participating teachers and teaching assistants demonstrated generalization of acquired skills when working with their students in the classroom. In conclusion, published practitioner journal articles may serve as effective and efficient training resources to increase teachers’ procedural integrity with reduced reliance on expert-facilitated staff training. Future research is warranted to replicate examination of the effects of published practitioner journal articles (a) on implementation of different assessment or instructional procedures, (b) among trainees with different professional or educational backgrounds, and (c) on critical professional behavior other than procedural integrity, such as identifying evidence-based interventions, integrating client values and context, and data-based decision-making.
References

Asterisks indicate the articles was included in the Chapter 2 literature review.


89


92


93


Appendix A: Two Case Scenarios for the PS Assessment
Alpha is a 4-year-old boy with moderate intellectual disability. Alpha is non-verbal. He can follow one-step verbal instructions and point objects as means of communication.

The behavior support team would like to identify highly-preferred snack and incorporate it/them to Alpha’s behavior support plan.

Mom reports that Alpha seems to like the following food items:

- Gold fish cracker
- Banana chips
- Pretzel sticks
- Raisins
- Peanuts

Please administer a paired-stimulus preference assessment and list Alpha’s preference hierarchy for snack. The role-played Alpha will let you know when to start and end.
Beta is a high school student with Down Syndrome. Beta can communicate using speech and follow one-step verbal instructions. The behavior support team would like to identify highly-preferred leisure activities and incorporate it/them to Beta’s behavior support plan.

The teaching assistant reports that Beta seems to enjoy the following activities:

- Playing smartphone
- Playing Connect 4
- Drawing pictures
- Reading books
- Listening to music

Please administer a **paired-stimulus preference assessment** and list Beta’s preference hierarchy for leisure activities. The role-played Beta will let you know when to start and end.
Appendix B: Two Case Scenarios for the MSWO Assessment
Gamma is a 6-year-old elementary school student with Autism Spectrum Disorders.

Gamma is non-verbal. She can follow one-step verbal instructions and point objects as means of communication. The behavior support team would like to identify highly-preferred snack and incorporate it/them to Gamma’s behavior support plan.

Mom reports that Gamma seems to like the following food items:

- Gold fish cracker
- Banana chips
- Pretzel sticks
- Raisins
- Peanuts
- Popcorn

Please administer a **multiple stimulus without replacement preference assessment** and list Gamma’s preference hierarchy for snack. The role-played Gamma will let you know when to start and end.
Delta is a 10-year-old student with moderate-to-severe intellectual disability. Delta can communicate using single words and follow one-step instructions. The behavior support team would like to identify highly-preferred leisure activities and incorporate it/them to Delta’s behavior support plan.

The teaching assistant reports that Delta seems to enjoy the following activities:

- Playing Smartphone
- Playing Connect 4
- Drawing
- Reading books
- Listening to music
- Playing Spiderman puzzle

Please administer a multiple-stimulus without replacement preference assessment and list Delta’s preference hierarchy for leisure activities. The role-played Delta will let you know when to start and end.
Appendix C: An Example Script for the PS Assessment Role-Play Sessions
**PS—Alpha (pointing)**

prep:

- Written expectations/Scenario
- Pens and Paper
- Timer (cell phone)
- Calculator
- Plates and snack items
  - Goldfish
  - Bana chips
  - Pretzel sticks
  - Raisins
  - Peanuts

## Script 1

<p>| | | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
</table>
| 1 | Left only      | 6 | No selection (**Atypical**)
| 2 | Right, wait 3s, Left (**Atypical**) | 7 | Left only   |
| 3 | Right only     | 8 | Left only      |
| 4 | Left only      | 9 | Right only     |
| 5 | Right, wait 3s, Left (**Atypical**) | 10 | No selection (**Atypical**) |
Appendix D: An Example Script for the MSWO Assessment Role-Play Sessions
**MSWO—Delta (single word)**

**prep:**

- Written expectations/Scenario
- Pens and Paper
- Timer (cell phone)
- Calculator
- Leisure activities
  - Smartphone
  - Connect 4
  - Crayon and paper
  - Book – German Village
  - MP3 Players and headphone
  - Spiderman puzzle

**Script 1**

<table>
<thead>
<tr>
<th></th>
<th>Connect 4</th>
<th></th>
<th>Connect 4, wait 3s, Another Item (<strong>Atypical</strong>)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
<td>6</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Smartphone, wait 3s, Another Item (<strong>Atypical</strong>)</td>
<td>7</td>
<td>Smartphone</td>
</tr>
<tr>
<td>3</td>
<td>Book</td>
<td>8</td>
<td>Book</td>
</tr>
<tr>
<td>4</td>
<td>MP3</td>
<td>9</td>
<td>Spiderman puzzle</td>
</tr>
<tr>
<td>5</td>
<td>Crayon</td>
<td>10</td>
<td>MP3</td>
</tr>
</tbody>
</table>
Appendix E: An Example Data Sheet for Evaluating PS Implementation
Assessment Criteria for PS—*SCRIPT1 Alpha*

Name of observer:  
Participant:  
Video tape label:  
Phase and session:  

1= correct 0 = incorrect

If one of the target behaviors is incorrect, the entire trial should be counted as incorrect. In some trials, Post-selection response OR Trial termination may be N/A (based on the simulated client’ scripts).

<table>
<thead>
<tr>
<th>Trials-PS</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stimulus Presentation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Placing two corresponding items on the table (Empty data sheets with items listed will be ready)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Incorrect if the teacher placed a smaller or larger number of stimuli on the tabletop.</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stimulus Position</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Corresponding positions as indicated in the data sheet “Within the client’s reach”; “Approx. 1 foot apart” (6 inch – 1.5 ft).</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Saying “Pick one”</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct if the response occurs after stimulus presentation and position are completed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Postselection response</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removing unselected stimulus and THEN circling the corresponding number</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
| *Correct if “removing unselected stimulus” occurs immediately after “providing access”, within 0-2 sec  
*Incorrect if recording occurs before removing items* |   |   |   |   |   |   |   |   |   |    |
| **Trial termination** |   |   |   |   |   |   |   |   |   |    |
| Ending trial after 4-7 sec of no response (counting from “pick one”) by removing both items and scoring the trial as none |   |   |   |   |   |   |   |   |   |    |
| **Listing the preference hierarchy** |   |   |   |   |   |   |   |   |   |    |
| *If two items receive the same points, they can be in either order as long as the items higher and lower than them are correct.* |   |   |   |   |   |   |   |   |   |    |
Appendix F: An Example Data Sheet for Evaluating MSWO Implementation
Assessment Criteria for MSWO—SCRIPTS 1 Delta

Name of observer: 
Video tape label: 
Participant: 
Phase and Session: 

1 = correct 0 = incorrect

If one of the target behaviors is incorrect, the entire trial should be counted as incorrect. The 10 trials are 2 MSWO sessions, each session with the same 6 items. The last response occurs after hypothetical data are provided.

<table>
<thead>
<tr>
<th>Trials-MSWO</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
<th>7</th>
<th>8</th>
<th>9</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Stimulus Presentation</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial 1 and 6: placing all six different items in front of the consumer “at the same time, generally in a straight line”</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other trials: placing the <strong>unselected</strong> items in an array</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Incorrect if the teacher placed a smaller or larger number of stimuli on the tabletop</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Incorrect if replacing the selected item in the stimulus array or removing a nonselected item from the array</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Stimulus Position</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Trial 1 and 6 are not applicable.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other trails: <strong>rearranging the order</strong> (i.e., at least changing the positions of two items) and presenting the array on the table</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Saying “Pick one”</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Correct if the response occurs after stimulus presentation is completed.</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Postselection response</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Removing unselected stimulus and THEN recording the order in the session column next to the item the client chose</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Correct if “removing unselected stimulus” occurs immediately after “providing access”, within 0-2 sec</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><em>Incorrect if recording occurs before removing items</em></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Listing the preference hierarchy</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix G: A Cover Page with Reading Log—Preference Assessments
Hello ______________________ (please fill in your name here)!

Please take one week to closely read the attached article “Pick Me! Conducting Preference Assessments With Students With Significant Disabilities.”

One week later, you will be asked to use paired stimulus preference assessment and multiple stimulus without replacement assessment to assess the preferences of a role-played student. One of the two data sheets will be provided to you at the beginning of the role play sessions; one is similar to Figure 3 (Page 20), for paired stimulus preference assessment and the other is similar to Figure 4 (Page 21), for multiple stimulus without replacement assessment.

Please use the table below to track the dates and durations you read this article.

<table>
<thead>
<tr>
<th>Date</th>
<th>Length of time spent in reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>Example →</td>
<td>2/29/2016 45 min</td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix H: PS Data Sheets for Participant Use
### Items (circle selection)

<table>
<thead>
<tr>
<th>Items</th>
<th>Number of Times Chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gold Fish Cracker</td>
<td></td>
</tr>
<tr>
<td>Banana chips</td>
<td></td>
</tr>
<tr>
<td>Pretzel sticks</td>
<td></td>
</tr>
<tr>
<td>Raisins</td>
<td></td>
</tr>
<tr>
<td>Peanuts</td>
<td></td>
</tr>
</tbody>
</table>

### Items (circle selection)

<table>
<thead>
<tr>
<th>Items (circle selection)</th>
<th>Left</th>
<th>Right</th>
<th>None</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>2</td>
<td>1</td>
<td>None</td>
</tr>
<tr>
<td>Trial 2</td>
<td>3</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>Trial 3</td>
<td>5</td>
<td>3</td>
<td>None</td>
</tr>
<tr>
<td>Trial 4</td>
<td>1</td>
<td>3</td>
<td>None</td>
</tr>
<tr>
<td>Trial 5</td>
<td>4</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>Trial 6</td>
<td>5</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>Trial 7</td>
<td>1</td>
<td>5</td>
<td>None</td>
</tr>
<tr>
<td>Trial 8</td>
<td>1</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>Trial 9</td>
<td>2</td>
<td>5</td>
<td>None</td>
</tr>
<tr>
<td>Trial 10</td>
<td>3</td>
<td>4</td>
<td>None</td>
</tr>
</tbody>
</table>

Rank items from the highest total to the lowest total to obtain a preference hierarchy, where the higher number indicates a higher preference.

1. __________________
2. __________________
3. __________________
4. __________________
5. __________________

Note: Data sheet for the paired-stimulus preference assessments that include five items to be assessed.
<table>
<thead>
<tr>
<th>Items</th>
<th>Number of Times Chosen</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Smartphone</td>
<td></td>
</tr>
<tr>
<td>2. Connect 4</td>
<td></td>
</tr>
<tr>
<td>3. Drawing</td>
<td></td>
</tr>
<tr>
<td>4. Book</td>
<td></td>
</tr>
<tr>
<td>5. Music (Headphones)</td>
<td></td>
</tr>
</tbody>
</table>

**Items (circle selection)**

<table>
<thead>
<tr>
<th></th>
<th>Left</th>
<th>Right</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Trial 1</td>
<td>2</td>
<td>1</td>
<td>None</td>
</tr>
<tr>
<td>Trial 2</td>
<td>3</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>Trial 3</td>
<td>5</td>
<td>3</td>
<td>None</td>
</tr>
<tr>
<td>Trial 4</td>
<td>1</td>
<td>3</td>
<td>None</td>
</tr>
<tr>
<td>Trial 5</td>
<td>4</td>
<td>2</td>
<td>None</td>
</tr>
<tr>
<td>Trial 6</td>
<td>5</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>Trial 7</td>
<td>1</td>
<td>5</td>
<td>None</td>
</tr>
<tr>
<td>Trial 8</td>
<td>1</td>
<td>4</td>
<td>None</td>
</tr>
<tr>
<td>Trial 9</td>
<td>2</td>
<td>5</td>
<td>None</td>
</tr>
<tr>
<td>Trial 10</td>
<td>3</td>
<td>4</td>
<td>None</td>
</tr>
</tbody>
</table>

Rank items from the highest total to the lowest total to obtain a preference hierarchy, where the higher number indicates a higher preference.

1. ___________________
2. ___________________
3. ___________________
4. ___________________
5. ___________________

*Note: Data sheet for the paired-stimulus preference assessments that include five items to be assessed.*
Appendix I: MSWO Data Sheets for Participant Use
Rank items from the lowest total to the highest total to obtain a preference hierarchy, where the lower number indicates a higher preference.

1. _______________________
2. _______________________
3. _______________________
4. _______________________
5. _______________________
6. _______________________

Note. Data sheet the multiple stimulus without replacement (MSWO) preference assessment that includes six items to be assessed.
Rank items from the lowest total to the highest total to obtain a preference hierarchy, where the lower number indicates a higher preference.

1. _______________________
2. _______________________
3. _______________________
4. _______________________
5. _______________________
6. _______________________

Note. Data sheet the multiple stimulus without replacement (MSWO) preference assessment that includes six items to be assessed.
Appendix J: Integrity Checklists for Training Preference Assessments
Baseline

1. _______ Items on table (all or nothing)
   a. Plain paper ______
   b. Writing utensil ______
   c. A pool of items to be assessed_____
      i. 5 items for PS assessments
      ii. 6 items for MSWO assessments
   d. Calculator____
   e. Timer____
2. _______ Role-played student sitting across the participant
3. _______ Provided a written case scenario
4. _______ Terminated the session after 10 trials
   Percentage = _____/4

Training via Practitioner Journal Article

1. _______ Asked the participant to self-report the time(s) spent reading the article
2. _______ Items on table (all or nothing)
   a. Data sheet for participant use _____
   b. Writing utensil ______
   c. A pool of items to be assessed_____
      i. 5 items for PS assessments
      ii. 6 items for MSWO assessments
   d. Calculator____
   e. Timer____
3. _______ Role-played student sitting across the participant
4. _______ Provided a written case scenario
5. _______ Terminated the session after 10 trials
   Percentage = _____/5
Appendix K: Social Validity Questionnaire for Participants—Preference Assessments
Social Validity Questionnaire (Anonymous)

Preference Assessment

On a scale of 1–5 with 1 = strongly disagree and 5 = strongly agree, please rate the following statements (circle):

1. The assessment focuses on an important student behavior.

   1   2   3   4   5

2. I believe that this assessment will produce useful information for planning intervention or instructional programs.

   1   2   3   4   5

3. The assessment can be easily incorporated into my classroom system.

   1   2   3   4   5

4. The procedural descriptions in the articles are clear and explicit, and I understand the assessment steps.

   1   2   3   4   5

5. After reading the procedural descriptions from the published journal article, I believe I can accurately implement the assessment with my students.

   1   2   3   4   5

6. I have the necessary materials to implement this assessment accurately.

   1   2   3   4   5

7. The time requirements of reading the journal article are reasonable.

   1   2   3   4   5

8. If I were going to provide training materials to my colleagues who had no previous exposure to this topic, I would recommend using this article.

   1   2   3   4   5

9. If I were going to provide training materials to parents who had no previous exposure to this topic, I would recommend using this article.

   1   2   3   4   5

Thank you very much!
Appendix L: Social Validity Questionnaire for the Director—Preference Assessments
Social Validity Questionnaire

Preference Assessments

On a scale of 1–5 with 1 = strongly disagree and 5 = strongly agree, please rate the following statements (circle):

1. Conducting systematic preference assessments is an acceptable way to identify things that will motive students with special needs.
   
   1 2 3 4 5

2. Other staff members in my school should be trained to conduct preference assessments.
   
   1 2 3 4 5

3. After participating this study, the staff learned how to conduct preference assessments.
   
   1 2 3 4 5

4. Training via the practitioner journal article was practical and feasible.
   
   1 2 3 4 5

5. The published practitioner article served as an effective method for training staff members on the target procedure (preference assessments), in the absence of an expert trainer.
   
   1 2 3 4 5

6. Based on the outcomes, I would use the practitioner journal article if I were going to provide training to staff members who had no previous exposure to the target procedure.
   
   1 2 3 4 5

7. Based on the outcomes, I would recommend other colleagues to use the practitioner journal article if they were going to provide training to staff members who had no previous exposure to the target procedure.
   
   1 2 3 4 5

8. Practitioner journal articles would be helpful in training staff to implement other assessment or instructional procedures.
   
   1 2 3 4 5
Please answer the following open-ended questions:

1. What do you like about using practitioner journal articles as a method of training?

2. What do you dislike about using practitioner journal articles as a method of training?

Additional comments:

Thank you very much!
Appendix M: The Case Scenario for System of Least Prompts
Alpha is a 4-year-old boy with moderate intellectual disability. He loves adult attention and high-fives. He can follow some one-step instructions and sometimes needs prompts to complete tasks/directions. When working with Alpha, full physical prompts usually result in correct response if he is not able to respond correctly to other prompts.

This semester a new skill has been added to Alpha’s IEP to promote independent living—\textbf{fastening a zipper on a jacket.}

The target skill has been broken down into small, teachable steps (see below for the task list of “fastening a zipper on a jacket”). Please use the system of least prompts (aka. least-to-most prompting) to teach the skill to Alpha on TWO occasions, collect data on his responses, and use a weighted point system to provide a graphic display of his progress.

The role-played Alpha will let you know when to start and end.

1. Bring the zipper to the bottom of jacket
2. Latch the two parts of the zipper together
3. Grasp the bottom of jacket near the zipper
4. Grasp the zipper
5. Pull the zipper all the way up
Appendix N: An Example Script for System of Least Prompts Role-Play Sessions
SLP—Alpha

♥ Adult attention and high-gives ♥

**Full physical prompts**

“Fastening a zipper on a jacket” prep:

- Task analysis and written expectations
- Pens and Paper
- Timer (cell phone)
- Calculator
- The role-played student is wearing the Jacket (zipper on top!!)

**Script 1**

<p>| | | | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Bring the zipper to the bottom of jacket</td>
<td>1</td>
<td>Most prompt</td>
<td>6</td>
</tr>
<tr>
<td>2. Latch the two parts of the zipper together</td>
<td>2</td>
<td>Most prompt</td>
<td>7</td>
</tr>
<tr>
<td>3. Grasp the bottom of jacket near the zipper</td>
<td>3</td>
<td>Least prompt</td>
<td>8</td>
</tr>
<tr>
<td>4. Grasp the zipper</td>
<td>4</td>
<td>Least prompt</td>
<td>9</td>
</tr>
<tr>
<td>5. Pull the zipper all the way up</td>
<td>5</td>
<td>Intermediate prompt</td>
<td>10</td>
</tr>
</tbody>
</table>
Appendix O: An Example Data Sheet for Evaluating System of Least Prompts

Implementation
Assessment Criteria for SLP (Systems of Least Prompting) — *SCRIPT 1 Alpha*

Name of observer: 
Participant: 
Video tape label: 
Phase and Session: 

1 = correct 0 = incorrect

If one of the target behaviors is incorrect, the entire trial should be counted as incorrect. Each trial is a step in the task analysis. The ending step in the behavior chain requires the participant to calculate and plot the total weighted points. In some trials, some prompt levels may be N/A (based on the simulated client’s scripts).

<table>
<thead>
<tr>
<th>Trial</th>
<th>Vocal Direction</th>
<th>Least Prompt</th>
<th>Intermediate Prompt</th>
<th>Most Prompt</th>
<th>Reinforcer Delivery</th>
<th>Data Collection</th>
<th>Graphing</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bring the zipper to the bottom of jacket</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Latch the two parts of the zipper together</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Grasp the bottom of jacket near the zipper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Grasp the zipper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Pull the zipper all the way up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6/1</td>
<td>Bring the zipper to the bottom of jacket</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7/2</td>
<td>Latch the two parts of the zipper together</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8/3</td>
<td>Grasp the bottom of jacket near the zipper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9/4</td>
<td>Grasp the zipper</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10/5</td>
<td>Pull the zipper all the way up</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Appendix P: A Cover Page with Reading Log—System of Least Prompts
Hello ______________________ (please fill in your name here)!

Please take one week to closely read the attached article “Teaching with the System of Least Prompts: An Easy Method for Monitoring Progress.”

One week later, you will be asked to use least-to-most prompting (the system of least prompts) to teach a skill to a role-played student and collect data to track his/her responses/progress. A data sheet that is similar to Figure 3 (Page 50) will be provided to you at the beginning of the role play sessions.

Please use the table below to track the dates and durations you read this article.

<table>
<thead>
<tr>
<th>Date</th>
<th>Length of time spent in reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/29/2016</td>
<td>45 min</td>
</tr>
</tbody>
</table>
Appendix Q: System of Least Prompts Data Sheet for Participant Use
Determine the prompt hierarchy (from least to most intrusive) and point values:

- **Independence**: ____ points
- ______________: ____ points
- ______________: ____ points
- ______________: ____ points

<table>
<thead>
<tr>
<th>Step</th>
<th>Task Analysis</th>
<th>Student Responding</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Bring the zipper to the bottom of jacket</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>Latch the two parts of the zipper together</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>Grasp the bottom of jacket near the zipper</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>Grasp the zipper</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>Pull the zipper all the way up</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Session</th>
<th>Total session points</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Hello ______________________ (please fill in your name here)!

Please take one week to closely read the attached article “Measuring Fidelity to Improve Intervention Effectiveness.”

During this week, please use the information in the least-to-most prompting (the system of least prompts) article to develop a fidelity checklist on how to accurately use least-to-most prompting and track student progress/responses. To find examples of fidelity checklist, you may refer to Figure 1 (page 17) or Figure 2 (page 18) of the attached article. You will be asked to self-assess the fidelity during next role-play sessions.

Please use the table below to track the dates and durations you read this article AND develop the fidelity checklist.

<table>
<thead>
<tr>
<th>Date</th>
<th>Length of time spent in reading</th>
</tr>
</thead>
<tbody>
<tr>
<td>2/29/2016</td>
<td>45 min</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Date</th>
<th>Length of time spent in developing fidelity checklist</th>
</tr>
</thead>
</table>
Appendix S: Integrity Checklists for Training System of Least Prompts
Baseline

1. _____ Items on table (all or nothing)
   a. Plain paper ______
   b. Writing utensil ______
   c. Material(s) necessary for the chained skill_____
      i. A jacket with a zipper
   d. Calculator_____
   e. Timer ______
2. _____ Role-played student sitting/standing across the participant
3. _____ Provided a written case scenario
4. _____ Asked the participant to state the prompt sequence to be used
5. _____ Terminated the session after 10 trials
   Percentage = _____/5

Training via Practitioner Journal Article

1. _____ Asked the participant to self-report the time(s) spent reading the article
2. _____ Items on table (all or nothing)
   a. Data sheet for participant use ______
   b. Writing utensil ______
   c. Material(s) necessary for the chained skill_____
      i. A jacket with a zipper
   d. Calculator_____
   e. Timer ______
3. _____ Role-played student sitting/standing across the participant
4. _____ Provided a written case scenario
5. _____ Asked the participant to state the prompt sequence to be used
6. _____ Terminated the session after 10 trials
   Percentage = _____/6
Training via Practitioner Journal Article (Self-Monitoring)

1. ______ Asked the participant to self-report the time(s) spent reading the articles and developing the procedural checklist
2. ______ Items on table (all or nothing)
   a. Data sheet for participant use ______
   b. Procedural checklist for self-monitoring ______
   c. Writing utensil ______
   d. Material(s) necessary for the chained skill_____
      i. A jacket with a zipper
   e. Calculator_____
   f. Timer _____
3. ______ Role-played student sitting/standing across the participant
4. ______ Provided a written case scenario
5. ______ Asked the participant to state the prompt sequence to be used
6. ______ Asked the participant to self-monitor his/her implementation of the procedures
7. ______ Terminated the session after 10 trials

Percentage = _____/7
Appendix T: Social Validity Questionnaire for Participants—System of Least Prompts
Social Validity Questionnaire (Anonymous)

The System of Least Prompt: Prior to Training

On a scale of 1–5 with 1= strongly disagree and 5 = strongly agree, please rate the following statements (circle):

1. The intervention focuses on an important student behavior.
   
   1 2 3 4 5

2. I believe that this intervention will improve the student’s behavioral outcomes.

   1 2 3 4 5

3. The intervention can be easily incorporated into my classroom system.

   1 2 3 4 5

4. I believe I can accurately implement the intervention with my students.

   1 2 3 4 5

5. I have the necessary materials to implement this intervention accurately.

   1 2 3 4 5

Thank you very much!
Social Validity Questionnaire (Anonymous)

The System of Least Prompt: After Reading the Article

On a scale of 1–5 with 1 = strongly disagree and 5 = strongly agree, please rate the following statements (circle):

1. The intervention focuses on an important student behavior.
   1   2   3   4   5

2. I believe that this intervention will improve the student’s behavioral outcomes.
   1   2   3   4   5

3. The intervention can be easily incorporated into my classroom system.
   1   2   3   4   5

4. The procedural descriptions in the articles are clear and explicit, and I understand the intervention steps.
   1   2   3   4   5

5. After reading the procedural descriptions from the published journal article, I believe I can accurately implement the intervention with my students.
   1   2   3   4   5

6. I have the necessary materials to implement this intervention accurately.
   1   2   3   4   5

7. The time requirements of reading the journal article are reasonable.
   1   2   3   4   5

8. If I were going to provide training materials to my colleagues who had no previous exposure to this topic, I would recommend using this article.
   1   2   3   4   5

9. If I were going to provide training materials to parents who had no previous exposure to this topic, I would recommend using this article.
   1   2   3   4   5

Thank you very much!
Social Validity Questionnaire

The System of Least Prompt: *After Self-Monitoring*

Please answer the following open-ended questions:

1. What do you think about the article about self-monitoring? (e.g., theme/content, writing style, and supporting materials)

2. What do you like about using self-monitoring to support the implementation of the prompting procedures?

3. What do you dislike about using self-monitoring to support the implementation of the prompting procedures?

4. To what extent would you consider using self-monitoring to support the implementation of other assessment or intervention procedures?

**Additional comments:**
Appendix U: Social Validity Questionnaire for the Director—System of Least Prompts
Social Validity Questionnaire

The System of Least Prompts

On a scale of 1–5 with 1 = strongly disagree and 5 = strongly agree, please rate the following statements (circle):

1. Using the system of least prompts is an acceptable way to promote skill acquisition and reduce prompt dependency for students with special needs.

   1  2  3  4  5

2. Other staff members in my school should be trained to implement the system of least prompts.

   1  2  3  4  5

3. After participating this study, the staff learned how to implement the system of least prompts.

   1  2  3  4  5

4. Training via the practitioner journal articles was practical and feasible.

   1  2  3  4  5

5. These published practitioner articles (pertaining to the system of least prompts and self-monitoring) served as an effective method for training staff members on the target procedure (the system of least prompt), in the absence of an expert trainer.

   1  2  3  4  5

6. Based on the outcomes, I would use the practitioner journal articles if I were going to provide training to staff members who had no previous exposure to the target procedure.

   1  2  3  4  5

7. Based on the outcomes, I would recommend other colleagues to use the practitioner journal articles if they were going to provide training to staff members who had no previous exposure to the target procedure.

   1  2  3  4  5

8. Practitioner journal articles would be helpful in training staff to implement other assessment or instructional procedures.

   1  2  3  4  5
Please answer the following open-ended questions:

9. What do you like about using practitioner journal articles as a method of training?

10. What do you dislike about using practitioner journal articles as a method of training?

Additional comments:

Thank you very much!