Education and Training in Autism and Developmental Disabilities

Focusing on individuals with autism, intellectual disability and other developmental disabilities

Volume 55 Number 3 September 2020
Education and Training in Autism and Developmental Disabilities

Editor: Stanley H. Zucker
Arizona State University
Mary Lou Fulton Teachers College

Consulting Editors

Martin Agran
Kevin Ayres
Devender Banda
Juliet Hart Barnett
Laura Basette
Kyle D. Bennett
Emily Bouck
Amanda Boutot
Michael P. Brady
Stacy Carter

Consulting Editors

Mary Lou Fulton Teachers College
Arizona State University

HELENA MEADAN-KAPLANSKY
Pam Mims
Kristi Morin
Wendy Oakes
Jenny Root
Karrie Shogren
Tom E.C. Smith
Scott Sparks
Fred Spooner
Robert Stodden
Keith Storey
William Therrien
Matt Tincani
Jason Travers
Toni Van Laarhoven
Elizabeth West
John Wheeler
Mark Wolery
Leah Wood
Dahun Zhang

Education and Training in Autism and Developmental Disabilities is sent to all members of the Division on Autism and Developmental Disabilities of The Council for Exceptional Children. All Division members must first be members of The Council for Exceptional Children. Division membership dues are $30.00 for regular members and $15.00 for full-time students. Memberships are on a yearly basis. All inquiries concerning membership, subscription, advertising, etc. should be sent to the Division on Autism and Developmental Disabilities, 3100 Clarendon Blvd., Suite 600, Arlington, Virginia 22201-5332. Advertising rates are available upon request.

Manuscripts should be typed, double spaced, and emailed (etadd@asu.edu) to the Editor: Stanley H. Zucker, Mary Lou Fulton Teachers College, Box 871811, Arizona State University, Tempe, AZ 85287-1811. Each manuscript should have a transmittal letter and a cover sheet that gives the names, affiliations, and complete addresses of all authors.

Editing policies are based on the Publication Manual, the American Psychological Association, 2009 revision. Additional information is provided on the inside back cover. Any signed article is the personal expression of the author; likewise, any advertisement is the responsibility of the advertiser. Neither necessarily carries Division endorsement unless specifically set forth by adopted resolution.

Education and Training in Autism and Developmental Disabilities is abstracted and indexed in Psychological Abstracts, PsyclINFO, e-psyche, Abstracts for Social Workers, International Journal of Rehabilitation Research, Current Contents/Social and Behavioral Sciences, Excerpta Medica, ISI Social Sciences Citation Index, Adolescent Mental Health Abstracts, Educational Administration Abstracts, Educational Research Abstracts, and Language and Language Behavior Abstracts. Additionally, it is annotated and indexed by the ERIC Clearinghouse on Handicapped and Gifted Children for publication in the monthly print index Current Index to Journals in Education and the quarterly index, Exceptional Child Education Resources. Access is also available in EBSCO, ProQuest, and JSTOR.


Division on Autism and Developmental Disabilities
The Council for Exceptional Children

Board of Directors

Officers

PAST PRESIDENT Michael Wehmeyer
PRESIDENT Ginevra Courtade
Vice President Peggy J. Schaefer-Whitby
SECRETARY Angi Stone-MacDonald
TREASURER Lynd Stansberry Brusahan

Members

Megan Carpenter (Student Representative)
Elizabeth Harkins
Kimberly Match
Jenny Root

Executive Director
Emily Bousk
Publications Chair
James Thompson
Communications Chair
Bee Jimenez
Conference Coordinator
Cindy Perras

The purposes of this organization shall be to advance the education and welfare of persons with autism and developmental disabilities, research in the education of persons with autism and developmental disabilities, competency of educators in this field, public understanding of autism and developmental disabilities, and legislation needed to help accomplish these goals. The Division shall encourage and promote professional growth, research, and the dissemination and utilization of research findings.

EDUCATION AND TRAINING IN AUTISM AND DEVELOPMENTAL DISABILITIES (ISSN 2154-1647) (USPS 0016-8500) is published quarterly, by The Council for Exceptional Children, Division on Autism and Developmental Disabilities, 3100 Clarendon Blvd., Suite 600, Arlington, Virginia 22201-5332. Members’ dues to The Council for Exceptional Children Division on Developmental Disabilities include $8.00 for subscription to EDUCATION AND TRAINING IN AUTISM AND DEVELOPMENTAL DISABILITIES. Subscription to EDUCATION AND TRAINING IN AUTISM AND DEVELOPMENTAL DISABILITIES is available without membership. Individual—U.S. $100.00 per year; Canada, PUAS, and all other countries $104.00; Institutions—U.S. $249.00 per year; Canada, PUAS, and all other countries $254.00; single copy price is $40.00. U.S. Periodicals postage is paid at Arlington, Virginia 22204 and additional mailing offices.

POSTMASTERS: Send address changes to EDUCATION AND TRAINING IN AUTISM AND DEVELOPMENTAL DISABILITIES, 3100 Clarendon Blvd., Suite 600, Arlington, Virginia 22201-5332.
Secondary Transition of Students with Autism Spectrum Disorder: Recommendations for Researchers 247
DAVID W. TEST, ERIK CARTER, JENNIFER COYLE, RACHEL SEAMAN-TULLIS, DANA RUSHER, and SAMUEL L. ODOM

What is the Proof Now? An Updated Methodological Review of Research on Social Stories 264
CHRISTINE M. MILNE, JUSTIN B. LEAF, JOSEPH H. CIHON, JULIA L. FERGUSON, JOHN MCEACHIN, and RONALD LEAF

Enhancing Early Numeracy Skills of Children with Severe Disabilities and Complex Communication Needs 277
JOHN C. WRIGHT, CHRISTOPHER J. LEMONS, VICTORIA F. KNIGHT, ESTHER LINDSTRÖM, and JULIA STRAUSS

Evidence for a Literacy Intervention for Students with Intellectual and Developmental Disabilities 290
JILL H. ALLOR, PAUL YOVANOFF, STEPHANIE AL. OTAIBA, MIRIAM B. ORTIZ, and CARLIN CONNER

Paraprofessional-Implemented Systematic Instruction for Students with Disabilities: A Systematic Literature Review 303
VIRGINIA L. WALKER, KAREN H. DOUGLAS, SARAH N. DOUGLAS, and SOPHIA R. D’AGOSTINO

Effects of General Education Teacher-Delivered Embedded Instruction to Teach Students with Intellectual Disability to Solve Word Problems 318
JESSICA BOWMAN, JOHN MCDONNELL, JOANNA RYAN, OLIVIA FUDGE COLEMAN, LYNDSEY AIONO CONRADI, and CARRIE EICHELBERGER

Dissemination and Decision-Making: Factors Related to Pre-Service Practitioners’ Selection of Practices for Students with Autism 332
MARIA L. HUGH, LEANNE D. JOHNSON, and VERONICA P. FLEURY

Increasing the Provision of Choices within an Adult Transition Program 348
RACHEL J. SCHWARTZ, RACHEL E. ROBERTSON, and SARAH WESTERFIELD

Manuscripts Accepted for Future Publication in Education and Training in Developmental Disabilities 246

The Division on Autism and Developmental Disabilities retains literary property rights on copyrighted articles. Up to 100 copies of the articles in this journal may be reproduced for nonprofit distribution without permission from the publisher. All other forms of reproduction require permission from the publisher.
Manuscripts Accepted for Future Publication in Education and Training in Autism and Developmental Disabilities

December 2020

Evaluating the use of video-and computer-based technology supports to facilitate small group instruction. Erinn E. Whiteside, Kevin M. Ayres, Jennifer R. Ledford, and Cary Trump, 3600 Vienna Drive, College Station, TX 77845.

Multi-component literacy intervention with science expository text for students with moderate intellectual disability. Carly A. Roberts, Jacob Tandy, So Yeon Kim, and Nancy Meyer, College of Education, University of Washington, 2012 Skagit Lane, Miller Hall, Box 353600, Seattle, WA 98195-3600.


Direct instruction of text-to-speech software for students with intellectual disability. Lauren Bruno, Amanda M. Lewis, Erica R. Kaldenberg, Patti Bahr, and Jordan Immerfall, University of Iowa, REACH Program, 240 S. Madison St., S229 LC, Iowa City, IA 52242.

Increasing opportunities for persons with IDD to live in their own homes. David Westling and Kelly R. Kelley, 15 Walnut Lane, Fletcher, NC 28723.


Effects of three combined reading instruction devices on the reading achievement of adolescents with mild intellectual disability. Celine Chatenoud, Catherine Turcotte, and Rebeca Aldama, Department d’education et formation specialisees, Universite du Quebec a Montreal, Case postale 8888, succursale Centre-Ville, Montreal, Quebec H3C 3P8 CANADA.
Abstract: Given the continuing poor post-school outcomes for students with ASD, combined with the limited number of current interventions to address the problem, the U.S. Department of Education, Office of Special Education and Rehabilitation Services’ (OSERS) charged the National Technical Assistance Center on Transition (NTACT) with identifying and disseminating evidence- and research-based practices and predictors for transition-age students with ASD. During this process of identification, we identified several areas that directly impact the research and its efficacy when looking at what works for students with ASD. The purpose of this article is to provide recommendations to researchers in the areas of: a) identifying effective interventions, b) tailoring available interventions, c) addressing mixed samples, d) improving measurement, e) equipping professionals, and f) developing and evaluating programs.

Approximately 1 in 59 children are identified as having autism spectrum disorder (ASD) according to the Center for Disease Control and Prevention’s Autism and Developmental Disabilities Monitoring Network (Baio et al., 2018). Between the 2005–2006 and 2014–2015 school years, the number of children nationwide identified as having ASD rose 165% (Samuels, 2016). In 2016, 578,099 children with ASD ages 6–21 received special education services under the Individuals with Disabilities Education Act (IDEA), Part B (IDEA Data Center, 2018).

Although the number of students with ASD has increased, their post-school outcomes have not. A series of recent reports illustrates the challenges these young people face. For example, the 2015 National Autism Indicators Report: Transition into Young Adulthood (Roux et al., 2015) examined 12 services youth with ASD commonly use (i.e., speech-language therapy, occupational/life skills therapy, physical therapy, vocational services, diagnostic medical services, psychological/mental health/social work, transportation, assistive technology/devices, personal assistant, in-home/classroom aide, respite care, case management). Results indicated: (a) nearly all youth (97%) received at least one of these 12 services while in high school (b) service usage dropped after high school; (c) 37% of youth in their early 20s were defined as “disconnected” (i.e., neither worked outside the home nor continued education after high school); and (c) 28% of youth were neither employed nor attending postsecondary school or training and were receiving no services or supports. The 2016 National Autism Indicators Report: Vocational Rehabilitation (Roux et al., 2016) also depicts a disappointing portrait. While approximately 60% of people with ASD who used vocational rehabilitation (VR) services left with a job (i.e., about the same rate as those with intellectual disability or...
other types of disabilities), about 80% of individuals with ASD worked part-time at a median weekly rate of $160. These earnings placed most workers with ASD below the federal poverty line.

Likewise, longitudinal studies depict the varied outcomes experienced among young people with autism. For example, the National Longitudinal Transition Study-2 (Newman et al., 2011) provided information about youth with disabilities over a 10-year period using a nationally representative sample of secondary students with disabilities. Up to eight years after leaving high school (a) only 43.9% of individuals with ASD were enrolled in a post-secondary education program (compared to 67% of all individuals with disabilities), (b) only 63.2% of individuals with ASD had been employed at some point since high school (compared to about 91% of all individuals with disabilities), and (c) only 17% of individuals with ASD lived independently (compared to 45% of all individuals with disabilities).

Most recently, the National Longitudinal Transition Survey 2012 (Lipscomb et al., 2017) reported the high school experiences of transition-age youth from approximately 13,000 students, ages 13 to 21, from more than 400 school districts. In a third volume of the NLTS 2012 reports, Liu et al. (2018) examined trends for in-school youth ages 15 to 18 from 2003 to 2012, using the NLTS2 and NLTS 2012. From 2003 to 2012, significant increases were found in the proportion of youth (ages 15–18) with ASD (a) living in households in which no parent had a paid job (9% to 17%), (b) living in households that received Supplemental Nutrition Assistance Program benefits in the past two years (6% to 17%) and (c) who had a school-sponsored work activity in the past year (11% to 21%). At the same time, a significant decrease was found in the percentage of parents meeting with school staff to develop transition plans (78% to 65%).

Given these statistics, it is not surprising that both federal reports (U.S. Government Accountability Office [GAO], 2017; U.S. Department of Education [USDOE], 2017) and researchers (Hong et al., 2017; Wehman et al., 2014; Westbrook et al., 2015) have called for increased resources to assist students with ASD as they transition from school to adulthood. Wehman et al. (2014) noted the paucity of research-based transition packages and models available to change the pattern of poor post-school outcomes for youth with ASD. In their review of behavioral and social interventions designed to prepare students with ASD for employment upon graduation, Westbrook et al. (2015) were not able to find any experimental or quasi-experimental studies from 1943 to 2011 that provided a clear link between interventions and employment outcomes for young adults with ASD. However, based on the qualitative and descriptive research during this time, they suggested promising characteristics of school-based programs for students with ASD (e.g., modeling and shaping techniques, social skills interventions, family-centered approaches).

To date, few articles have summarized evidence-based practices and predictors specifically for secondary students with ASD. For example, a meta-analysis by Lee et al. (2007) identified self-management as an effective intervention for increasing appropriate behavior for students with ASD. However, their results were disaggregated only for preschool (3-5) and school age (6-17) students. A 2014 special issue of *Remedial and Special Education* focused on Autism, Adolescence, and High School (Hume, 2014) included a series of papers offering intervention recommendations related to families, peer and social competence, academic performance, and independence and self-management. However, these recommendations were based on expert knowledge rather than systematic literature reviews. More recently, a review by Hong et al. (2017) examined functional living skills interventions for adolescents and adults with ASD. Although their review identified four interventions with strong effects (i.e., visual cues, audio cueing, video modeling, and behavioral in-vivo), the authors emphasized the need for future research to identify interventions to improve the functional living skills of these students. Given the continuing poor post-school outcomes for students with ASD, combined with the limited number of current interventions to address the problem, the U.S. Department of Education, Office of Special Education and Rehabilitation Services’ (OSERS) charged the National Technical Assistance Center on...
Transition (NTACT) with identifying and disseminating evidence- and research-based practices and predictors for transition-age students with ASD. During this process of identification, we identified several areas that directly impact the research and its efficacy when looking at what works for students with ASD. The purpose of this article is to provide recommendations to researchers in the areas of: a) identifying effective interventions, b) tailoring available interventions, c) addressing mixed samples, d) improving measurement, e) equipping professionals, and f) developing and evaluating programs.

Summary of Existing Knowledge

Current Practices

The National Standards Project (NSP; National Autism Center, 2009) and the National Professional Development Center on Autism Spectrum Disorder (NPDC; Odom et al., 2010), conducted comprehensive reviews of the ASD intervention literature in the mid-2000s and updated their findings five and six years later (National Autism Center, 2015; Wong et al., 2014). In contrast, NTACT examined its database of practices for all transition-aged students with disabilities (https://transitionta.org/effectivepractices) to extract practices that included students with ASD. In this next section, we provide a description of findings related to transition-age students with ASD from these three projects.

The National Standards Project. In 2015, the National Autism Center at the May Institute published Phase 2 of their NSP (National Autism Center, 2015). This comprehensive review addressed the need for evidence-based practice guidelines for individuals with ASD, and answered the question of how to effectively treat this population. For individuals with ASD under 22 years of age, the NSP identified 14 established interventions, although only nine of those interventions included transition-age participants. The practice categories included behavioral interventions, cognitive behavioral intervention, modeling, parent training package, peer training package, scripting, self-management, social-skills package, and story-based interventions. In terms of evidence strength for interventions for adults with ASD, there was only one established intervention (i.e., behavioral interventions), and one emerging intervention (i.e., vocational training package).

National Professional Development Center on ASD. In 2014, NPDC identified 27 focused interventions as evidence-based practices. The center’s goal is to promote the use of evidence-based practices (EBPs) for children and youth with ASD, birth to 22 years. Although NPDC identified 27 EBPs, only 14 were related to education/training (i.e., academic outcomes), seven were related to employment (i.e., vocational outcomes), and 16 were related to independent living (i.e., adaptive outcomes). Table 1 includes the interventions, outcomes and age of proposed student use.

National Technical Assistance Center on Transition. Through their systematic review, NTACT identified 47 practices for practitioners who work with secondary students with ASD to incorporate into their teaching and practice (see Table 2). Practices were categorized using the Taxonomy for Transition Programming 2.0 (Kohler et al., 2016) as an organizational framework. Thirty-eight practices were identified in the area of Student Development, four in Student-Focused Planning, and five in Interagency Collaboration. No practices were found in the areas of Family Engagement and Program Structures that included secondary students with ASD. Each practice was also tagged with the relevant post-secondary area (i.e., education, employment, independent living) to assist practitioners in transition planning for students with ASD.

Current Predictors

In 2009, Test and colleagues identified variables or predictors correlated with improved post-school outcomes for students with disabilities using the quality indicators suggested by Thompson et al. (2005). Predictors are macro level variables that influence systems, programs, and general practices and skills students need to be successful after high school. In 2016, using the same research procedures and inclusion criteria as Test et al. (2009), four additional predictors emerged (Mazzotti et al., 2016), bringing the total to 20. To identify predictors of post-school success in the areas of education, employment, and inde-
pendent living for students with ASD, NTACT analyzed its database of all predictor studies to determine if students with ASD were included in the sample. Table 3 provides the predictor, its definition, and the postsecondary outcome(s) for students with ASD, if they were a part of the study. While all the studies in Table 3 included at least one student with ASD, only two studies (Chiang et al., 2012; Chiang et al., 2013) included only students with ASD.

**Recommendations for Research**

As emphasized throughout this paper, high-quality research can play an important role in identifying how practices, programs, partner-

---

**TABLE 3**

NPDC on ASD: Findings Relevant to Outcomes for Transition-age Individuals with ASD

<table>
<thead>
<tr>
<th>Evidence-based Practices</th>
<th>Outcome</th>
<th>Age</th>
</tr>
</thead>
<tbody>
<tr>
<td>Antecedent-based intervention</td>
<td>Academic</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Cognitive-behavioral intervention</td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Differential reinforcement of alternative, incompatible, or other behavior</td>
<td>Adaptive</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td>Discrete trial training</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Extinction</td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Functional behavior assessment</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Functional communication training</td>
<td>Adaptive</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td>Modeling</td>
<td>Vocational</td>
<td>15–22</td>
</tr>
<tr>
<td>Parent-implemented intervention</td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Peer-mediated instruction and intervention</td>
<td>Academic</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td>Prompting</td>
<td>Academic</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td></td>
<td>Vocational</td>
<td>15–22</td>
</tr>
<tr>
<td>Reinforcement</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td></td>
<td>Vocational</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td>Response interruption/redirection</td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Scripting</td>
<td>Vocational</td>
<td>6–14</td>
</tr>
<tr>
<td>Self-management</td>
<td>Vocational</td>
<td>15–22</td>
</tr>
<tr>
<td>Social narratives</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Structured play group</td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Task analysis</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
<tr>
<td>Technology-aided instruction and intervention</td>
<td>Academic</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td></td>
<td>Academic</td>
<td>15–22</td>
</tr>
<tr>
<td></td>
<td>Vocational</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td>Time delay</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td>Video modeling</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td></td>
<td>Vocational</td>
<td>6–14, 15–22</td>
</tr>
<tr>
<td>Visual Supports</td>
<td>Academic</td>
<td>6–14</td>
</tr>
<tr>
<td></td>
<td>Adaptive</td>
<td>6–14</td>
</tr>
</tbody>
</table>

Note: Academic = education/training outcomes; Adaptive = independent living outcomes; Vocational = employment outcomes.
## TABLE 2

**NTACT Practices for Secondary Students with ASD**

<table>
<thead>
<tr>
<th>Student Development Practice</th>
<th>Related Post-School Outcome Area(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Anchored Instruction to Teach Math</td>
<td>Education</td>
</tr>
<tr>
<td>Collaborative Strategic Reading-High School to teach reading comprehension</td>
<td>Education</td>
</tr>
<tr>
<td>Computer-based video instruction to teach students to read grocery aisle sign words and locate items in actual grocery store in response to a photograph shopping list and a typed word shopping list</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Differential reinforcement to teach response latency and task completion</td>
<td>Education, Independent Living</td>
</tr>
<tr>
<td>Envision IT Curriculum to teach information technology skills</td>
<td>Education</td>
</tr>
<tr>
<td>Graphic organizers to teach reading comprehension</td>
<td>Education</td>
</tr>
<tr>
<td>Graphic organizers to teach science</td>
<td>Education</td>
</tr>
<tr>
<td>Individual work system to teach on-task behavior and work completion skills</td>
<td>Employment</td>
</tr>
<tr>
<td>Mnemonics to teach social studies vocabulary</td>
<td>Education</td>
</tr>
<tr>
<td>Mobile video modeling to teach interview skills</td>
<td>Employment</td>
</tr>
<tr>
<td>Multi-component (i.e., choice embedding, functional communication training, building tolerance for delay of reinforcement, and presentation of discriminative stimuli for non-problem behaviors) to teach to complete shopping trip in community</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Multi-component PMI (i.e., peer training and participant instruction on the use of text cues) to teach assertive conversational skills (initiating, asking follow-up questions, and commenting)</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Multi-media, computer-based program using video captions and still photographs to teach students to read aisle signs and locate items in a grocery store</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Peer assisted instruction/support to teach social interactions</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Peer directed novel question training to teach conversational skills</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Peer mediated instruction to teach conversational skills during lunch</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Peer network interventions to teach peer interactions, social engagement</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Peer networks to facilitate increased social interactions</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Project Search to teach employment status, hours worked, benefits, adaptive behavior</td>
<td>Employment</td>
</tr>
<tr>
<td>Self-determined learning model of instruction to teach goal attainment</td>
<td>Education</td>
</tr>
<tr>
<td>Self-Regulated Strategy Development + POW-Tree to teach self-advocacy through persuasive writing</td>
<td>Education</td>
</tr>
<tr>
<td>Simulated instruction to teach basic finance (purchasing with debit card, tracking expenses, deposits, financial decision making)</td>
<td>Independent Living</td>
</tr>
</tbody>
</table>

(Continued on next page)
ships, and policies shape the transition experiences and outcomes of students with ASD. Although research in the field of transition has certainly grown - both in quantity and quality - over the past thirty years (Carter et al., 2013; Cushing et al., 2019), a substantive focus on ASD has emerged only recently (Test et al., 2014). Much more must be learned about the services and supports that will best prepare these young people for adulthood. In this sec-

<table>
<thead>
<tr>
<th>Student Development Practice</th>
<th>Related Post-School Outcome Area(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Social skills and sports program (direct instruction, modeling and process training) to teach eye contact, turn taking, relevant information used in conversation</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Take Action: Making Goals Happen curriculum to teach goal-setting and attainment</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Task analysis with prompting hierarchies to teach on-task behavior</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Technology to teach math</td>
<td>Education</td>
</tr>
<tr>
<td>Time delay to teach science skills</td>
<td>Education</td>
</tr>
<tr>
<td>Touch Math to teach mathematic computation skills</td>
<td>Education</td>
</tr>
<tr>
<td>Video modeling to teach leisure skills</td>
<td>Employment, Independent Living</td>
</tr>
<tr>
<td>Video modeling to teach vocational skills</td>
<td>Employment, Independent Living</td>
</tr>
<tr>
<td>Video modeling to teach home maintenance skills</td>
<td>Employment, Independent Living</td>
</tr>
<tr>
<td>Video modeling to teach vocational tasks</td>
<td>Employment, Independent Living</td>
</tr>
<tr>
<td>Video modeling to teach social skills</td>
<td>Employment, Independent Living</td>
</tr>
<tr>
<td>Video modeling to teach promote generalization of purchasing skills to community stores</td>
<td>Employment, Independent Living</td>
</tr>
<tr>
<td>Video modeling to teach employment skills</td>
<td>Employment, Independent Living</td>
</tr>
<tr>
<td>Video prompting via an iPad to teach independent living skills</td>
<td>Independent Living</td>
</tr>
<tr>
<td>Working at Gaining Employment Skills (WAGES)</td>
<td>Employment</td>
</tr>
<tr>
<td>Person Centered Planning to teach future expectations</td>
<td>Education</td>
</tr>
<tr>
<td>Self-Advocacy Strategy to teach student involvement in the IEP Meeting</td>
<td>Education</td>
</tr>
<tr>
<td>Self-Directed IEP to teach student involvement in the IEP meeting</td>
<td>Education</td>
</tr>
<tr>
<td>Whose Future, Is It? to teach self-determination skill</td>
<td>Education</td>
</tr>
<tr>
<td>CIRCLES to teach self-determination and IEP meeting participation</td>
<td>Education</td>
</tr>
<tr>
<td>Project SEARCH Plus ASD Supports to teach independence in the employment setting</td>
<td>Employment</td>
</tr>
<tr>
<td>Project SEARCH to teach employment status, hours worked, benefits, adaptive behavior</td>
<td>Employment</td>
</tr>
<tr>
<td>Project SEARCH to teach social, communication, and job skills</td>
<td>Employment</td>
</tr>
<tr>
<td>Video modeling (VidCoach) to teach interviewing skills</td>
<td>Employment</td>
</tr>
</tbody>
</table>

TABLE 2

(Continued)
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Definition*</th>
<th>Education</th>
<th>Employment</th>
<th>Independent living</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Career Awareness</td>
<td>Learning about opportunities, education, and skills needed in various occupational pathways to choose a career that matches one's strengths and interests.</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Benz et al., 1997 (mixed sample)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Carter et al., 2012 (mixed sample)</td>
</tr>
<tr>
<td>Career Technical Education</td>
<td>A sequence of courses that prepares students for a specific job or career at various levels from trade or craft positions to technical, business, or professional careers.</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Baer et al., 2003 (mixed sample)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chiang et al., 2013 (autism only)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Halpern, 1985 (mixed sample)</td>
</tr>
<tr>
<td>Community Experiences</td>
<td>Activities occurring outside of the school setting, supported with in-class instruction, where students apply academic, social, and/or general work behaviors and skills.</td>
<td></td>
<td></td>
<td></td>
<td>The predictor studies for community experiences did not include students with ASD.</td>
</tr>
<tr>
<td>Exit Exam Requirements/High School Diploma Status</td>
<td>Exit Exam Requirements: Standardized state tests, assessing single content areas (e.g. Algebra, English) or multiple skill areas, with specified levels of proficiency that students must pass in order to obtain a high school diploma. Diploma Status: Achieved by completing the requirements of the state awarding the diploma including the completion of necessary core curriculum credits.</td>
<td></td>
<td></td>
<td>X</td>
<td>Heal &amp; Rusch, 1994 (mixed sample)</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Definition*</th>
<th>Education</th>
<th>Employment</th>
<th>Independent living</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Goal-Setting</td>
<td>*Goal directed behavior involves actions that enable a person to reach a specified preferred outcome. Teaching goal setting and attainment skills involves teaching students to define and articulate a goal, identify current status in relation to the goal, develop an action plan, and evaluate progress toward achieving the goal (Wehmeyer &amp; Schwartz, 1998).</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Chiang et al., 2012 (autism only)</td>
</tr>
<tr>
<td>Inclusion in General Education</td>
<td>Requires students with disabilities to have access to general education curriculum and be engaged in regular education classes with peers without disabilities.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Baer et al., 2003 (mixed sample)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Chiang et al., 2012 (autism only)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Halpern, 1985 (autism only)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heal &amp; Rusch, 1994 (mixed sample)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Heal et al., 1997 (mixed sample)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Lombardi et al., 2012 (mixed sample)</td>
</tr>
<tr>
<td>Interagency Collaboration</td>
<td>A clear, purposeful, and carefully designed process that promotes cross agency, cross program, and cross disciplinary collaborative efforts leading to tangible transition outcomes for youth.</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Bullis et al., 1995 (mixed sample)</td>
</tr>
<tr>
<td>Occupational Courses</td>
<td>Individual courses that support career awareness, allow or enable students to explore various career pathways, develop occupational specific skills through instruction, and experiences focused on their desired employment goals.</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Repetto et al., 2002 (mixed sample)</td>
</tr>
<tr>
<td>Paid Employment/Work Experience</td>
<td><strong>Paid Employment:</strong> Includes existing standard jobs in a company or organization or customized work assignments negotiated with the employer, but these activities always feature competitive pay (e.g., minimum wage) paid directly to the student by the employer.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Benz et al., 1997 (mixed sample)</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Definition*</th>
<th>Education</th>
<th>Employment</th>
<th>Independent living</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Work Experience</td>
<td>Any activity that places the student in an authentic workplace, and could include: work sampling, job shadowing, internships, apprenticeships, and paid employment.</td>
<td></td>
<td>X</td>
<td>X</td>
<td>Bullis et al., 1995 (mixed sample) Carter et al., 2012 (mixed sample) Doren &amp; Benz, 1998 (mixed sample)</td>
</tr>
<tr>
<td>Parent Expectations</td>
<td>*Parent expectations typically mean having high expectations for their children (Pleet-Odle et al., 2016). Parent expectations include parents and family members planning, articulating an expectation that their child will participate in integrated and inclusive postsecondary education, and be employed in integrated and inclusive settings in the community after high school (Doren, Gau, &amp; Lindstrom, 2012).</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Carter et al., 2012 (mixed sample) Chiang et al., 2012 (autism only) Doren et al., 2012 (mixed sample)</td>
</tr>
<tr>
<td>Parental Involvement</td>
<td>Parents/families/guardians are active and knowledgeable participants in all aspects of transition planning (e.g., decision-making, providing support, attending meetings, and advocating for their child).</td>
<td>X</td>
<td>X</td>
<td></td>
<td>The predictor studies for parental involvement did not not include students with ASD.</td>
</tr>
<tr>
<td>Program of Study</td>
<td>An individualized set of courses, experiences, and curriculum designed to develop students’ academic and functional achievement to support the attainment of students’ desired post-school goals.</td>
<td>X</td>
<td></td>
<td></td>
<td>Shandra &amp; Hogan, 2008 (mixed sample)</td>
</tr>
<tr>
<td>Self-Advocacy/Self-Determination</td>
<td>The ability to make choices, solve problems, set goals, evaluate options, take initiative to reach one’s goals, and accept consequences of one’s actions.</td>
<td></td>
<td>X</td>
<td>X</td>
<td>Halpern, 1985 (mixed sample)</td>
</tr>
</tbody>
</table>

(Continued on next page)
<table>
<thead>
<tr>
<th>Predictor</th>
<th>Definition*</th>
<th>Education</th>
<th>Employment</th>
<th>Independent living</th>
<th>Reference(s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Self-Care/Independent Living</td>
<td>Skills necessary for management of one’s personal self-care and daily independent living, including the personal management skills needed to interact with others, daily living skills, financial management skills, and the self-management of healthcare/wellness needs.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Carter et al., 2012 (mixed sample) Heal &amp; Rusch, 1994 (mixed sample)</td>
</tr>
<tr>
<td>Social Skills</td>
<td>Behaviors and attitudes that facilitate communication and cooperation (e.g., social conventions, social problem-solving when engaged in a social interaction, body language, speaking, listening, responding, verbal and written communication).</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Benz et al., 1997 (mixed sample) Carter et al., 2012 (mixed sample) Chiang et al., 2013 (autism only) Halpern, 1985 (mixed sample)</td>
</tr>
<tr>
<td>Student Support</td>
<td>A network of people (e.g., family, friends, educators, and adult service providers) who provide services and resources in multiple environments to prepare students to obtain their annual transition and post-secondary goals aligned with their preferences, interests, and needs.</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>Doren &amp; Benz, 1998 (mixed sample) Halpern, 1985 (mixed sample)</td>
</tr>
<tr>
<td>Transition Program</td>
<td>Prepares students to move from secondary settings (e.g., middle school/high school) to adult-life, utilizing comprehensive transition planning and education that creates individualized opportunities, services, and supports to help students achieve their post-school goals in education/training, employment, and independent living.</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Benz et al., 1997 (mixed sample) Halpern, 1985 (mixed sample) Repetto et al., 2002 (mixed sample)</td>
</tr>
<tr>
<td>Travel Skills</td>
<td>Travel skills are defined as the ability to get to places outside home independently (Carter et al., 2012; McDonnall, 2011).</td>
<td>X</td>
<td>X</td>
<td></td>
<td>Carter et al., 2012 (mixed sample)</td>
</tr>
</tbody>
</table>

(Continued on next page)
TABLE 3  
(Continued)

<table>
<thead>
<tr>
<th>Predictor</th>
<th>Definition*</th>
<th>Education</th>
<th>Employment</th>
<th>Independent living</th>
<th>Reference(s)</th>
</tr>
</thead>
</table>
| Work Study             | A specified sequence of work skills instruction and experiences designed to develop students' work attitudes and general work behaviors by providing students with mutually supportive and integrated academic and vocational instruction.  
*Wehmeyer (1997) suggested autonomy occurs when an individual acts in relation to their own interests, preferences, and abilities without the undue influence of others and Wehmeyer and Schwartz (1998) defined decision-making as a process of selecting or coming to a conclusion about which set of potential solution is the best by teaching students to utilize problem-solving skills. | X         | X           |                    | Baer et al., 2003 (mixed sample) Carter et al., 2012 (mixed sample)          |
| Youth Autonomy/        |                                                                             |           |             |                    | Berry et al., 2012 (mixed sample) Carter et al., 2012 (mixed sample) Doren et al., 2012 (mixed sample) |
| Decision-Making        |                                                                             | X         | X           |                    |                                                                             |

Note. *Definitions are from Rowe et al. (2014) unless otherwise noted.
tion, we highlight several research implications that may help move this area of the field forward.

Identifying Effective Interventions

This article began by emphasizing the considerable heterogeneity among students with ASD. Although they share a common special education label, these students vary in myriad ways. Examples include the diverse diploma paths they are pursuing, the manifestation of co-occurring disabilities, the severity of autism symptoms, the degree of challenging behaviors, or the presence of a cognitive impairment. Moreover, each student possesses a unique combination of strengths, needs, interests, preferences, and post-school goals. Such variations make it doubtful that a given intervention will have the same impact on or relevance for every student with ASD.

One need is to evaluate existing transition interventions across a much wider range of students. For example, few academic-related interventions have been evaluated with secondary students with ASD who have cognitive impairments. Likewise, approaches for addressing the early work experiences of students with ASD who are pursuing a regular diploma have received limited empirical attention. Identifying practices that have never or rarely been evaluated with certain subgroups of students with ASD could provide a roadmap for potential research. Well-designed replication studies could help identify any boundaries associated with particular interventions. A related need, however, involves providing better descriptions of the students with ASD who participate in these studies. The clarity and consistency with which participating students with ASD are described in research reports makes it difficult to interpret and aggregate this literature. For example, researchers vary widely in the ways in which they define ASD and the relevant characteristics of students. In addition to reporting disability labels, it is important to include assessments of social and communication abilities, behavioral needs, and intellectual functioning. Addressing both of these needs puts the field in a much better position to identify which practices may work well for which students.

Tailoring Available Interventions

Individualization remains a hallmark of both special education and transition services. As researchers identify practices that have strong empirical support, it also is important to determine whether and how such practices should be adapted to meet the individualized needs of particular students. Indeed, it is unlikely that the practices described in Tables 1 to 3 (e.g., career awareness, community experiences, self-management, or social skills instruction) should be implemented in exactly the same way for any two students with ASD. This tailoring of interventions certainly requires a theoretical understanding of how interventions impact students. But, it also requires empirical guidance.

One way researchers can address this need is by using comparative designs. To date, most studies rely on demonstration designs in which a new intervention is contrasted with prevailing practices or no intervention at all (e.g., teaching self-determination is found to have more impact than not teaching self-determination; Carter et al., 2013; Gilson et al., 2017). Although this approach can establish the promise of an intervention, it says little about how an intervention might be configured for different students. Studies that adopt parametric or component analyses can provide much-needed insights into how much of an intervention is needed or which elements of a multi-component intervention are essential or unnecessary.

A second way researchers can address this need is by identifying data-driven approaches for adapting interventions. For example, a systematic review by Huber and Carter (2016) identified a variety of approaches researchers used when individualizing peer-mediated interventions to meet the needs of students with ASD and match the settings in which they were served. Future studies could delineate and evaluate the steps practitioners should take when tailoring transition practices to meet the needs of youth and young adults with ASD.

Addressing Mixed Samples

One challenge with interpreting the existing literature is the extent to which studies in-
clude participant samples comprised of both students with and without ASD. As reflected in Table 3, almost all of the available predictor studies include mixed samples in which students with ASD reflected a small proportion or students. Likewise, systematic reviews of intervention studies highlight this same challenge (Gilson et al., 2017; Lee et al., 2018). Both group-design and single-case intervention studies frequently include students with ASD alongside students without ASD, usually individuals with intellectual or other developmental disabilities. Although there are many reasons researchers might be interested in mixed samples, such sampling approaches can complicate claims about whether a given practice has sufficient support for certain subgroups of participants.

One solution is to design studies that focus narrowly on students with ASD. For example, many large-scale, extant datasets (e.g., NTLS2; NLTS 2012; National Core Indicators; National Survey of Children’s Health) are designed such that researchers could focus their analyses on the subset of students who have ASD. Within intervention studies, the inclusion criteria could be narrowed to focus specifically on individuals with ASD. A second approach involves incorporating analytic approaches that allow the impact of autism to be examined within mixed samples. This could mean examining ASD as a moderator in experimental studies or analyzing the association between ASD and effect size in meta-analyses and systematic reviews.

**Improving Measurement**

The field of transition is replete with measurement challenges. Documenting the outcomes of students with ASD during and after high school requires careful consideration of both what and when we measure (Trainor et al., in press). As reflected in the introduction to this article and throughout Table 3, it is important to understand the outcomes young people with ASD experience in the areas of education, employment, and independent living. Yet, the field’s understanding of these outcomes is often only at surface level. For example, more is known about whether young adults with ASD live after graduation than about the quality of their life in those setting. Quality of life also includes consideration of physical and psychological health, social relationships, self-determination, spirituality, sexuality, sense of community, and many others. Additional research is needed to explore how young people with ASD experience these other domains and the interventions that lead to better outcomes in each area.

The temporal aspects of measurement in transition also require further consideration. Although correlational predictor studies are helpful in connecting the in- and post-school experiences of young people with ASD, they cannot establish causal links. At this point, most experimental studies in the area of transition examine only the short-term impact of interventions (e.g., outcomes within one semester). With the identification of predictors, the opportunity now exists of conducting experimental studies of variables in high school that have long-term impacts through transition and into adulthood.

**Equipping Professionals**

Delivering the breadth of needed transition services and supports to students with ASD requires educators and other professionals to have the capacity and commitment to implement research-based practices. Although Tables 1 to 3 identify a range of important research-based practices and predictors, the extent to which each is implemented within schools remains uneven and uncertain (Holzberg et al., 2018; Knight et al., 2019). In other words, the pathways for equipping professionals to deliver best practices in transition are underdeveloped and underevaluated. Future research must address the quality and reach of training and ongoing support provided to transition professionals.

One implication is that the field should ex-
amine multiple approaches to professional development to identify how each impacts implementation and enhances the outcomes of students with ASD. The complexity and variety of interventions within the field of transition suggests that single-event workshops, in-service trainings, transition conference sessions, webinars, or other approaches that lack opportunities for practice or performance feedback may be insufficient to produce high levels of implementation fidelity. Likewise, the field needs professional development approaches that distinguish between implementation of discrete instructional practices (e.g., modeling, scripting, video modeling), comprehensive programs or multi-component interventions (e.g., community-based transition programs, inclusion in general education), and collaborative or cross-agency interventions (e.g., interagency collaboration, Pre-Employment Transition Services). A second implication is that researchers should more actively involve practitioners within their evaluation studies. In many (if not most) studies, members of the research team are involved in delivering transition interventions. This leaves unanswered the question of whether a given intervention is sufficiently feasible and understandable for practitioners to implement the approaches with adequate fidelity to benefit students.

Developing and Evaluating Programs

The research reviewed in this article addresses a wide range of individual practices and predictors associated with improved outcomes for students with ASD. This work provides an important foundation for educators and other professionals charged with serving young people with ASD. However, much less research has focused on how these interventions are combined and sequenced within comprehensive transition programs for students with disabilities (Trainor et al., 2019). In other words, schools and districts need clear guidance on how transition education should be designed and delivered to multiple students by multiple staff across multiple years. What combination of instruction and experiences should students with ASD have prior to graduation? How should it all be ordered? How much time should be devoted to each? How should these programs be resourced and staffed? The research literature includes few descriptions of comprehensive intervention packages designed to meet the needs of students with ASD (e.g., Luecking & Luecking, 2015; Odom et al., 2014). In addition, it would be helpful to determine critical program features that are associated with improved post-school outcomes for students with ASD. Although several transition program assessment approaches have been developed (Brewer, 2006; Kohler, 1996; Morningstar et al., 2016), none have been examined in relation to student outcomes.

Conclusion

For secondary transition programs for students with ASD, the research on post-school outcomes has established the need; the dramatic increase in prevalence has established the urgency; and the range of abilities for individuals with ASD has established the complexity. Yet, when one examines the life course of individuals with ASD and their families, there may be no more important phase of the educational experience than the high school years and preparation for adulthood. Fortunately, the intervention research literature, as described in this article, yields strong recommendations for practices that teachers and school practitioners can employ, and this is the first step in having an impact on the lives of young adults with ASD. Building the capacity for public school systems to support professionals in learning and utilizing such practices is the next step. The principles of implementation science that have been applied to other programs for children and youth with ASD (Odom et al., in press) and discussed in this paper (e.g., organizational fit, professional development, adaptation, individualization) may provide guidance on how to move these research-based interventions into effective daily practice in schools and communities.

References


Baer, R. M., Flexer, R. W., Beck, S., Amstutz, N.,


Received: 12 June 2019
Initial Acceptance: 21 August 2019
Final Acceptance: 7 November 2019

"This document was developed under funding by Cooperative Agreement Number H326E140004 with the U.S. Department of Education, Office of Special Education and Rehabilitative Services (OSERS). Opinions expressed herein do not necessarily reflect the position or policy of the U.S. Department of Education nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Department of Education."
What is the Proof Now? An Updated Methodological Review of Research on Social Stories

Christine M. Milne, Justin B. Leaf, and Joseph H. Cihon
Autism Partnership and Endicott College

Julia L. Ferguson, John McEachin, and Ronald Leaf
Autism Partnership

Abstract: One intervention that has been evaluated in numerous empirical investigations and has been implemented clinically is Social Stories. Social Stories are a systematic form of intervention where a text is written describing a situation (e.g., social situation) and how the learner should respond during that situation. Previous reviews and analysis have shown that the literature on Social Stories have serious methodological flaws (e.g., Leaf et al., 2015; Styles, 2011) and/or are ineffective (e.g., Leaf et al., 2015; Reynhout & Carter, 2011). One reason for the methodological flaws is the lack of functional relations shown between the Social Story and the targeted behavior; usually with poor implementation of the experimental design. The purpose of this review was to evaluate if researchers implement both single subject designs and group designs in the correct manner, which show a convincing functional relation. Additionally, we evaluated studies where Social Stories were compared to another intervention (e.g., the Teaching Interaction Procedure) and if Social Stories were added to another intervention (e.g., differential reinforcement). Clinical implications as well as future research is discussed.

Social Stories™ are an intervention developed by Gray and Garand (1993) as a strategy to teach desired social behavior(s) (e.g., Bledsoe et al., 2003; Kuoch & Mirenda, 2003) or reduce inappropriate behaviors (e.g., Lorimer et al., 2002; Scattone et al., 2002). Social Stories™ involve developing an individualized story to address a specific target behavior. Each story is composed of a combination of specific sentence types (i.e., descriptive, perspective, affirmative, cooperative, control, and directive sentences) that have a specific purpose for addressing the target behavior. Guidelines for developing a Social Story™ have changed over time resulting in some Social Stories™ involving only text (e.g., Quirmbach et al., 2009) while others have included pictures and text (e.g., Ozdemir, 2008). Social Stories™ can also be presented in several formats including a book, multimedia presentation (e.g., Hagiwara & Myles, 1999), or even sung to the individual (e.g., Brownell, 2002). Social Stories™ are a commonly implemented intervention for individuals diagnosed with autism spectrum disorder (ASD; Reichow & Volkmar, 2010) and have been implemented by behavior analysts, teachers (Kuttler et al., 1998; Scattone et al., 2002), and parents (Kuoch & Mirenda, 2003). While there are an ever increasing number of published studies that have evaluated Social Stories™, there are also reviews questioning the effectiveness and methodological rigor of Social Stories™ (e.g., Leaf et al., 2015; Styles, 2011).

Styles (2011) reviewed 51 studies using Social Stories™ from 1993 to 2011 that were published in peer reviewed journals. Styles found that there were numerous methodological limitations in the research when considering the results of these studies. In 2011, Reynhout and Carter conducted an analysis of Social Stories™ that included more studies than in previous reviews. Results indicated that there was higher quality research (e.g., quality of participant description, baseline predictability, experimental control, external validity) on Social Stories™ than previously identified; however, there was variability in the results and “only a small clinical effect on behavior” (Reynhout & Carter, 2011, p. 897).
Leaf and colleagues (2015) evaluated 41 published studies on Social Stories™ conducted between January 1993 and December 2012. The rationale behind this evaluation was to critically examine the implementation of the research design used within the studies to determine if functional control was demonstrated. In this review, Leaf and colleagues evaluated studies using AB case designs, reversal designs, and multiple baseline designs. Within each design the authors defined quality indicators (e.g., type of data collected, length and trend of baseline, immediacy of effects, overlapping data, and combination of procedures) to determine if the research design was implemented in a way that permitted a demonstration of functional control. Based upon these quality indicators, the authors categorized each study as having convincing evidence (i.e., that the implementation of Social Stories™ was clearly responsible for the behavior change), partially convincing evidence (i.e., that the implementation of Social Stories™ could be responsible for the behavior change), or no convincing evidence (i.e., that the implementation of the Social Stories™ was not responsible for the change the behavior change). Of the 41 studies evaluated, only 3 studies (i.e., 7.3%) demonstrated convincing evidence that the Social Story™ was responsible for the behavior change.

Despite the numerous reviews showing that Social Stories™ have serious methodological flaws (e.g., Leaf et al., 2015; Styles, 2011) and/or are ineffective (e.g., Leaf et al., 2015; Reynhout & Carter, 2011), professionals continue to endorse and implement Social Stories™ for individuals diagnosed with ASD. It is important for reviewers to continue to evaluate if the research on Social Stories™ shows clear functional control, i.e. if the independent variable (i.e., Social Stories™) was responsible for the change in the dependent variable. It is also important to review the results of studies that have compared Social Stories™ to other interventions. Therefore, the purpose of this study is to update and extend the review conducted by Leaf et al. (2015) by evaluating the research methodology of Social Story™ research conducted with individuals diagnosed with ASD from January 2013 to December 2018. In so doing, three questions will be evaluated. First, if the researchers were able to demonstrate functional control when using single subject designs and group designs for studies solely evaluating Social Stories™ as a standalone procedure. Second, for those studies that compared Social Stories™ to another procedure (e.g., the teaching interaction procedure), which procedure was more effective? Third, for studies evaluating the addition of Social Stories™ to other procedures (e.g., video modeling) did Social Stories™ actually have a beneficial effect?

**Method**

**Inclusion Criteria**

To be included for the analysis, articles needed to meet the following four criteria. First, articles had to be published in a peer reviewed journal (i.e., dissertations, theses, and book chapters were excluded). Second, articles had to be experimental (i.e., no reviews, commentaries, or programmatic descriptions). Third, the intervention needed to include the implementation of either Social Stories™ (i.e., stories that follow the guidelines provided by Gray) or social stories (i.e., story-based interventions that do not follow the guidelines provided by Gray, but were still classified as social stories by the authors). Fourth, articles had to include a participant diagnosed with ASD, a developmental disability, or an intellectual disability.

**Search Procedures**

We conducted a systematic review of social stories in accordance with the PRISMA guidelines (Moher et al., 2009; see Figure 1). The PRISMA guidelines consists of four broad levels: (1) identification of articles, (2) screening of articles; (3) eligibility of articles, and (4) inclusion of articles. Two reviewers (referred to as primary and secondary reviewer hence forward) were used across all levels.

First, the primary researcher conducted a search for articles published in peer-reviewed journals from January 2013 to November 2018 using the PsychINFO database. The primary researcher used the following keywords: “Social Stories,” “Social Story,” “Social Narratives,” “story based,” “story based interven-
tion,” “developmental disabilities,” and “autism” in all possible combinations. All articles identified through the initial search of PsychINFO were reviewed for the inclusion criteria. The PsychINFO search yielded 1654 articles. The two reviewers read the title and abstract of the 1654 articles and retrieved the full article, of any article (n = 27), which appeared to implement social stories. Interobserver agreement (IOA) was taken on which articles were going to be excluded and which articles were going to be further analyzed. To calculate interrater agreement the number of agreements (included or excluded) was divided by the total number of articles screened and multiplied by 100. Interrater agreement on this measure was 99.9%.

Next, two reviewers evaluated the 27 articles to identify if the studies met the four inclusion criteria. A total of 23 articles were included,

Figure 1. Flowchart of protocol for systematic review of social stories based on PRISMA guidelines (Moher et al., 2009).
while 4 were excluded (see Figure 1). IOA was taken on which articles were going to be excluded and which articles were going to be synthesized. To conduct IOA the number of agreements (included or excluded) was divided by the total number of articles screened and multiplied by 100. IOA on this measure was 100%. Thus, a total of 23 articles were included in this review for evaluation.

**Level of Evidence**

Each study was categorized based upon three levels of evidence (i.e., convincing, partially convincing, no convincing evidence) based on the methodological rigor of the experiment. Each level of evidence was defined by several variables of the experiment depending on the research design implemented (see Tables 1 to 4). For a study to be classified as one of the three levels it needed to meet all criteria for that level of evidence (e.g., convincing evidence across all variables). If a study was scored as multiple levels of evidence (e.g., convincing evidence and partially convincing), then the level of evidence was determined by the lowest score earned across the applicable variables for the research design implemented.

**Evaluations of the Designs**

**Case study designs.** There were six variables used to evaluate studies using a case study design. The six variables included (1) the type of data collected (2) the length of the baseline condition, (3) the stability and direction of the baseline trend, (4) if the treatment effect occurred immediately, (5) the number of overlapping data points between baseline and intervention, and (6) if the social story intervention was combined with any other procedure. Table 1 shows the three levels of evidence within each measure for case study designs. The type of data collected was defined as either objective (i.e., observable and measurable) or subjective (e.g., verbal report of self or others). The length of baseline was considered to be convincing if there were three or more sessions, partially convincing if there were one to two sessions conducted, and not convincing if there were zero sessions of baseline or baseline was not reported. Base-
line was considered stable and trending in the correct direction if (a) there were two consecutive days of data trending in the correct direction or (b) two consecutive days of stability or (c) three out of four sessions of baseline data were at a stable rate without the last data point heading in the incorrect direction (d) across all skills and/or participants. Immediate effect was defined as (a) two of the first three data points during intervention higher than all data points during baseline or, (b) the third data point of intervention was higher than all data points during baseline. The number of overlapping data points was defined as the percentage of data points during which an intervention data point was the same as any baseline level data point. The percentage of overlapping data points was calculated and was considered to be convincing if there were 0–20% overlapping data points, partially convincing if there were 21–40% overlapping data points, and not convincing if there were 41–100% overlapping data points (all partial percentages were rounded up) Finally, each article was evaluated for any other procedures that were combined with the social story intervention. If the study did not combine another procedure with social stories then it was considered to be convincing, if the study did combine another procedure with social stories then it was considered to be not convincing. Changing criterion. Five variables (see Table 2) were considered when evaluating the level of convincing evidence of the study that used a changing criterion design. The five variables included (1) type of data collected, (2) the stability and direction of the baseline trend, (3) how many sessions until behavior change occurred, (4) variability of data, and (5) if the social story intervention was combined with other procedures. The type of data collected, stability and direction of the baseline trend, and if the social story intervention was combined with any other procedures was defined the same as within case study designs. The number of sessions until behavior change occurred was considered convincing if behavior change occurred within three sessions, partially convincing if behavior change occurred within five sessions, and not convincing if behavior change occurred within seven or more sessions. Little variability (i.e., 80% or more of sessions are within 10% of the mean) in the data was considered to be convincing evidence, moderate variability (i.e., 40–79% of sessions are within 10% of the mean) in the

<table>
<thead>
<tr>
<th>TABLE 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures and Evidence Levels for Changing Criterion Design</td>
</tr>
<tr>
<td>Level of Evidence</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Convincing evidence</td>
</tr>
<tr>
<td>Partial evidence</td>
</tr>
<tr>
<td>No convincing evidence</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>TABLE 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>Measures and Evidence Levels for Multiple Baseline Designs</td>
</tr>
<tr>
<td>Level of Evidence</td>
</tr>
<tr>
<td>---</td>
</tr>
<tr>
<td>Convincing evidence</td>
</tr>
<tr>
<td>Partial evidence</td>
</tr>
<tr>
<td>No convincing evidence</td>
</tr>
</tbody>
</table>

268 / Education and Training in Autism and Developmental Disabilities-September 2020
data was considered to be partially convincing, and high variability (i.e., 0–39% of sessions are within 10% of the mean) in the data was considered to be not convincing.

Multiple baseline designs. There were five variables considered when evaluating studies using a multiple baseline design (see Table 3). The five variables were (1) type of data collected, (2) baseline trend and stability, (3) staggering the intervention correctly, (4) clear change in behavior, and (5) if the social story procedure was combined with other procedures. The type of data, baseline trend and stability, and if procedures were combined were defined the same as within case study and multiple baseline designs. Staggering the intervention correctly was defined as (a) the data trending the correct way on the panel directly above without the previous two data points trending in the incorrect way and (b) the last two data points being higher than 80% of baseline data points. Clear behavior change was defined as (a) 75% of all data points in a condition higher than all of the baseline data points or (b) clear level change through visual analysis.

Group designs. Table 4 displays each variable with each level of convincing demonstration for group designs. The six variables analyzed for a study using a group design were (1) type of data taken, (2) if a comparison group was included, (3) randomization of groups, (4) the type of evaluator used, (5) if a pre-test occurred for the group(s), and (6) if a post-test occurred for the group(s). The type of data was categorized as objective (i.e., observable and measurable), standardized assessment (i.e., a normed/standardized assessment was used), or subjective (e.g., verbal report of behavior). Second, we evaluated if a comparison group was used. For this variable, if a control group or second treatment group was used then it was considered convincing and if there was not a second group then it was considered not convincing. Third, we evaluated if participants in the groups were randomized, quasi-randomized or had a matched sample, or if there was no randomization of participants. Fourth, we evaluated if the evaluators used were blind evaluators, a researcher or teacher was used as the evaluator, or the child or caregiver was evaluator. Fifth, we evaluated if there was pre-test condition
and if the pre-test occurred with both groups, occurred with only the treatment group, or if there was no pre-test. Sixth, we evaluated if there was a post-test condition and if the post-test occurred with both groups, occurred with only the treatment group, or if no post-test was conducted.

**Analysis of Comparative Studies**

There were five studies that compared social stories to another procedure. For these five studies, we evaluated the number of participants in the study, what procedure(s) were compared, what design was implemented to compare the procedure(s), and which procedure was the most effective within the study. Effectiveness was determined by visual analysis of the data and conclusions made by the authors of each of the studies.

**Analysis of Combination Studies**

There were four studies found that evaluated the benefits of adding social stories to another procedure. For these four studies we evaluated the number of participants in the study, the procedure(s) combined with social stories, what design was implemented to evaluate the combination of procedures, and what procedures were found to be the most effective. Effectiveness was determined by visual analysis of the data and conclusions made by the authors of each of the studies.

**Results**

**Evaluation of Research Design**

Table 5 provides the results of the level of evidence for the 13 articles that evaluated social stories in the absence of other procedures. Of these 13 studies, 10 (i.e., 76.9%) demonstrated no convincing evidence, 2 (i.e., 15.3%) demonstrated partially convincing evidence, and 1 study (i.e., 7.8%) demonstrated convincing evidence that the social story intervention was responsible for the behavior change. Tables 6 through 9 provide information found across the variables analyzed for each type of design. These tables provide an overview of how each article was scored across the previously defined variables and where the article would fall in terms of evidence level.

One study used a case study design (i.e., Fees et al., 2014) and was found to demonstrate no convincing evidence. One study used a changing criterion design (i.e., Laprime & Dittrich, 2014) and was also found to demonstrate no convincing evidence. Seven studies used a multiple baseline design. Of the seven studies, one (i.e., Kim et al., 2014) was found to demonstrate convincing evidence, two demonstrated partially convincing evidence (i.e., Halle et al., 2016; Olcay-Gül & Tekin-Iftar, 2016), and four were found to provide no convincing evidence (i.e., Hutchins & Prelock, 2012, 2013; Olçay-Gül, 2016; Vandermeer et al., 2015). Four studies used group designs and all four were found to provide no convincing evidence (i.e., Jeeratok et al., 2014; Mahasneh et al., 2017; Pop et al., 2013; Schwartzberg & Silverman, 2013).

**Results of Comparative Studies**

Seven studies compared social stories to other interventions. The results of the analysis for comparative studies are depicted in Table 10. Across the seven studies, there were a total of
### TABLE 6
Results for Studies Using a Case Study Design

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Type of Data</th>
<th>Length of Baseline</th>
<th>Baseline Trend</th>
<th>Immediate Effect</th>
<th>Overlapping Data</th>
<th>Other Procedures</th>
<th>Level of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fees et al. (2014)</td>
<td>Subjective</td>
<td>N/A (Baseline data not reported)</td>
<td>N/A (Baseline data not reported)</td>
<td>Behavior change occurred after 3 sessions</td>
<td>N/A (Baseline data not reported)</td>
<td>Yes (Social story sang in a song)</td>
<td>Not Convincing</td>
</tr>
</tbody>
</table>

### TABLE 7
Results for Changing Criterion Design

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Type of Data</th>
<th>Baseline Trend</th>
<th>Immediate Effect</th>
<th>Variability</th>
<th>Other Procedures</th>
<th>Level of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laprine &amp; Dittrich (2014)</td>
<td>Objective</td>
<td>100% stable</td>
<td>Behavior change occurred within 3 sessions</td>
<td>High variability</td>
<td>Yes (Discrimination training, differential reinforcement, &amp; response cost)</td>
<td>Not Convincing</td>
</tr>
</tbody>
</table>
# TABLE 8
Results for Multiple Baseline Design

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Type of Data</th>
<th>Baseline Trend</th>
<th>Staggering</th>
<th>Clear Behavior Change</th>
<th>Other Procedures</th>
<th>Level of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Halle et al. (2016)</td>
<td>Objective</td>
<td>100% stable</td>
<td>50-74% correctly staggered</td>
<td>80-100% of all conditions</td>
<td>Yes (Video Modeling)</td>
<td>Partially Convincing</td>
</tr>
<tr>
<td>Hutchins &amp; Prelock (2012)</td>
<td>Subjective</td>
<td>67-99% stable</td>
<td>0-49% correctly staggered</td>
<td>0-49% of all conditions</td>
<td>Yes (Comic Strip)</td>
<td>Not Convincing</td>
</tr>
<tr>
<td>Hutchins &amp; Prelock (2013)</td>
<td>Subjective</td>
<td>N/A</td>
<td>N/A</td>
<td>N/A</td>
<td>No</td>
<td>Not Convincing</td>
</tr>
<tr>
<td>Kim et al. (2014)</td>
<td>Objective</td>
<td>100% stable</td>
<td>75-100% correctly staggered</td>
<td>80-100% of all conditions</td>
<td>No</td>
<td>Convincing</td>
</tr>
<tr>
<td>Olçay-Gül (2016)</td>
<td>Objective</td>
<td>100% stable</td>
<td>75-100% correctly staggered</td>
<td>80-100% of all conditions</td>
<td>Yes (Video modeling)</td>
<td>Not Convincing</td>
</tr>
<tr>
<td>Olçay-Gül &amp; Tekin-Iftar (2016)</td>
<td>Objective</td>
<td>67-99% stable</td>
<td>75-100% correctly staggered</td>
<td>80-100% of all conditions</td>
<td>No</td>
<td>Partially Convincing</td>
</tr>
<tr>
<td>Vandermeer et al. (2015)</td>
<td>Objective</td>
<td>66-0% stable</td>
<td>74 to 50% correctly staggered</td>
<td>49 to 0% of all conditions</td>
<td>No</td>
<td>Not Convincing</td>
</tr>
</tbody>
</table>

# TABLE 9
Results for Group Designs

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Type of Data</th>
<th>Groups</th>
<th>Randomized</th>
<th>Evaluator</th>
<th>Pre-Test</th>
<th>Post-Test</th>
<th>Level of Evidence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jeekratok et al. (2014)</td>
<td>Subjective</td>
<td>No second group</td>
<td>No randomization</td>
<td>Teachers and caregivers</td>
<td>Occurred with treatment group</td>
<td>Occurred with treatment group</td>
<td>Not Convincing</td>
</tr>
<tr>
<td>Mahameh et al. (2017)</td>
<td>Subjective</td>
<td>No second group</td>
<td>No randomization</td>
<td>Researcher</td>
<td>Occurred with both groups</td>
<td>Occurred with both groups</td>
<td>Not Convincing</td>
</tr>
<tr>
<td>Pop et al. (2013)</td>
<td>Subjective</td>
<td>Control groups</td>
<td>Randomized</td>
<td>Blind evaluator</td>
<td>No pre-test</td>
<td>Occurred with both groups</td>
<td>Not Convincing</td>
</tr>
<tr>
<td>Schwartzberg &amp; Silverman (2013)</td>
<td>Subjective</td>
<td>Control group</td>
<td>Randomized</td>
<td>Researcher</td>
<td>Occurred with both groups</td>
<td>Occurred with both groups</td>
<td>Not Convincing</td>
</tr>
</tbody>
</table>
25 participants for whom social stories were compared to an alternative intervention. Social stories were compared to a photo activity schedule (Daneshvar et al., 2018), Cool versus Not Cool™ (Leaf et al., 2016), the teaching interaction procedure (Kassardjian et al., 2014), video modeling alone (Acar et al., 2017; Kagohara et al., 2013; O’Handley et al., 2015) and video modeling in combination with other interventions (i.e., prompting or reinforcement; Malmberg et al., 2015). In all seven studies, the alternative intervention was found to be more effective than social stories.

### TABLE 10
Results of Comparative Studies

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Number of Participants</th>
<th>Compared Intervention</th>
<th>Design</th>
<th>Most Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Acar et al. (2017)</td>
<td>3</td>
<td>Video Modeling</td>
<td>Adapted alternating treatments design</td>
<td>Video modeling</td>
</tr>
<tr>
<td>Daneshvar et al. (2018)</td>
<td>4</td>
<td>Photo activity schedule</td>
<td>Adapted alternating treatments design</td>
<td>Photo activity schedule</td>
</tr>
<tr>
<td>Kagohara et al. (2013)</td>
<td>2</td>
<td>Video modeling</td>
<td>Multiple baseline design across participants</td>
<td>Video modeling</td>
</tr>
<tr>
<td>Kassardjian et al. (2014)</td>
<td>3</td>
<td>Teaching interaction procedure</td>
<td>Adapted alternating treatment design</td>
<td>Teaching interaction procedure</td>
</tr>
<tr>
<td>Leaf et al. (2016)</td>
<td>1</td>
<td>Cool vs. not cool procedure</td>
<td>Adapted alternating treatments design replicated across skills</td>
<td>Cool vs. not cool procedure</td>
</tr>
<tr>
<td>Malmberg et al. (2015)</td>
<td>4;2</td>
<td>Video modeling; prompting and reinforcement</td>
<td>Multielement design and concurrent multiple baseline design across participants; Multiple baseline design across participants and within participants across behaviors</td>
<td>Video modeling; Prompting and Reinforcement</td>
</tr>
<tr>
<td>O’Handley et al. (2015)</td>
<td>6</td>
<td>Video modeling</td>
<td>Multiple baseline design with A/B/B+/C condition sequence</td>
<td>Video modeling</td>
</tr>
</tbody>
</table>

### Results of Combined Studies

Three studies evaluated the effectiveness of combining social stories with another intervention. The results of this analysis are depicted in Table 11. Across the three studies a total of seven individuals participated in studies that combined social stories with Animal Assisted Therapy (AAT), differential reinforcement of zero rates of behaviors (DRO), video modeling, or functional communication training (FCT). Grigore and Rusu (2014) found social stories in combination with AAT to be more effective than social stories alone. Iskander and Rosales (2013) found that social

### TABLE 11
Results of Combined Treatments

<table>
<thead>
<tr>
<th>Author(s) and Year</th>
<th>Number of Participants</th>
<th>Combined Intervention</th>
<th>Design</th>
<th>Most Effective</th>
</tr>
</thead>
<tbody>
<tr>
<td>Grigore &amp; Rusu (2014)</td>
<td>3</td>
<td>Animal Assisted Therapy</td>
<td>Reversal design</td>
<td>Social stories + AAT</td>
</tr>
<tr>
<td>Iskander &amp; Rosales (2013)</td>
<td>2</td>
<td>Differential reinforcement of zero rates of behaviors (DRO)</td>
<td>Multiple baseline design across behaviors</td>
<td>Social stories + DRO</td>
</tr>
<tr>
<td>Pane et al. (2015)</td>
<td>2</td>
<td>Functional communication training (FCT)</td>
<td>Alternating treatment design</td>
<td>Social stories + FCT</td>
</tr>
</tbody>
</table>
stories combined with DRO was more effective than social stories alone. Finally, Pane et al. (2015) found that social stories combined with FCT was more effective than social stories combined with non-functional based communication training.

Discussion

The purpose of this review was threefold. First, in an effort to provide consumers with information on whether researchers have demonstrated functional relationships between social stories and the desired behavior change, we examined the experimental rigor of studies evaluating the effectiveness of social stories implemented in isolation. Second, to provide consumers with a summary of evidence on whether social stories are more effective than other procedures, we evaluated research that compared social stories to other interventions. Finally, to provide consumers with evidence on the additive effects of combining social stories with other procedures, we evaluated the research which combined social stories with other interventions.

Overall across 13 studies evaluating Social Stories™, the results of this analysis demonstrated that only 1 (7.8%) of the 13 studies had convincing evidence that the social story was responsible for the change in behavior, 2 (15.3%) showed partially convincing evidence, and 10 (76.9%) showed no convincing evidence. These findings are consistent with the evaluation of the literature conducted by Leaf et al. (2015) which evaluated the published literature on social stories from 1993 to 2012 and found that only 7.3% of articles demonstrated convincing evidence, 41.5% of articles demonstrated partially convincing evidence, and 51.2% of articles demonstrated no convincing evidence that the social story was responsible for the change in behavior. In other words, a functional relationship has not been established in the majority of studies leaving it unknown if social stories or if some other confounding variables were responsible for the behavior change.

This review found that when Social Stories™ are compared to other interventions, the other interventions have been found to be more effective. For example, Kassardjian et al. (2013) found that social stories resulted in no behavior change and was not even as effective as the no-teaching control condition, whereas the teaching interaction procedure was far more effective. This review also found that social stories have no additive value with other interventions. For example, Iskander and Rosales (2013) found that social stories alone are not as effective as when social stories are combined with differential reinforcement procedures. Given numerous studies have found differential reinforcement to be effective as a standalone procedure (Jessel & Ingvarsson, 2016) and that numerous studies have found social stories to be ineffective as a standalone procedure (e.g., Leaf et al., 2015); it is safe to assume that the differential reinforcement procedure was more responsible for the behavior change.

Given the wide use and popularity of Social Stories™, a clinician may encounter a request from a parent or school administrator to use this procedure. However, it is the clinician’s obligation to kindly and professionally inform others that Social Stories™ has yet to provide sufficient methodological support for their use, and provide alternative interventions that are evidence-based and conceptually systematic with applied behavior analysis. Furthermore, Social Stories™ should not be recommended through informal forums such as social media.

The results of this review have several implications for future researchers. If future researchers evaluate the effectiveness of social stories, it is pertinent to ensure appropriate experimental rigor. This would include, but is not limited to, displaying clear experimental control, developing clear operational definitions, and reporting treatment fidelity and interobserver agreement data. Single-subject research methodology could provide researchers with the tools to evaluate the conditions under which social stories are and are not effective. If single-subject designs are effective at identifying certain conditions under which social stories are effective, researchers could make use of comparative studies of social stories and other established interventions (e.g., video modeling, behavioral skills training, script fading) under the same conditions that social stories are effective in isolation. The results of this line of research could inform when and when not to use social stories.
If the results from these comparative studies identify that social stories are more efficacious, then, larger scale studies evaluating social stories using group design would be warranted. Finally, editors and action editors should only accept research on social stories that demonstrate clear functional control and have other characteristics of quality research.

Given the results of this as well as other reviews of the literature (e.g., Leaf et al., 2015; Reynhout & Carter, 2011; Styles, 2011), that demonstrate the majority of published studies do not support the efficacy of Social Stories™ in isolation or enhancing the effects of other procedures, standards projects should reconsider identifying social stories as an established procedure (e.g., National Autism Center, 2015). Until future researchers using methodically sound designs demonstrate function control, Social Stories™ should not be implemented in practice, recommended, or endorsed, especially when other procedures have been found to be more efficacious.

References


Received: 2 May 2019
Initial Acceptance: 3 July 2019
Final Acceptance: 7 August 2019
Enhancing Early Numeracy Skills of Children with Severe Disabilities and Complex Communication Needs

John C. Wright  
Marist College

Christopher J. Lemons  
Stanford University

Victoria F. Knight  
University of British Columbia

Esther Lindström  
Lehigh University

Julia Strauss  
Peabody College of Vanderbilt University

Abstract: There is a lack of research on effective interventions to improve the early numeracy skills of children with severe disabilities—autism spectrum disorder, developmental disability, and intellectual disability—and complex communication needs. While preliminary research suggests the Early Numeracy curriculum is effective for teaching children with severe disabilities, efficacy of this curriculum has not yet been examined for students with co-occurring complex communication needs. Using a multiple probe across participants research design, we evaluated the efficacy of the Early Numeracy curriculum for increasing targeted math skills of four children with severe disabilities and complex communication needs. Results indicated a functional relation between the intervention and improvements in early math skills. Limitations and future research needs are discussed.

Increasing the mathematical competence of school-aged children remains an essential need in U.S. schools. Gersten and Chard (1999) equate the importance of early numeracy with early literacy skills and note that explicit instruction of foundational numeracy (e.g., counting, set making, number identification) is critical to placing children on a positive trajectory for later academic development. The National Council of Teachers of Mathematics (NCTM) indicated that building a strong foundation in numeracy in the early grades provides significantly enhanced opportunities for work and leisure for all students, including those with disabilities (NCTM, 2000). Early numeracy skills are the mathematical thinking and reasoning skills that provide the foundation for developing more complex skills and include skills such as number identification, rote counting, number conservation, composing and decomposing numbers, magnitude of numbers, early measurement concepts, basic operations, and patterning (Browder et al., 2012). Unfortunately, results from a recent National Center in Education Statistics report show a trend of decreasing mathematics scores on standardized assessments during the last three years and no progress in the last ten years for all children (Kena et al., 2016). This problem is magnified for children with severe disabilities (i.e., autism spectrum disorder, developmental disability, intellectual disability) who have limited opportunities to access mathematics content. Although researchers and practitioners advocate for children with severe disabilities to receive intensive early numeracy math instruction (Browder et al., 2012), few studies have focused on math interventions for children with the most complex needs (King et al., 2016). Although the National Council of Teachers of Mathematics (NCTM, 2000) emphasizes equity and access to effective intervention as essential for all children, Browder et al. (2012) identified a lack of exposure to early numeracy skills in school as a key factor.
limiting the progress of skill acquisition in math for children with severe disabilities. Further, for children with severe disabilities and co-occurring complex communication needs (CCN), math learning is often confined to discrete trials of number identification or rote learning experiences (Jimenez & Kemmery, 2013).

A systematic review of 68 studies found that school-aged children with severe disabilities are able to learn basic math through systematic instruction (Browder et al., 2008). The review documented a limited scope of research often focused on number recall and basic computation. A more recent review demonstrated an increased emphasis on more advanced numeracy skills, including algebra, geometry, and data analysis (Spooner et al., 2019). Information specifically pertaining to children with CCN, however, is lacking in both reviews. For the purpose of the present study, CCN will be used to describe children with minimal to no comprehensible verbal speech, which reflects predominant definition of CCN present in reviewed studies.

With increased focus on high stakes testing, special educators are required to expand the breadth of topics covered in their classrooms (Spooner et al., 2011). Towles-Reeves et al. (2009) and Kearns, et al. (2011) examined the alternate assessments based on alternate achievement standards (AA-AAS) scores for children with severe disabilities in a combined seven states. Students included in these analyses demonstrated significant deficiencies in expressive language, basic reading, and numeracy skills. Of more than 12,000 students in grades 3–12, the results of these analyses demonstrated that no more than 12% of children were able to rote count to five, and only 4% of children could use an operation in context (Kearns et al., 2011).

While math outcomes of children with severe disabilities are often poor, it is likely that children with co-occurring CCN exhibit even greater challenges. First, research focusing on academic interventions for children with CCN often examines functional communication for life skills rather than academic skills. Light and McNaughton (2015) noted the absence of research targeting skills other than communication and advocated for interventions in which communication is a tool for learning content in school. Thus, many educators may be prioritizing communication intervention over academic intervention for this population. Second, no studies have specifically examined mathematic interventions or related outcomes of children with CCN. Additional explorations are needed to understand how mathematical learning can be promoted for learners with severe disabilities and CCN.

In response to weak national scores and the ongoing need for quality math instruction, Browder et al. (2012) developed an early numeracy conceptual model for teaching math to children with severe disabilities. The goal of the model was to highlight practices that would increase access to high quality early numeracy instruction for children with severe disabilities. The model included four key components: (a) systematic prompting and feedback, (b) embedded instruction in general math classes, (c) small group contextualized math stories, and (d) early numeracy targeted skills (number identification, counting, set making, symbolic use, patterns, measurement, and calendar/time). Browder and colleagues conducted a pilot study involving seven children with severe disabilities and demonstrated positive effects of early numeracy instruction on early math outcomes. All children showed an immediate response to the intervention and a positive trend across increasingly difficult content; however, a functional relation between the intervention and student outcomes could not be evaluated due to the use of A-B single case designs.

Based on this pilot study, Jimenez et al., (2013) developed the Early Numeracy (EN) curriculum. Key features were similar to the conceptual model (Browder, 2012) and included: (a) systematic prompting and feedback, (b) contextualized math stories, (c) task analytic instruction, (d) multiple exemplar training, and (e) opportunities for embedded instruction in grade-level math classes. Jimenez and Kemmery (2013) completed an initial efficacy study on the EN curriculum treatment package using a multiple probe across participants research design. All participants had a moderate intellectual disability (ID), and two of the participants also had autism diagnoses. Results showed that all five children increased math skills from baseline to intervention. This study demonstrated a functional relation be-
tween the EN curriculum and the acquisition of foundational math skills (e.g., one-to-one correspondence, number identification, making sets of objects).

Due to demonstrated efficacy of EN for students with severe disabilities and limited guidance on teaching early math skills to children with CCN, the purpose of the present study was to evaluate early numeracy instruction for children with severe disabilities and CCN. The research question guiding the study was, “Is there a functional relation between the EN curriculum and improvements in targeted math skills of elementary-aged children with severe disabilities and co-occurring CCN?”

Method

Participants

Using teacher recommendations and targeted flyers, four students with autism spectrum disorder (ASD), developmental disability, or intellectual disability and co-occurring complex communication needs were recruited to participate in this study (see Table 1). Aiden is a six year old diagnosed with ASD. He receives speech therapy three times a week, has approximately 10 intelligible words in his vocal-verbal repertoire, and uses picture exchange communication system (PECS) for most social and academic requests. Darius is a six year old diagnosed with ASD, ID, and epilepsy. He receives speech therapy twice weekly, has approximately 20 words in his vocal-verbal repertoire, and relies mostly on gestures and PECS to articulate his wants and needs. He had no occurrences of seizures during the study. Ricky is a six year old with a primary diagnosis of Down syndrome. He receives speech therapy twice weekly at school, has approximately four intelligible vocal-verbal words in his repertoire, and communicates primarily via gestures and unintelligible sounds. Sierra is a 10 year old diagnosed with ID and epilepsy. She has no spontaneous vocalizations but has an echoic repertoire of 10 words. She receives speech therapy three times a week and is learning PECS. She had on occurrences of seizures during the study.

Criteria for selection for inclusion in this study included: (a) having a CCN, including usage of augmentative and alternative communication (AAC) device, picture communication board, or minimal/emerging vocal verbal repertoire; (b) ability to independently and consistently respond to prompts and communicate selections with verbal approximations, pointing, gesturing, or accurate manipulation of EN materials; and (c) having mastered less than 50% of targeted math skills prior to intervention. The Peabody Picture Vocabulary Test (PPVT), researcher observations, and teacher report were used to examine participants’ ability to respond to verbal and visual prompts (Abkarian et al., 1987). EN screening probes (Browder et al., 2009) were used to determine pre-intervention mastery of early numeracy skills. After consent was received for all children, screening indicated that one student did not meet eligibility requirements due to lack of ability to communicate choice given a visual and verbal prompt.

Implementers

The first implementer was the principal researcher. He was Caucasian with 13 years teaching experience and pursuing a doctoral degree in special education. The second implementer was Caucasian with two years experience as a registered behavior technician and was completing her master’s degree in special education, concentrating in severe disabilities.

Settings

This study took place in two classrooms at the same public elementary school in a large urban school district in the southern United States. The classrooms were both approximately 10x10m and held a variety of child-sized tables, chairs, and bookshelves. The classrooms were divided into sections or centers (e.g., library, one-on-one worktables, computer area, group work area) using carpets and movable dividers. One classroom had an interactive smartboard, and the other had a digital projector.

Most sessions occurred inside the classroom during small group math instruction time. One student worked at a hallway workstation approximately five feet from the classroom door. This included one child-sized trapezoidal table and two chairs. The primary teacher
<table>
<thead>
<tr>
<th>Child</th>
<th>Gender/Ethnicity</th>
<th>Grade (Age)</th>
<th>Educational Label</th>
<th>Relevant Testing</th>
<th>Primary Communication</th>
<th>Math ability (Teacher report)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Aiden</td>
<td>Male, African American</td>
<td>1 (6)</td>
<td>Autism, Intellectual Disability</td>
<td>Autism Spectrum Rating Scale: Total score = 70 (very elevated) &amp; social communication = 75 (very elevated); Childhood Autism Rating Scale: Verbal = 30, Nonverbal = 32</td>
<td>Emerging vocal verbal repertoire, intelligible vocabulary of 10–15 words, gestures and PECS</td>
<td>Demonstrates 1:1 correspondence (as teacher counts to 20)</td>
</tr>
<tr>
<td>Darius</td>
<td>Male, African American</td>
<td>1 (6)</td>
<td>Autism, Intellectual Disability, Epilepsy (with frequent partial seizures)</td>
<td>Composite score = 73 (CTONI); ABAS-2 Composite score = 76 (parent), 54 (teacher); limited verbal</td>
<td>Emerging vocal verbal repertoire, intelligible vocabulary of 20–30 words, gestures and PECS</td>
<td>Receptively identifies numbers up to 40 from a field of 2, expressively identifies 2D and 3D shapes, matches number to quantity for numbers 1–5</td>
</tr>
<tr>
<td>Ricky</td>
<td>Male, African American</td>
<td>1 (6)</td>
<td>Developmental Delay/ Down Syndrome</td>
<td>Reading = 73, Math = 50 (Woodcock Johnson IV, average range = 90–110); ABAS-2 Composite score = 49 (parent), 61 (teacher)</td>
<td>Non-vocal verbal, verbal repertoire, intelligible vocabulary 2–4 words, gestures</td>
<td>Receptively identifies numbers 1–5 from a field of three. Not able to identify number 6–10 or match number to quantity</td>
</tr>
<tr>
<td>Sierra</td>
<td>Female, Hispanic</td>
<td>4 (10)</td>
<td>Intellectual Disability/Epilepsy</td>
<td>Composite score = 70; all scores in ‘poor’ or ‘very poor’ range (CTONI-2) ABAS-2 (all subscores fell in the extremely low category)</td>
<td>Minimal vocal verbal repertoire, intelligible vocabulary of 5–10 words, gestures, PECS</td>
<td>Counts (with verbal approximations) to 20 with 80% accuracy; receptively identifies numbers 1–9</td>
</tr>
</tbody>
</table>

*Note: ABAS-2 = Adaptive Behavior Assessment System; CTONI-2 = Comprehensive Test of Nonverbal Intelligence – Second Edition; PECS = Picture exchange communication system.*
and paraeducators worked with groups of children on math skills while participating children were in intervention with a researcher. The primary teachers in both classrooms were certified special education teachers with bachelor’s degrees. There were at least two other paraeducators in the classrooms at all times providing support.

The primary researcher and one graduate assistant conducted all intervention sessions at a child-sized table. Each worked with two of the four children during one-on-one sessions. For classroom-based sessions, movable dividers separated the workspace from the rest of the class.

Materials

During baseline and intervention sessions, children used materials provided by the researcher from the EN curriculum; foundational support components include math storybook, visual organizers, number lines, student response book, assessments, and story-specific manipulatives produced by the Attainment Company (Jimenez et al., 2013). All baseline and intervention sessions were recorded using Vixia mini video recorder.

Response Definitions and Measurement Systems

The primary dependent variable was the percentage correct of targeted math skills listed in the objectives of the EN curriculum (Jimenez et al., 2013). The EN curriculum consists of four units, with six lessons in each unit. Each unit targets 12 standards-based early numeracy skills.Probe questions were asked of all skills listed in the unit objectives. These skills included counting, making sets, using symbols, identifying and extending patterns, measuring, using a calendar, and identifying numbers. Because verbal responses were not expected due to participants’ CCN, researchers allowed the following modifications to the response: independently pointing to correct numbers (number identification), one-to-one correspondence using number lines or manipulatives (counting, making sets, calendars), independently pointing to or creating patterns with manipulatives (patterns), independently pointing to symbols ( =, <, >), and accurate manipulation of non-standard measurement items (small cubes, paper clips, etc.) and standard measuring tools (rulers). Non-examples of targeted responses included randomly touching numbers or number line, echoic imitations (approximations) of researcher prompt, prompted correct answers, or touching visual organizers without looking at them.

Baseline sessions lasted approximately 10–15 minutes and ended when all probe questions had been asked. Each probe session had 10 or 11 questions with a short break (three minutes) after approximately five questions had been asked. Immediately after each question, the researcher recorded whether or not the student emitted the correct response. In the event that the participant’s pace a set made the session too long or a student demonstrated problem behavior, the researcher ended the session. When problem behavior arose, researchers first used the classroom teacher’s prompt, “Show me you’re ready to work” with a model of being seated with hands in lap. After two attempts for using this prompt, sessions were ended. In any of these cases, answers were still recorded and documented, and the next session started where the previous session ended.

Intervention sessions lasted approximately 20 minutes. Sessions included the same procedure for collecting data as in baseline, plus an additional teaching component from the EN curriculum. A sample interaction for a targeted math skill during intervention occurred as follows: (a) probed targeted skill, (b) recorded response on data collection sheet, (c) taught same targeted skill with math story and hands-on materials from EN curriculum using constant time delay (CTD), and (d) probed next targeted skill.

Interobserver Agreement

Interobserver Agreement (IOA) was collected for at least 25% of the baseline and intervention conditions for every student. IOA sessions were chosen at random and were calculated using a point-by-point method (Ayres & Ledford, 2014). Percentage agreement was calculated by counting the number of agreements and dividing this by the sum of agreements and disagreements. The result was multiplied by 100.
A graduate student unfamiliar with the participants conducted the IOA sessions. During training on IOA procedures, he was given a coding manual for IOA training and taught how to code probe sessions using videos of children not participating in the study. The primary researcher jointly coded sessions with him to gain consensus. The observer then coded two videos independently and became reliable when IOA was at least 80% with the primary researcher. Retraining would have occurred if IOA fell below 80%, but this did not happen.

Experimental Design

This study used a multiple probe (MP) design (days) across participants (Gast, Lloyd, & Ledford, 2014). MP was chosen because (a) learning new math skills was not a reversible behavior; and (b) this study focuses on an academic intervention during which intermittent collection of data prior to intervention is not prolonged nor aversive. Additionally, a MP across participants design allowed for a systematic demonstration of experimental control by sequentially introducing the EN intervention across individuals in a staggered, tiered approach. This approach helped control for threats to internal validity by requiring stability in level and trend prior to introduction to the next condition. Through visual analysis, confidence in the intervention increased when an immediate and abrupt therapeutic change in dependent variable was observed upon introduction of the independent variable.

Visual analysis was used to examine the functional relation between the implementation of the EN curriculum and the acquisition of targeted math skills. A minimum of three stable or therapeutically positive data points were required before a student could enter a new condition of the study. In order to ensure that all participants were able to enter intervention before the end of the school year, mastery of Unit 1 skills were not required for the student in the preceding tier. However, researchers did focus on each participant having a therapeutic trend and stable responding prior to subsequent tiers entering intervention. If a student entered into intervention and failed to show improvement in targeted math skills after four consecutive sessions, individual modifications to the prompting system and adjustments to token board reinforcements occurred.

Procedure

Modifications of EN curriculum for children with complex communication needs. Researchers modified the targeted response for the skill of numeral identification, allowing children to point to an EN number line instead of emitting a vocal verbal response. Additionally, the rote counting skill and number naming skill were removed from all probes for all children across conditions.

Probe Procedures

Written procedures were used during screening and baseline, intervention, and generalization probes. Sessions occurred once a day, four days a week. Two graduate students implemented this intervention for all participants, and each graduate student worked with two children.

Prior to baseline, the primary implementer observed each student a minimum five times. Each observation was approximately 20 min. Due to the fact that each student had CCN, the goals of the observations were to: (a) assess the receptive communication skills of each student; (b) observe student use of communication (e.g., AAC devices, communication boards, and teacher-specific methods); (c) understand use of each student’s token economy system; (d) understand current math instruction for participants; and (e) build rapport with children.

Screening

Prior to baseline probes, four scripted screening probes—one probe for each EN unit—were conducted to assess participants’ ability on the entire set of skill taught in the EN curriculum. Numerals used for each objective in the probes were randomized (e.g., “Point to the numeral 3,” “Make a set of 4 cubes”). Due to the fact that each unit builds in complexity, the screening probes were completed in order to verify that Unit 1 was the most appropriate starting point for each child. One screening
probe was administered each day. To avoid frustration and minimize the chance of becoming aversive, screening probes were terminated if participants made an error or had no response for five consecutive probe questions. Screening indicated that Unit 1 was appropriate for all participants.

**Baseline**

Each baseline probe took 10–20 min to administer, and all probes were scripted (e.g., “Which group of apples has more,” “Point to the numeral 1”). Breaks in administration of the probes were incorporated into the procedure using each child’s individual token economy system. Breaks lasted approximately 3 min, at which point token boards were reset. Initially, one token per response was used as reinforcement. To mirror token economy systems in participants’ classrooms, implementers incrementally increased the number of responses required for reinforcement to 2:1 or 3:1, modifying reinforcement depending on the needs of each student. Feedback was given based on engagement only (e.g., “Your eyes are looking at the picture. You are very focused.”), with no math-specific praise or error correction. Initial probes indicated that Unit 1 was the appropriate starting point for each student, with each student demonstrating less than 45% mastery of Unit 1 skills.

Unit 1 baseline probes consisted of 11 questions across the domains of counting, sets, symbol use, measurement, patterns, calendar, and number identification. A minimum of three baseline probes for Unit 1 was required for all children prior to the first student beginning intervention. Baseline stability in level, trend, and variability was required for all children prior to intervention.

**Intervention**

Once baseline stability was obtained, the first student began intervention while the remaining children received intermittent probes for Unit 1. Intervention probe procedures were the same as baseline. Based on the recommendations from the EN curriculum, each intervention session included the following components: warm up, probe skill one, teach skill one using EN scripted curriculum, probe skill two, teach skill two using EN, etc. Neither error correction nor prompting occurred during probes. All skills were taught with a modified version of constant time delay (CTD), as dictated by the EN script. During warm up, 0 s CTD was used to activate prior knowledge and engage students in the session by ‘waking up’ the numbers. These consisted of one or two scripted probes (e.g., “Wake up the number 2 by pointing to it,” or “Count to 6 with me.”).

Intervention sessions on each of the targeted skills began with the “test” phase of CTD procedure (5 s delay round). After 5 s without a response from the student or after an incorrect response, the implementer used the least restrictive prompt to elicit the target behavior for error correction. For example, to assess number identification, the instructor delivered an attentional cue, followed by the task direction, “Point to the number 4.” If the student failed to respond or incorrectly responded, the implementer modeled while pointing, “This is the number 4. (removes pointing gesture) Your turn.” If the student failed to respond or incorrectly responded after this prompt, the instructor led the student by saying, “If you don’t know, wait and I will help you. This is the number 4. Touch 4 with me.” While individual needs of the children dictated the level of prompting needed, all children responded to verbal and gestural prompts.

When the first student showed a consistent, therapeutic change, ideally in level, trend, variability/stability, immediacy of effect, consistency of data, and non-overlapping data to prior condition, the remaining participants’ data required three additional Unit 1 baseline probes to demonstrate their continued stability prior to the second student entering intervention. To show experimental control, this procedure was repeated for the remaining children in order for them to enter intervention. Due to the variety of new skills required to show mastery of Unit 1, the severity of participants’ intellectual disabilities and the CCN, mastery of Unit 1 was not required for the next student to enter into intervention. A clear, therapeutic change, as described above, was necessary before the next student could enter into the intervention condition. Additionally, leaving children in an intermittent probe condition while other children were
engaged in intervention posed potential ethical concerns; therefore, we minimized the time children were not receiving math intervention.

Mastery criteria for Unit 1 skills required children to correctly respond to a minimum of nine out of 11 probe questions in three consecutive sessions. Because many skills from Unit 2 were similar (but increasing in difficulty), these probes served as generalization probes. Subsequent Units would follow identical procedures; however, none of the children progressed past Unit 2. Mastery criteria for Units 2, 3, and 4 were 9 out of 11, 8 out of 10, and 8 out of 10, respectively.

Procedural Fidelity

Implementer training was conducted in a quiet classroom on a university campus in the southeastern United States. Implementers had a checklist of all EN components, strategies, and materials needed for each lesson as outlined in the EN curriculum. Similar to the training used by Jimenez and Kemmery (2013), implementers were trained using a model-lead-test (MLT) format (Engelmann & Becker, 1982). MLT is also known as I Do-We Do-You Do and is effective in promoting active engagement and increasing student achievement. Training lasted approximately five hours and ended when each implementer was 100% accurate using the training checklist. After training, implementers were required to pass a fidelity check with the primary implementer by demonstrating how to deliver the entire first lesson without using the checklist with at least 95% accuracy.

During the study, the research team recorded a minimum of 20% of all sessions for all implementers across children and conditions. A graduate student who successfully completed procedural fidelity training coded videos for implementation fidelity. If fidelity was below 80% for any implementer, he/she would be retrained according to the procedure above; however, this was not necessary.

Social Validity

A social validity questionnaire, developed by the primary researcher, was given to the educators (two special education teachers; six paraeducators) familiar with the study participants. As a part of this assessment, educators watched researchers deliver one baseline and one intervention session to assess session procedures and outcomes. After watching each session, educators were asked about the feasibility and appropriateness of the procedures and content.

Results

Effectiveness Data

Darius’ baseline accuracy on the EN probes was approximately 38%. However, his relatively stable level was slightly higher than the other children, and he was the first student to enter the EN intervention. He had an immediate increase in level upon entering intervention (approximately 54% correct), and consistently held this level of effect through the next three sessions of intervention (see Figure 1). After one session with a decrease to baseline levels, Darius continued to demonstrate a steady increasing trend of math skill acquisition. He met mastery criteria for Unit 1 after 22 sessions. Following the procedures outlined above, Darius was the only student who moved into Unit 2 intervention.

After ten baseline sessions, Sierra began intervention. Although Darius was not yet at mastery level, his consistency of effect, immediacy of response, and increasing trend merited starting the intervention with the next student. Sierra’s initial baseline probes showed stability with low levels of accuracy on math probes (approximately 6%). Upon entering intervention, Sierra had an initial, positive therapeutic response (33% accuracy); however, this effect was not maintained with two of the next three sessions, resulting in 0% accuracy. After four intervention sessions, the intervention was modified to teach fewer math skills, as it was hypothesized that there were too many skills being taught at once, confusing the participant. The modified intervention included only counting, set making, symbol usage, and number identification. This constituted a total of six probe items, compared to the initial eleven probe items. After the intervention was modified, Sierra’s percentage of correct items immediately increased, with the first three modified interven-
Figure 1. Participant Baseline and Intervention Data.
tion sessions resulting in 17%, 33%, and 50% accuracy, respectively. After thirteen intervention sessions, Sierra’s level had stabilized (approximately 42% correct) and showed a clear change in level from baseline. Additionally, her trend was accelerating. In graphing Sierra’s data, percentage of correct responding in baseline data reflects the reduced number of skills taught during intervention.

With intermittent probing, the final two children were clearly still performing at low levels of accuracy on the probes. Although neither Darius nor Sierra met mastery criteria for Unit 1, Aidan was brought into intervention. Darius and Sierra both demonstrated a clear response to intervention, and implementers believed that repeated baseline probing might soon become aversive to Aidan, whose baseline accuracy was approximately 17%. Initially, Aidan had a minimal, but immediate response to the modified intervention; however, he returned to baseline levels for several sessions. His teacher reported that he did not receive his prescribed attention medication for three sessions (sessions 24, 25, and 26); his decreased performance on these days may have been due to this change.

Due to time constraints in the school year, Ricky was brought into intervention at session 25. His baseline accuracy was approximately 17%. He showed an immediate increase in level during four sessions of the intervention, and his average correct item percentage increased to over 50% for his last four sessions (more than double his baseline average).

**Generalization.** Generalization probes of Unit 2 during baseline and intervention show that all participants were able to apply EN skills to tasks beyond the current level of instruction. Darius could complete 1 of 6 of the Unit 2 probes during baseline and an average of 4 (range = 2–6) after intervention; Sierra could complete none of the Unit 2 probes during baseline and an average of 2.3 (range = 1–3) after intervention began; Ricky could complete none of the Unit 2 probes during baseline and an average of 1.5 (range = 1–2) after intervention; and Aidan could complete none of the Unit 2 probes during baseline and an average of 1 (range = n/a) after intervention.

**Social Validity**

Eight elementary special education teachers and paraeducators completed social validity questionnaires to rate the efficacy and feasibility of the EN curriculum. All respondents independently marked ‘strongly agree’ (= 1) or ‘agree’ (= 2) to all questions (average scores on a five point scale are in parentheses): (a) procedures in the video were appropriate for children with moderate or severe disabilities (M = 1.2); (b) procedures in the video were appropriate for children with CCN (M = 1.9); (c) procedures were effective in teaching math to children with disabilities (M = 1.7); (d) teachers would find this intervention helpful (M = 1.7); (e) outcomes of this intervention were evident and effective (M = 1.9); and (f) I would want access to this intervention in my classroom (M = 1.6).

**Reliability**

IOA ranged from 88–100%, with an average of 94% across all children, implementers, and conditions. Procedural fidelity averaged 99.4% (97.5-100%) across implementers, children, and all conditions. If procedural fidelity had dropped below 80%, implementers would have been retrained. However, this was not needed due to consistently high fidelity.

**Discussion**

The purpose of this study was to evaluate the efficacy of the EN curriculum to improve early math skills of four elementary students with severe disabilities and CCN. We extended previous efficacy studies of students with severe disabilities (e.g., Jimenez & Kemmery, 2013) to evaluate efficacy of the curriculum on students with severe disabilities and concomitant CCN. We made research-based adaptations to the curriculum to reflect needs specific to students with CCN and administered intervention four times per week.

Visual analysis of student data indicates clear demonstration of effect (i.e., improvement in level and trend) for three participants and minimal response to intervention in the final participant. Increases in level and trend indicated positive response to intervention on proximal measures including number identi-
fication, counting, and making sets. This evidence of a functional relation between the EN curriculum and early numeracy outcomes suggests promise in teaching math to students with concomitant severe disabilities and CCN. Intervention effects may be attributed to the systematic nature of the intervention: special education teacher reports indicated that the EN curriculum was more structured and focused than business-as-usual math instruction students received prior to intervention. Teacher reports of high social validity support this potential explanation of effects and the feasibility of implementing the EN curriculum.

Relation to Previous Studies

Our findings align with those from previous efficacy studies of early numeracy instruction for students with severe disabilities. Previous meta-analyses support the instruction of early numeracy skills for students with severe disabilities (see Browder et al., 2008), including use of the EN curriculum, in particular (Jimenez & Kemmerly, 2013). Given the lack of research specifically focused on math intervention for students with co-occurring CCN (see Browder et al., 2008; Spooner et al., 2019), we cannot draw comparisons with existing research on this topic. The present study extends knowledge of effective instructional methods for students with CCN, including CTD, prompting, and use of visual supports. Precise examination of these research-based methods and content contributes greater theoretical and practical knowledge regarding academic intervention for students with severe disabilities and CCN.

Limitations

A few limitations to the present study invite greater consideration. First, we assessed skills immediately prior to instruction on each targeted skill (versus conducting cold probes). This method of assessment may have affected learning, as students had an opportunity to practice the skill incorrectly prior to being taught the skill with error correction and reinforcement. This also extended sessions longer than the curriculum intended and may not be representative of how practitioners would implement the instructional protocol in typical classrooms. Similarly, the curriculum is designed to be used in small groups, and the one-on-one format may not be representative of what is feasible in schools. Finally, due to the end of the school year, we had to cease instruction prematurely, were not able to assess maintenance or fully realized generalization probes for all participants, and we were not able to assess most students’ skills beyond those in Unit 1. Future research may target these characteristics to examine intervention feasibility.

Implications for Research

Further studies are needed to evaluate early numeracy and math interventions for students with severe disabilities and CCN. Additionally, more studies are needed to determine whether the findings from our study generalize to other age groups and learner profiles. Individual differences among students with CCN (e.g., language, cognition, behavior) may result in differential outcomes in response to intervention. For example, one student may require specific training on an electronic communication device, whereas another may benefit from a visual schedule to anticipate lesson components. Additional research is needed on other math interventions and distal measures to understand early numeracy and math instruction for students with CCN more generally.

More intervention research will help to identify instructional content and methods that optimize academic learning for students with severe disabilities and CCN. Shown to be effective for students with higher-incidence disabilities, procedures such as data-based individualization (DBI; National Center on Intensive Intervention, 2012) may be appropriate for enhancing academic instruction for students in this population by targeting co-occurring language, cognitive, and behavioral needs. Future studies may examine systematic use of research-based qualitative and quantitative adaptations in the DBI framework to improve response to academic intervention.

Implications for Practice

The findings from our study have implications for special educators teaching academics to
students with severe disabilities and CCN. Recent calls have been made to increase research focus on improving academic outcomes for children with CCN (Dymond, 2015). Given higher academic standards for all students with disabilities in high-stakes assessments and the paucity of research available on academic instruction for students with CCN, our study contributes more precise knowledge regarding academic instruction for students in this population. Further, this research indicates that EN might provide a feasible alternative for special education teachers supporting students with CCN.

First, it is important that teachers ensure that children participating in math intervention have the means to clearly and efficiently communicate with the instructor and peers. As Light and McNaughton (2015) highlighted, communication is a tool for learning content in school. Supporting students to communicate academic content and personal needs (e.g., desire to take a break) may include enlisting low- and/or high-tech communication systems, as well as expertise from related service providers (e.g., speech-language pathologist, augmentative communication specialists). Because children are often learning new academic skills and methods to communicate about these skills (e.g., AAC, gestures, pointing) at the same time, teachers may consider whether communication methods allow the child to adequately demonstrate targeted learning. For example, in the present study, we opted to remove the skill of rote counting from our assessment because we were unable to determine how children could demonstrate this skill using AAC, gestures, or pointing. It appeared to us that if children used AAC with numerals to rote count they would be demonstrating number identification and sequencing instead of the targeted objective.

Second, the study’s findings on social validity underlined the need for more feasible academic interventions for this population. Considered in relation to feedback from study interventionists, special education teachers’ ratings reflect the importance of research-based academic instruction that is accessible in scope and materials. Fewer moving parts may be more appropriate for students with CCN exhibiting problem behaviors when presented with many manipulatives resembling toys. For students in this designation, more streamlined lessons and simpler materials may be beneficial. For this reason, teachers are encouraged to consider individual differences among students systematically.

Overall, our study extends knowledge of research and practice on early numeracy instruction for students with severe disabilities and complex communication needs. Our findings highlight potential promise in academic intervention for students in this population, as well as considerable need for further examination.

References


program for students with moderate and severe disabilities. Attainment Company.


Received: 2 May 2019
Initial Acceptance: 3 July 2019
Final Acceptance: 17 August 2019
Abstract: The importance of helping all children make meaningful progress in reading development is supported by legislative mandates and court rulings. The purpose of this research was to describe findings from a series of single-case design studies completed over three years to test the effects of a specific comprehensive literacy intervention, Friends on the Block, on word recognition. Participants were 18 students with intellectual and developmental disabilities (IDD) ranging in age from 6 to 13. Results were analyzed using Tau statistical procedures that indicate consistent moderate-to-strong positive effects across all participants for both sight words and decodable words. Across the 67 sight word studies the mean effect size was 0.73, and across the 8 decodable word studies the mean effect size was 0.74. Further research is needed to examine the impact of the intervention on a broader array of variables. Implications, limitations, and directions for future research are discussed.

Educators agree that literacy provides students with intellectual and developmental disabilities (IDD) improved opportunities and outcomes, including access to social relationships, employment, recreational activities, independent learning, self-care, and quality of life (Cihak & Smith, 2018; Conners, 2003). Low literacy is related to poor long-term outcomes including low employment, reduced earnings, crime, depression, and substance abuse (Reynolds et al., 2011). Although historically educators have not emphasized literacy for students with IDD, increasing accountability and focus on accessing content are raising expectations for students with IDD (Browder & Spooner, 2014; Katims, 2001). Legislative mandates (e.g., Every Student Succeeds Act, 2015; Individuals with Disabilities Act, 2004) require that schools raise expectations and the Supreme Court recently ruled that schools need to demonstrate students with special needs are making reasonable growth (Endrew F. v. Douglas County School District, 2017). Students with IDD experience the most pervasive challenges in learning to read and write (Wei et al., 2011).

The research described in this article was supported by Grant R324A130102 from the Institute of Education Sciences, U.S. Department of Education. Nothing in the article necessarily reflects the positions or policies of the federal government, and no official endorsement by it should be inferred. We also appreciate the support of the students, parents and school professionals who participated in this research. Disclosure: Allor, Cheatham, and Al Otaiba disclose that they have a financial interest in the books and lessons referred to in this article as Friends on the Block (FriendsOnTheBlock.com). Correspondence concerning this article should be addressed to Jill H. Allor, Southern Methodist University, Department of Teaching and Learning, Simmons School of Education and Human Development, Post Office Box 750455, Dallas, Texas 75275-0455. E-mail: jallor@smu.edu

© Division on Autism and Developmental Disabilities
For example, Lemons and colleagues (Lemons et al., 2012; Lemons et al., 2015; Lemons et al., 2017) studied the effects of two evidence-based reading programs, as well as adaptations of these programs based on the Down syndrome phenotype. They found moderate evidence for a functional relationship between two interventions and word reading (Lemons et al., 2012). For an adapted program targeting phonological awareness, a functional relationship was established and gains were modest (Lemons et al., 2015). For an adapted program targeting letter-sound knowledge, decodable words, and high-frequency words, a functional relationship was established for three out of seven participants. To build literacy skills for students who are nonverbal or who have extremely limited verbal ability, Browder and colleagues created an early literacy intervention, the Early Literacy Skills Builder (ELSB), designed to facilitate receptive responses (Browder et al., 2012). The researchers conducted a randomized control trial with ELSB and found statistically significant differences in favor of the treatment group on phonemic awareness, early phonics, and conventions of reading (Browder et al., 2012).

Although this existing research base is encouraging and should raise expectations for students with IDD, it also reveals challenges specific to teaching students with IDD to read (Allor et al., 2014). First, it is difficult to provide interventions with sufficient intensity given that students with IDD progress slowly and require extensive practice to master skills. Practice should include basic skills and strategies, as well as reading connected text. In our own prior research, we found that even with 45 minutes of daily instruction using a high-quality evidence-based program, students needed 1.5 to 3.5 academic years to make gains on oral reading fluency that typical students make in only half an academic year (Allor et al., 2014). Second, students with IDD struggle to transfer skills from one task to another (Conners et al., 2006). Thus, teachers must not only teach skills to students, but they must explicitly teach them when to apply skills (e.g., need to learn how to sound out words as they read, but also when they no longer need to use that strategy). Third, low levels of language that limit listening and reading comprehension are also characteristic of students with IDD. Appropriate reading material should provide scaffolds to support language and comprehension. There are not enough engaging and meaningful books to provide these scaffolds, support practice, build fluency and reading stamina, and promote transfer of skills (Cihak & Smith, 2018).

Researchers continue to explore ways to further tailor instruction to improve literacy outcomes for students with IDD (see Connor et al., 2014 and Lemons et al., 2017). In this article, we focus on a literacy intervention, Friends on the Block, that we recently developed for students with IDD (Allor, Cheatham, & Al Otaiba, 2018). The intervention combines evidence-based early literacy practices with evidence-based instructional practices for students with IDD, addressing the relative strengths and challenges common in students with IDD. Our goal was to provide instruction that would be consistent with principles of effective instruction and theoretical models of reading, while also adapting our techniques and materials to better address the needs of students with IDD. For example, we use explicit modeling, cumulative review and practice, and feedback when teaching basic decoding skills as is common in evidence-based interventions, but we also provide explicit instruction in the application of those skills to text, as well as constant time delay, an evidence-based practice for students with IDD (Cohen et al., 2008).

Although a thorough description of the intervention is outside the scope of this article, we describe several key instructional design features here and provide illustrative examples (see Allor et al., 2018 and friendsontheblock.com for further details). First, we created an efficient scope and sequence that supports explicit instruction, cumulative review and practice for transfer of skills across instructional strands. For example, the first levels of the intervention target small sets of sight words because evidence strongly supports that students with IDD have the ability to memorize sight words (see Browder et al., 2006). While learning sight words, lessons also include basic phonemic awareness and letter-sound learning. In later levels, decodable words are added. These decisions were based on our prior work in which we ob-
served that students had memorized sight words and letter sounds, but needed more time to practice the phonemic awareness skills of blending and segmenting (Allor et al., 2013; Allor et al., 2014). This also made it possible for students to read meaningful sentences as sight words are frequently used in common spoken language (e.g., I do not see dad). In our previous research, we found that many students needed to cycle back through lessons and that requiring mastery for a daily lesson was impractical (Allor et al., 2014). Similarly, Lemons and colleagues (Lemons et al., 2012) recommend focused review due to retention challenges. Therefore, we organized this new scope and sequence into levels, providing more flexible pacing and extensive cumulative review and practice within and across levels.

Second, we developed the curriculum to center around specially designed books with lessons that align with each book, enabling us to provide students with meaningful, scaffolded text and the opportunity to transfer skills from lessons to text. Sets of target words were carefully selected so they could be combined into simple sentences that resemble spoken language. We added several features to our texts to provide important scaffolds, enhancing word recognition, comprehension, and engagement: (a) picture words to scaffold imageable content words that had not been taught, (b) helper text that is read by the teacher/tutor, and (c) text bubbles in the early books. These scaffolds, along with supported reading techniques, such as echo reading, make it possible for students to begin reading meaningful books in the first level of the intervention. This is one way we address the need to explicitly teach students to transfer skills from one task to another (e.g., students practice sight words in lessons and in the corresponding book). Most literacy programs for struggling readers emphasize decoding and introduce sight words at a relatively slow pace; research is needed to determine optimal pacing of sight words and phonics skills (Castles et al., 2018). Including sight words early in our program provides students with the opportunity to transfer skills to text quickly, while they are learning to decode, which is particularly important for students with IDD who require extensive repetition over months, or even years to master decoding skills.

In an initial single-case design (SCD) study using a multiple baseline across levels of instruction, we demonstrated the initial promise and feasibility of the intervention with eight participants (ages 6–13 years; IQs 40–63) with IDD (Allor et al., 2018). Considering each participant as an independent replicated experiment, we tested for a positive relation between the intervention and word reading. Word reading was selected as the target variable for the SCD study because it is a primary outcome of the intervention and a key skill required for overall reading success; further, SCD necessitates focus on one primary outcome variable. For each participant, we obtained a statistically significant non-overlapping data effect size. The median effect overall was, .80 (considered to be a medium effect size), supporting the hypothesis of a functional relation between participating in the intervention and improved student outcomes. In addition to word reading, students also demonstrated growth and transfer of skills on several early literacy distal measures, although outcomes were somewhat variable. Feasibility and usability evidence were positive and demonstrated that Friends on the Block was acceptable to teachers and students as end users. Teachers reported high levels of satisfaction, indicating that they found the intervention was feasible, easy to use, engaging, and effective for their students. Similarly, families were highly satisfied, reporting positive changes in their child’s interest and ability in literacy and language. The final product incorporates suggestions by end users. Teachers and paraprofessionals implemented the intervention with high fidelity overall, though dosage was somewhat lower than our recommendation. Teachers delivered an average of only three, rather than our intended four 20-minute Instructional Lessons and paraprofessionals delivered an average of only two of four 20-minute Reinforcement Lessons per week.

**Purpose and Research Questions**

Adding to our initial study data, using the same SCD with an additional 10 participants, we have now completed a total of 18 independent replicated experiments. The purpose of
the present study is to describe the aggregated findings across these 18 participants to determine the effects of the Friends on the Block literacy intervention on word recognition. Our research question is: When provided to elementary-aged children with low IQs, including students with IDD, what is the overall effect of Friends on the Block (literacy intervention) on (a) sight word reading, and (b) decodable word reading across the 18 SCD studies?

Method

The research design was a within-case multiple baseline across level single-case design (SCD) study. After establishing a stable baseline, we placed students into an initial level commensurate with their word reading ability. Then, as the student mastered words taught at that level, they moved into the next level. The literature on estimation and interpretation of single-case effect sizes makes clear the need to consider (a) experimental design, (b) baseline trend and auto-correlation of measures, (c) estimation of effect, and (d) aggregation of effects both within and across cases (Kratochwill et al., 2010). In particular, researchers have been encouraged to expand the use of quantitative data analytic methods, including meta-analytic procedures which are used in reviews of literature, in order to report the aggregate of effects for SCD (e.g., Shadish, 2014). The purpose of this research was to ascertain the promise of a single intervention, Friends on the Block, summarizing the findings of a series of replicated studies using meta-analytic statistical techniques. Of the 18 SCD studies included here, 8 were reported previously (Allor et al., 2018).

Setting and Participants

Following Institutional Review Board approval, 18 student participants were recruited for the study from six schools (10 classrooms); two schools were private schools for students with special needs, in a large metropolitan city in the Southwest. Students were recruited who met the following criteria: (a) IQ scores ranging from 40 to 79; (b) in grades one to four; (c) verbal (used spoken English language as primary means of communication) and (d) limited literacy skills. The sample for the present study included eight students from our previously published study and 10 additional students. Participants received the intervention across one to two academic years. Fourteen students received the intervention across one academic year; four students received the intervention across two academic years, 12 weeks in the spring of one year and 21 weeks in the next academic year. The students, five of whom were girls, ranged in age from 6 to 12 at the start of intervention. IQs ranged from 40–70 for all students except one student who had an overall IQ of 88, with a verbal score of 61 and percent reasoning score of 119. We report classification information provided by the students’ schools (see Table 1).

Friends on the Block Literacy Intervention

The intervention includes a set of 58 books (i.e., early readers) arranged in 14 levels and designed to provide practice on target sight words and decodable words (i.e., decodable words begin in later levels). The books focus on a set of engaging characters, including students with disabilities, and their families and pets, as well as non-fiction books thematically linked to the stories. (See examples at friendsontheblock.com; also see Allor et al., 2018). For each book, there are Instructional and Reinforcement Lessons that provide detailed guidance on teaching the specific skills needed to read each book. Instructional Lessons include (1) warm-up activities (e.g., phonological awareness, letter-sound correspondence), (2) book reading/discussion, and (3) learning games (e.g., skill practice). Reinforcement Lessons, implemented by paraprofessionals or other adults, include book reading and games. Students continue reading the books and participating in the accompanying lessons for one level until the words for that level are mastered. The mastery assessment is a proximal measure of the target words, the dependent variable described below. Other skills (e.g., letter sounds, phonemic awareness, oral language) gradually increase in difficulty across the levels and extensive cumulative review is provided. Activities are linked thematically to books. Books also provide cumulative review.

In the early level books, students read only sight words they have been taught (e.g., Level
I, like, do, not, want, a and a small number of words with pictures beneath them for scaffolding (e.g., I like pizza; a picture of pizza is beneath the word pizza). Sentences are brief, but increase in length and complexity in higher levels; words are selected that are high frequency and likely to be part of students’ oral lexicons. Phonemic awareness and decoding practice is provided in the warm-up activities and learning games, but students are not expected to apply these skills to text during the early levels. Text in later levels includes both sight words and high frequency decodable words (e.g., CVC words such as am and run). In the highest levels, advanced vowel patterns are added, as well as more sight words. The topics included in the books were selected to be relevant and interesting to students and included common settings and themes (e.g., playground, food), and social stories (e.g., sharing, what to do if lost in a store).

**Intervention Implementation**

Prior to the intervention each year, teachers participated in two hours of professional development learning how to implement the curriculum. They practiced blending words, segmenting words, pronouncing letter sounds, and providing specific corrective feedback using an I do it, we do it, you do it approach. Students were recommended by the teachers, and project staff conducted a placement test that contained a subset of six words for each level (higher levels contained 12 words, six sight words and six decodable words). Students were placed in the first level for which they read less than 4/6 words correctly. If students did not respond in 2 seconds or were overtly off task (e.g., looking around the room), they were prompted with “What word?” Students were then given approximately 5 seconds to say the word. The test was discontinued after six con-

---

**TABLE 1**

Case Demographics and Dosage

<table>
<thead>
<tr>
<th>Case</th>
<th>Age</th>
<th>IQ</th>
<th>Starting Level</th>
<th>Ending Level</th>
<th>PPVT Age Eq</th>
<th>Disability</th>
<th>Total Dosage (hours)</th>
<th>Total Dosage (weeks)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
<td>6</td>
<td>59</td>
<td>1</td>
<td>3</td>
<td>2:04</td>
<td>Smith-Lemli Opitz</td>
<td>31.96</td>
<td>27 weeks</td>
</tr>
<tr>
<td>Brent</td>
<td>7</td>
<td>55</td>
<td>1</td>
<td>3</td>
<td>3:04</td>
<td>Autism Spectrum Disorder/Speech Impairment</td>
<td>20.50</td>
<td>21 weeks</td>
</tr>
<tr>
<td>Claire</td>
<td>9</td>
<td>41</td>
<td>3</td>
<td>5</td>
<td>2:07</td>
<td>Intellectual Disability/Speech Impairment</td>
<td>38.50</td>
<td>21 weeks</td>
</tr>
<tr>
<td>Dan</td>
<td>9</td>
<td>42</td>
<td>6</td>
<td>9</td>
<td>2:06</td>
<td>Autism Spectrum Disorder</td>
<td>26.48</td>
<td>27 weeks</td>
</tr>
<tr>
<td>Ellie</td>
<td>9</td>
<td>42</td>
<td>1</td>
<td>3</td>
<td>2:09</td>
<td>Down syndrome</td>
<td>30.46</td>
<td>27 weeks</td>
</tr>
<tr>
<td>Frank</td>
<td>9</td>
<td>70</td>
<td>1</td>
<td>3</td>
<td>5:05</td>
<td>Autism Spectrum Disorder/Intellectual Disability/Speech Impairment</td>
<td>32.10</td>
<td>21 weeks</td>
</tr>
<tr>
<td>Greg</td>
<td>10</td>
<td>40</td>
<td>1</td>
<td>3</td>
<td>2:05</td>
<td>Smith-Lemli Opitz/Brain Tumor</td>
<td>30.66</td>
<td>27 weeks</td>
</tr>
<tr>
<td>Henry</td>
<td>10</td>
<td>43</td>
<td>1</td>
<td>7</td>
<td>5:04</td>
<td>Intellectual Disability/Speech Impairment</td>
<td>57.45</td>
<td>12 weeks/21 weeks</td>
</tr>
<tr>
<td>Isabelle</td>
<td>10</td>
<td>47</td>
<td>6</td>
<td>7</td>
<td>3:06</td>
<td>Down syndrome</td>
<td>57.08</td>
<td>27 weeks</td>
</tr>
<tr>
<td>John</td>
<td>10</td>
<td>56</td>
<td>1</td>
<td>3</td>
<td>3:04</td>
<td>Down syndrome</td>
<td>50.15</td>
<td>27 weeks</td>
</tr>
<tr>
<td>Kevin</td>
<td>10</td>
<td>88</td>
<td>7</td>
<td>11</td>
<td>7:05</td>
<td>Speech Impairment</td>
<td>32.10</td>
<td>12 weeks/21 weeks</td>
</tr>
<tr>
<td>Larry</td>
<td>11</td>
<td>52</td>
<td>1</td>
<td>6</td>
<td>6:01</td>
<td>Intellectual Disability/Speech Impairment</td>
<td>32.50</td>
<td>21 weeks</td>
</tr>
<tr>
<td>Matt</td>
<td>11</td>
<td>63</td>
<td>5</td>
<td>7</td>
<td>4:09</td>
<td>Down syndrome</td>
<td>43.88</td>
<td>27 weeks</td>
</tr>
<tr>
<td>Neal</td>
<td>11</td>
<td>65</td>
<td>1</td>
<td>6</td>
<td>9:05</td>
<td>Speech Impairment</td>
<td>35.47</td>
<td>12 weeks/21 weeks</td>
</tr>
<tr>
<td>Oscar</td>
<td>11</td>
<td>70</td>
<td>1</td>
<td>4</td>
<td>5:11</td>
<td>Intellectual Disability/Speech Impairment</td>
<td>53.20</td>
<td>21 weeks</td>
</tr>
<tr>
<td>Pam</td>
<td>12</td>
<td>63</td>
<td>7</td>
<td>9</td>
<td>7:06</td>
<td>Intellectual Disability/Speech Impairment</td>
<td>49.75</td>
<td>21 weeks</td>
</tr>
<tr>
<td>Quinn</td>
<td>12</td>
<td>65</td>
<td>8</td>
<td>14</td>
<td>5:04</td>
<td>Intellectual Disability/Speech Impairment</td>
<td>40.58</td>
<td>12 weeks/21 weeks</td>
</tr>
<tr>
<td>Robert</td>
<td>13</td>
<td>40</td>
<td>1</td>
<td>3</td>
<td>4:07</td>
<td>Down syndrome</td>
<td>54.55</td>
<td>27 weeks</td>
</tr>
</tbody>
</table>

IQ scores were based on Full Scale 4 scores on the Wechsler Abbreviated Scale of Intelligence WASI-II; (Wechsler, 2011).

Kevin had a verbal score of 61 and a percent reasoning score of 119.
secutive errors. The intervention was designed to be implemented daily in two 20-minute sessions: one Instructional and one Reinforcement Lesson.

**Measures and Data Collection Procedures**

*Target word measures and data collection.* The dependent variable (word reading) was measured with a proximal test of word reading. Through either a Power Point or flashcards, depending on student response and teacher recommendation (Jones et al., 2019), students were presented with a word test including 6 testing words from their current level and 6 testing words from the subsequent level. Therefore, as we collected intervention data for the current level we also collected baseline data for the next/subsequent level. Once they correctly read at least 4/6 testing words in their current level over two testing sessions, students were moved to the next level. Inter-observer agreement was conducted for 20% of these testing sessions and calculated to be 96.97% (SD = 5.31).

*Fidelity of implementation.* Implementation was measured with a researcher created fidelity observation instrument. Teachers were videoed once per instructional level per student and each activity was coded on a scale of 1 to 3 (3 = excellent implementation, 2 = adequate implementation, and 1 = weak implementation). To reach 100% fidelity teachers needed to obtain a 3 on each intervention component as well as teacher quality characteristics. As needed, coaches met with teachers to clarify any procedures. Overall fidelity was 90.45% (SD = 12.16). The total number of hours of intervention for each student is listed on Table 1; average dosage was 40.26 hours (SD = 11.22). A master coder reviewed 20% of the videos for Inter-observer agreement; this was 97.16% (SD = 4.35).

**Data Analytic Procedures**

Across 18 cases (i.e., student participants) for sight words and 8 cases for decodable words, our multiple baseline across levels design resulted in 67 and 27 studies overall, respectively. Testing for an intervention effect, each study analytically requires a comparison of baseline phase and intervention phase per level.

*Baseline corrected Tau.* The data obtained from each study was used to estimate the aggregated intervention effect. Coefficient Tau is a rank-order correlation between the phase shift and the repeated measure of word reading, and is considered an acceptable single-case experiment effect size when properly adapted as a percent non-overlapping coefficient (Parker et al., 2011). Tarlow (2017) describes short-comings of the conventional effect size coefficient $\text{Tau-U}$ and offers improvements computationally. We capitalized on the recommendations and used the estimate ‘baseline-corrected Tau’ ($\text{Tau}_{bc}$) for several reasons. First, it complements conventional visual and baseline-to-intervention non-overlapping data analyses. Second, as described by Tarlow (2017), the coefficient has properties that render it easily interpretable, e.g., 0 is a null effect, it is bounded by −1 and +1. Third, our criteria for implementing the intervention does not require a stable baseline and therefore adjustment for a baseline trend is necessary. As a nonoverlap index, the coefficient $\text{Tau}_{bc}$ is interpreted as the proportion of nonoverlapping data taking into consideration phase nonoverlap and intervention phase trend after controlling for the baseline trend (Parker et al., 2011). As a correlation, $\text{Tau}_{bc}$ is easily interpreted as a standardized effect size and is suitable for meta-analysis of effects obtained across replicated studies in designs such as ours.

*Combined effects.* In the context of single-case research, one way to combine the effects obtained from the replications and test hypotheses is through meta-analytic statistical procedures. Although meta-analysis is most often used to synthesize statistical findings reported in studies found through a systematic review of the literature on a specific topic, that is not our intent here. Rather, we use these procedures to examine the combined effects of a series of 18 independent SCD studies we conducted, each using the same intervention and research design. By design, specifically our criteria for shifting from baseline to intervention and advancing to the next instructional level, the studies varied in length. In our research, the effect size $\text{Tau}_{bc}$ is weighted by the inverse of the variance ($SE\ \text{Tau}_{bc}$).
squared), which gives greater weight to more reliable studies. Note, reliability of the effect size tends to be related directly to the number of observations, particularly the number of baseline observations. Studies with more observations, particularly during baseline, tend to produce more reliable estimates (Tarlow, 2017). Multi-level modeling was used to analyze results obtained from the replicated studies (Beretvas & Chung, 2008; Noortgate & Onghena, 2007). Across 18 cases, our within case multiple baseline design resulted in 67 and 27 studies overall for sight words and decodable words, respectively.

**Results**

The effect size $\tau_{se}$, as described above, was estimated for each study (Table 2). Our first step was to meta-analytically combine and analyze effects obtained across replicated studies within each case (i.e., student). For each case, our results indicate consistent moderate-to-strong positive effects for both sight words and decodable words. Table 2 provides a summary of the sight word and decodable word studies per case. The mean effect size $\tau_{se}$ across the 18 sight word cases ranged from 0.44 to 0.86, with an overall mean $\tau_{se}$ of 0.73. The estimated $\tau_{se}$ across the 8 decodable word cases ranged from 0.47 to 0.90, with a mean of 0.74. Using the standard error of the estimated $\tau_{se}$ for each case the 95% confidence interval does not include the value 0, supporting the hypothesis of a positive intervention effect. The homogeneity statistic $Q$ reported in Table 2 is a test of the hypothesis that the variability of the effect sizes within a case varies randomly. For sight words and decodable

<table>
<thead>
<tr>
<th>Case</th>
<th>N Studies</th>
<th>$Q^a$</th>
<th>Mean $\tau_{se}$</th>
<th>Lower</th>
<th>Upper</th>
<th>N Studies</th>
<th>$Q^a$</th>
<th>Mean $\tau_{se}$</th>
<th>Lower</th>
<th>Upper</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ann</td>
<td>3</td>
<td>1.29</td>
<td>0.65**</td>
<td>0.34</td>
<td>0.96</td>
<td>3</td>
<td>2.25</td>
<td>0.47**</td>
<td>0.11</td>
<td>0.83</td>
</tr>
<tr>
<td>Brent</td>
<td>2</td>
<td>0.01</td>
<td>0.82**</td>
<td>0.52</td>
<td>1.12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Claire</td>
<td>3</td>
<td>0.97</td>
<td>0.68**</td>
<td>0.39</td>
<td>0.98</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dan</td>
<td>4</td>
<td>1.38</td>
<td>0.52**</td>
<td>0.12</td>
<td>0.84</td>
<td>3</td>
<td>2.25</td>
<td>0.47**</td>
<td>0.11</td>
<td>0.83</td>
</tr>
<tr>
<td>Ellie</td>
<td>3</td>
<td>0.24</td>
<td>0.73**</td>
<td>0.47</td>
<td>0.99</td>
<td>2</td>
<td>0.15</td>
<td>0.70**</td>
<td>0.37</td>
<td>1.04</td>
</tr>
<tr>
<td>Frank</td>
<td>2</td>
<td>0.02</td>
<td>0.85**</td>
<td>0.58</td>
<td>1.13</td>
<td>5</td>
<td>6.04</td>
<td>0.71**</td>
<td>0.49</td>
<td>0.92</td>
</tr>
<tr>
<td>Greg</td>
<td>4</td>
<td>2.41</td>
<td>0.62**</td>
<td>0.34</td>
<td>0.90</td>
<td>2</td>
<td>0.15</td>
<td>0.70**</td>
<td>0.37</td>
<td>1.04</td>
</tr>
<tr>
<td>Henry</td>
<td>5</td>
<td>0.61</td>
<td>0.86**</td>
<td>0.68</td>
<td>1.04</td>
<td>6</td>
<td>2.22</td>
<td>0.90**</td>
<td>0.76</td>
<td>1.05</td>
</tr>
<tr>
<td>Isabelle</td>
<td>3</td>
<td>0.29</td>
<td>0.71**</td>
<td>0.45</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>John</td>
<td>4</td>
<td>0.27</td>
<td>0.77**</td>
<td>0.54</td>
<td>1.01</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Kevin</td>
<td>4</td>
<td>1.10</td>
<td>0.78**</td>
<td>0.55</td>
<td>1.00</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Larry</td>
<td>4</td>
<td>0.09</td>
<td>0.79**</td>
<td>0.54</td>
<td>1.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Matt</td>
<td>4</td>
<td>2.63</td>
<td>0.43**</td>
<td>0.11</td>
<td>0.75</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neal</td>
<td>6</td>
<td>0.28</td>
<td>0.71**</td>
<td>0.49</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Oscar</td>
<td>4</td>
<td>1.78</td>
<td>0.83**</td>
<td>0.61</td>
<td>1.04</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Pam</td>
<td>3</td>
<td>1.20</td>
<td>0.45**</td>
<td>0.10</td>
<td>0.79</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Quinn</td>
<td>6</td>
<td>3.24</td>
<td>0.78**</td>
<td>0.60</td>
<td>0.97</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Robert</td>
<td>3</td>
<td>1.21</td>
<td>0.65**</td>
<td>0.37</td>
<td>0.93</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

*Statistically significant $Q$ indicates heterogeneity of the effect sizes across the N studies. For each case the variability of $\tau_{se}$ was not significant statistically. 

*a $p \leq 0.05$, ** $p \leq 0.01$.

**TABLE 2**

Effect Size Tau Estimates by Case
words, \( Q \) values were non-significant, indicating that within case effect sizes varied randomly. Using conventional standards for evaluating effect sizes (Lipsey & Wilson, 2001), the \( Tau_{se} \) reported in Table 2 indicate moderate to strong effects of the intervention. (See Figures 1–3 for sample graphs; see friendsontheblock.com for all graphs.)

The next step was a meta-analysis of the \( Tau_{se} \) values across all studies within and across cases. The range of \( Tau_{se} \) across the 67 sight word studies was 0.03 to 0.94. Across the 27 decodable word studies, the range of \( Tau_{se} \) was 0.02 to 0.89. The range, which included a few low values, indicates some variability in response to the intervention. Yet, the mean effect per case was consistently significant statistically, supporting the hypothesized positive effect (see Table 2). Using random effects multilevel modeling and weighting each study by the reliability of the effect size estimate, the mean estimated \( Tau_{se} \) for sight words and decodable words was 0.70 and 0.66, respectively. These values indicate that approximately 70% and 66% of the measures during intervention were higher than the measures obtained during baseline. As a correlation coefficient, the values of \( Tau_{se} \) indicate strong positive intervention effects.

Discussion

Our purpose was to examine the effects of a comprehensive literacy intervention, *Friends on the Block*, on word recognition across three years of SCD studies with 18 participants with IDD (ages 6 to 13). We examined the data from 18 cases (i.e., students) to determine the effects for sight words and in 8 of the 18 cases for decodable words. For each case, we used a multiple-baseline design across levels of the intervention. We used quantitative statistical methods, including multi-level modeling, to analyze the data across cases.

Summary of Findings

The answer to our research question is that there is a clear functional relation between
the intervention and increases for both sight words and decodable words. Within each case the effect sizes varied. Results indicate consistent moderate-to-strong positive effects for each case and across all cases for both sight words and decodable words using conventional standards for evaluating effect sizes (Lipsey & Wilson, 2001). Effects were clearer for sight words than decodable words, but this is likely because we conducted more individual studies of sight words than decodable words. We interpret these findings within the context of the amount of growth (see Table 1), not simply in terms of the differences between baseline scores and intervention scores.

**Individual Differences**

To illustrate individual differences in response we provide brief descriptions of two students, Pam and Henry, and their progress (see Figures 1–3). Pam, a 12-year-old with an IQ of 63, is a student for whom the intervention had relatively weaker effects according to our analyses. Though she began the study with some sight word knowledge, during the study she increased her sight word knowledge and her ability to read decodable short vowel words and long vowel words with a silent-e pattern (e.g., cake). Yet, our analyses indicated a somewhat inconsistent pattern; she had a moderate overall effect size of .45 on sight words and .53 on decodable words, but inconsistent effects across levels on sight words (.19, .63, and .28 for Levels 7, 8, and 9, respectively), and on decodable words across those same levels (.72, .17, and .63, respectively). This variability appears to be due to baseline scores indicating she knew some words prior to the intervention, thus, decreasing the percent of nonoverlapping data, and lowering the overall effect. In other words, she knew some of the words taught in a level before instruction in that level began. Pam began the study at Level 7 reading 77 words correctly on the placement test given in October. By the end of April, she had mastered Level 9 and read 102 words correctly on the
same word list. Thus, across the 21 weeks of intervention (49.75 hours of instruction), she demonstrated mastery of words in three intervention levels. To put this in context, Levels 7–9 represent skills typically learned around the middle of first grade, yet Pam was 12 years old at the beginning of the study.

By contrast, Henry, a 10-year-old with an IQ of 43, is a student for whom the intervention had consistently stronger effects. He began the study with very limited literacy skills (e.g., only read six words correctly on the placement test; knew about 15 letter sounds; had very low phonemic awareness) and advanced through the first six levels of the curriculum, ultimately reading 48 words on our proximal measure of a selection of words taught in the intervention. He had both a high overall ES of .86 on sight words and .74 on decodable words. Unlike Pam, Henry had consistently strong effects across levels, ranging from .66 to .94. He began the study in February at Level 1, reading only 6 words correctly on the placement test. He progressed through Level 3, reading 19 words correctly, by the end of May (across 12 weeks). He continued the following year progressing from Level 4 to Level 6 and growing from 25 words to 48 words correct across the 21 weeks. Although the number of words Henry read correctly grew relatively fast compared to the other students in our sample, his ending reading performance was similar to a typical student in the early part of first grade; he was able to read both the taught sight words and some decodable short vowel words.

**Fidelity of Implementation**

Summative data on fidelity of implementation indicate high levels of implementation and we can conclude that the dosages reported were sufficient to produce positive and clear effects for the intervention. Across the 18 cases, the average dose was 3.03 Instructional Lessons (SD = .45) and 1.83 Reinforcement Lessons (SD = 1.02) per week. With this dosage, all...
students progressed through at least two and up to six levels of the curriculum within a school year (see Table 1). Additional methods for increasing dosage need to be examined in future research, such as providing additional small group size sessions, supporting families to provide practice, supporting student practice in centers and peer-pairs, and using technology, such as e-books within a listening center.

Limitations and Future Research Directions

It is clear from this analysis that there is a positive functional relation between the intervention and word reading, providing strong support for the promise of the intervention, but these findings should be interpreted cautiously for several reasons. First, by virtue of SCD, the analysis does not address whether the rate or amount of overall reading progress made by the students in the study was educationally or functionally meaningful, because the study focused solely on one proximal dependent variable (i.e., word recognition), even though the intervention was comprehensive and addressed multiple skills. This was necessary to demonstrate a functional relation between the intervention and the dependent variable, though it does not fully address the impact of the intervention which also targeted phonemic awareness, fluency, and comprehension skills. Second, we only tested a sample of the words taught in the intervention (e.g., 6 of the 15 decodable words per level). This was necessary due to the extensive testing required to meet What Works Clearinghouse quality indicators as it would have been unreasonable to test all words repeatedly (Kratochwill et al., 2010). Although we collected data on multiple measures (see Allor et al., 2018), these are descriptive and not causal.

A second limitation is that we had fewer demonstrations of effects with decodable words than sight words and for the upper levels of the intervention as opposed to the lower levels. Finally, our findings are limited to English-speaking students who are elementary school age and who used spoken language as their primary means of communication (i.e., none were non-verbal, though all had very low language skills; see Table 1 for PPVT scores). A next step is to adapt the curriculum for students with more complex communication needs.

Future research involving a larger scale randomized-control trial (RCT) is needed to examine the impact of the intervention on multiple variables, particularly comprehension and engagement, and to further examine potential moderators, such as age, disability type, IQ and language skills. A deeper understanding of potential moderators has the potential to improve our understanding of how to adapt instruction to address individual differences. Although teachers reported increases in student engagement and comprehension, these outcomes could be examined empirically with a RCT. If the length of the intervention were increased, growth in fluency could also be examined. Research is also needed to examine the components and characteristics of the program, as well as exploring additional ways to meet individual needs and capitalize on student interests. For example, Lemons and colleagues have explored promising ways to adapt literacy instruction for students with DS, such as increasing use of mnemonic words and images to support working memory and phonological awareness (Lemons et al., 2015). Although we selected topics likely to be relevant and motivating, motivation may be further increased if students were provided more choices or if books related to specific individual interests.

Summary and Instructional Implications

In summary, this study builds on our previous research and the research of others that supports comprehensive literacy instruction for students with IDD, including students with IQs as low as 40 (Allor et al., 2014; Browder et al., 2012; Connor et al., 2014; Lemons et al., 2017). It provides support for raised expectations for reading growth for students with IDD, which is relevant to Endrew vs. Douglas County (2017). The study provides preliminary evidence for the use of a text-centered literacy curriculum for students with IDD. We also recommend the use of discussion questions similar to those used in Dialogic Reading (see Whitehurst & Lonigan, 1998) to build oral language and comprehension. Our aim was to study the promise of the intervention for students with IDD and these findings dem-
onstrate strong initial support for the effectiveness of Friends on the Block in teaching early word recognition to students with IDD. This study extends the findings from our initial study, in which we reported on the effectiveness of the intervention through SCD with the first eight participants, as well as descriptive measures of student progress on a variety of academic skills, and the feasibility of the intervention (Allor et al., 2018). The study reported here used quantitative methods to combine those eight cases with an additional 10 cases, increasing the number of sight word and decodable word studies to 67 and 27, respectively, and providing strong support for a statistically significant functional relation between the intervention and word recognition. Given the importance of literacy in improving opportunities and outcomes for students with IDD (Cihak & Smith, 2018; Conners, 2003) and the pervasive challenges they experience in learning to read and write, it is important that we improve our understanding of how to best teach students with IDD.

References


Paraprofessional-Implemented Systematic Instruction for Students with Disabilities: A Systematic Literature Review

Virginia L. Walker
University of North Carolina at Charlotte

Karen H. Douglas
Radford University

Sarah N. Douglas
Michigan State University

Sophia R. D’Agostino
Hope College

Abstract: The purpose of this systematic literature review was to summarize single-case intervention studies involving paraprofessional-implemented systematic instruction for students with disabilities. In the 19 reviewed studies, 60 paraprofessionals received training to implement systematic instruction with most learning naturalistic language strategies, least-to-most prompting, pivotal response training, or discrete trial training. Researchers delivered paraprofessional training in a majority of cases using both didactic and experiential training methods. Paraprofessionals primarily taught students with autism spectrum disorder and focused on social/communication skills in a one-to-one instructional arrangement. The majority of studies demonstrated positive effects on paraprofessional implementation of systematic instruction and student outcomes. Implications for practice, limitations, and areas for future research are addressed.

Paraprofessionals work under the direction of certified teachers and other professionals to provide special education and related services to students with disabilities in school settings (Council for Exceptional Children [CEC], 2015; Individuals with Disabilities Act, 2004). The number of paraprofessionals supporting students in this capacity has grown over the past few decades, with federal data reporting over 488,000 paraprofessionals supporting students with disabilities in public schools (U.S. Department of Education, 2018) and more paraprofessionals serving students with disabilities than special education teachers in the U.S. (U.S. Department of Education, 2018). As the number of paraprofessionals has increased, the role of paraprofessionals also has shifted. In the past, paraprofessional supports were clerical in nature, but paraprofessionals now engage in instructional roles, including implementation of evidence-based practices (EBPs) under the supervision of teachers (Carter et al., 2009). In response to the increasing reliance on paraprofessionals, CEC developed a set of paraprofessional preparation standards outlining knowledge and skill competencies, including knowledge in “evidence-based practice” for students with disabilities (CEC, 2015, p. 144).

A number of challenges have been and continue to be identified concerning the implementation and utilization of paraprofessional supports. Although federal law clearly conveys a requirement for appropriately trained and supervised paraprofessionals (Every Student Succeeds Act, 2015), research overwhelmingly indicates that paraprofessionals are under prepared and under trained to work with students with disabilities (Douglas et al., 2016; Fisher & Pleasants, 2012; Giangreco et al., 2002). Further, teachers typically are not well prepared for their supervisory responsibilities with paraprofessionals (Biggs et al., 2019; Douglas et al., 2016). This lack of preparation and training often leads to delivery of supports that are not appropriate or effective for students with disabilities (Fisher & Pleasants, 2012). However, an emerging body of research suggests that, with adequate training and supervision, paraprofessionals can imple-
ment a wide range of EPBs for students with disabilities (Brock & Carter, 2013; Rispoli et al., 2011; Walker & Smith, 2015). Despite this knowledge, however, little is known about paraprofessional implementation of specific EPBs and the methods that are most effective in preparing paraprofessionals.

**Systematic Instruction**

Systematic instruction is an EBP that draws on the principles of applied behavior analysis and includes prompting students in a defined, consistent manner to perform an observable and measurable behavior and providing feedback contingent on student responses (Browder et al., 2014; Collins, 2012; Spooner et al., 2017). Several prompting systems have been effective with students with disabilities across domains related to academics, essential life skills, and communication/social skills (e.g., Browder et al., 2009; Browder et al., 2008; Dogoe & Banda, 2009; Hudson et al., 2018; Morse & Schuster, 2004). These include the system of least prompts, most-to-least prompting, time delay, simultaneous prompting, and graduated guidance (Brown et al., 2016; Collins, 2012). Each involves systematically delivering prompts to encourage student responding and fading instructor support once the student demonstrates targeted performance levels. Other instructional practices that rely on systematic prompting include naturalistic language strategies, pivotal response training (PRT), and discrete trial training (DTT; Odom et al., 2010; Wong et al., 2015). Table 1 provides a description of each of these approaches.

Systematic instruction is regarded as an EBP for students with extensive support needs (i.e., “severe disabilities”; Browder et al., 2014; Spooner et al., 2017). These students are the 1% of students with intellectual disability, autism spectrum disorder (ASD), or multiple disabilities who have extensive and pervasive needs across domains and who participate in alternate assessments due to these needs (Kurth et al., 2018). Professionals working with this population have identified systematic instruction as a critical skill area that is necessary to effectively educate students with extensive support needs (Ruppar et al., 2018). In addition, various prompting strategies, naturalistic language strategies, PRT, and DTT all have been identified as EPBs for students with ASD (McLeskey et al., 2017; Odom et al., 2010; Wong et al., 2015). It should be noted that, although these practices are established EPBs for students with ASD and extensive support needs, systematic instruction also can be effective for other learners and help support instruction across a wide range of skill domains (e.g., Britton et al., 2017; Head et al., 2011). Considering the promise that systematic instruction holds, it is imperative that all educators who support students with disabilities, including paraprofessionals, are trained to use systematic instruction. Because systematic instruction is often multifaceted and therefore requires adherence to specific procedures, paraprofessionals may need more support and training to ensure implementation fidelity (Brock & Carter, 2013; Walker et al., 2020). As such, the field of special education would benefit from a literature review aimed at gathering and interpreting data on the topic of paraprofessional-implemented systematic instruction.

**Study Purpose**

Emerging evidence suggests that paraprofessionals can implement EPBs for students with disabilities. However, there are no reviews to our knowledge that summarize the current status of paraprofessional-implemented systematic instruction. Other literature reviews have examined paraprofessional-implemented practices, but none have targeted specific EPBs, such as systematic instruction. For example, Brock and Carter (2013) found that paraprofessionals are capable of implementing a range of instructional practices for students with intellectual and developmental disabilities. However, the scope of their review extended beyond systematic instruction to address multiple strategies including social narratives, interventions involving augmentative and alternative communication, and visual supports. Two other paraprofessional-focused literature reviews examined intervention studies involving paraprofessional training to support students with any type of disability (Walker & Smith, 2015) and those with ASD (Rispoli et al., 2011). Both reviews also included a wide variety of instructional practices, without providing information specific to
Although these three literature reviews provide important information about paraprofessionals and their implementation of interventions, a focused literature review is necessary to understand the characteristics of participants and intervention conditions associated with paraprofessional-implemented systematic instruction, the paraprofessional fidelity of implementation, and the quality and effectiveness of studies. In addition, a literature review allows for identification of issues within existing literature to inform practice and future research directions (Cooper et al., 2009). As such, we conducted a literature review of intervention studies involving paraprofessional-implemented systematic instruction for students with disabilities, with a goal to inform best practice for supporting paraprofessionals in delivering systematic instruction. Our review addressed the following research questions: (1) What are the characteristics of paraprofessionals who have implemented systematic instruction and the participating students?; (2) What are the characteristics of paraprofessional-implemented systematic instruction interventions?; and (3) What is the quality of the research on paraprofessional-implemented systematic instruction?

**Method**

We conducted a systematic literature review to descriptively summarize intervention research.
involving paraprofessional-implemented systematic instruction for students with disabilities. Each phase of the review is described in greater detail in the sections that follow.

**Literature Search**

To identify studies to consider for inclusion in the review, we searched the literature in several ways (see Figure 1). First, we conducted a search of four online databases (PsycINFO, ERIC, Medline, ProQuest) in January 2017 using key search terms from two broad categories: (1) paraprofessional (e.g., paraeducator, teaching assistant, classroom aide) and (2) systematic instruction (e.g., time delay, system of least prompts, DTT). We limited the online database search to references available in English and considered both published and unpublished studies to avoid the threat of publication bias (Gage et al., 2017). Second, we conducted a database search focused on the paraprofessional literature (see database at https://www.uvm.edu/cess/cdci/selected-paraprofessional-references). Third, we conducted an electronic search of 21 peer-reviewed journals that commonly published research in special education, disability studies, and behavioral intervention. We searched for the aforementioned key terms in the title and abstract of each article across all volumes and issues. Fourth, we conducted a reference list search of existing research reviews pertaining to paraprofessionals and the reference lists of studies included in those reviews (i.e., Brock & Carter, 2013; Rispoli et al., 2011; Walker & Smith, 2015). After eliminating duplicates and references that were not empirical in nature (e.g., book chapters, position papers, conference proceedings), a total of 244 references remained to be considered for inclusion in the review. A complete list of search terms and detailed search procedures is available from the first author.

![Figure 1. This figure depicts the overall process for searching the literature, screening potentially relevant references, and determining the eligibility of articles for inclusion in the review.](image-url)
Study Selection

To determine whether the 244 identified studies qualified for inclusion in the review, we applied the following selection criteria to each study: (a) included one or more students with a disability; (b) involved paraprofessional-implemented systematic instruction with one or more dependent measures to assess paraprofessional implementation; and (c) utilized an experimental single-case research design to assess intervention effectiveness. If a study involved a multiple-component intervention, we included the study only if there was clear evidence that systematic prompting and feedback was part of the intervention package. We excluded studies if student participants did not have a disability, dependent measures for paraprofessional implementation were not described, or information about paraprofessional participants was missing such that coding at the paraprofessional level was not possible. Studies also were excluded when information related to paraprofessional participants was not available or when qualitative descriptions of participant characteristics or group designs were utilized, as we were interested in evaluations of systematic instruction at the participant level (Ledford & Gast, 2018).

Two doctoral students in special education reviewed abstracts of the identified studies and applied selection criteria to determine inclusion in the review. A total of 33 studies met inclusion criteria after this initial round of review. Next, two different reviewers, the first and second authors, both special education faculty, reviewed the full text of the 33 studies identified as having met the inclusion criteria through the abstract review. After this second round of review, 14 additional studies were excluded as the reviewers determined that the intervention did not reflect systematic instruction, leaving 19 studies meeting the inclusion criteria. To assess inter-rater reliability during the first round of review, the primary reviewer applied the selection criteria to all studies and the secondary reviewer evaluated 30% of the studies selected at random. In the second round of review, both reviewers independently applied the criteria to all studies. After independently completing reviews, agreement was calculated on a code-by-code basis by dividing the total number of agreed codes by the total number of codes and multiplied by 100. Agreement was 97% (range: 75–100%) for the first round of review (abstract review), and 100% for the second round of reviews (full text review).

Descriptive Coding

We reviewed included studies to gather descriptive information about the paraprofessional and student participants and interventions involving paraprofessional-implemented systematic instruction. Coding was completed for each qualifying paraprofessional participant, as paraprofessionals were treated as the unit of analysis for the purposes of this review. Coding categories and individual coding items and procedures are discussed next. The coding form used in this review is available from the first author.

Coding instrument. Based on content from previous paraprofessional reviews (i.e., Brock & Carter, 2013; Rispoli et al., 2011; Walker & Smith, 2015) and literature on systematic instruction (i.e., Brown et al., 2016; Collins, 2012), the coding instrument was comprised of five general descriptive coding categories: (1) paraprofessional participants, (2) student participants, (3) paraprofessional training, (4) paraprofessional-implemented systematic instruction, and (5) quality measures. Coding items for paraprofessional participants included gender, age, educational level, time as a paraprofessional or in related role, race/ethnicity, and prior training in systematic instruction. Student participant coding items focused on gender, age, school level, race/ethnicity, disability diagnosis, and previous intervention for targeted skills. The coding instrument included the following coding items pertaining to the training paraprofessionals received to implement systematic instruction: trainer role, training format, training duration, implementation fidelity, research design to measure effects of training, and paraprofessional behavior outcomes. Coding items for paraprofessional-implemented systematic instruction were target skills, setting, instructional arrangement, other implementers, prompting strategies (see Table 1), schedule of reinforcement, plans for thinning reinforcement, intervention frequency, research design for effects of systematic instruc-
tion on student behavior, and student outcomes. For paraprofessional and student outcomes, we visually analyzed graphs of outcome measures using guidelines set forth by Ledford and Gast (2018) to determine if the intervention had a positive effect, mixed effect, or no effect. Finally, we extracted information about the following four quality characteristics in single-case research studies: dependent measure reliability, generalization, maintenance, and social validity (Horner et al., 2005).

Coding procedure. We used the coding instrument to extract information across each qualifying paraprofessional participant. When information was not reported or unclear, reviewers selected “cannot determine”. We contacted authors for clarification when prompting strategies were not clearly described. A total of six authors confirmed or clarified the prompting strategy. In some cases, multiple codes were selected under a coding item due to multiple applicable codes. The individuals who previously applied the inclusion criteria independently reviewed and coded the 19 studies. After independently completing reviews, they met to discuss any discrepancies and agreed upon codes. To assess inter-rater reliability, secondary reviewers evaluated 30% of the studies selected at random. We calculated agreement on a code-by-code basis, dividing the total number of agreed upon codes by the total number of codes and multiplied by 100. Overall agreement was 95% (range: 88–100%).

Data Analysis
To analyze coding items containing closed-ended response options, we calculated the number and percentage of instances the response option was selected. Some coding items allowed for multiple codes to be selected. For coding items containing opened-ended response options (e.g., age, time as a paraprofessional, duration of intervention), we calculated a range.

Results
Across the 16 journal articles and three dissertations that met the inclusion criteria, a total of 60 paraprofessionals received training to deliver systematic instruction to students with disabilities. Overall, paraprofessional training resulted in positive outcomes in relation to paraprofessional implementation of systematic instruction. Of the 13 studies that included student outcomes, positive outcomes were reported for 85% of students. Only one study failed to demonstrate a positive effect for student participants (Ryan et al., 2008). Results are described next, with a summary of studies provided in Table 2. It should be noted that, in some cases, articles did not provide sufficient information to permit coding for specific coding categories; as such, the percentages below do not always total to 100%.

Paraprofessional and Student Participants
Table 3 contains detailed information about paraprofessional characteristics related to gender, race/ethnicity, educational level, and training related to systematic instruction. A majority of paraprofessionals were female, with ages ranging between 18 and 71 years across the 39 paraprofessionals for whom information about age was reported. When educational level was reported, paraprofessionals typically had either a four-year college degree or some experience in college without having earned a degree. The amount of time as a paraprofessional or in a related role ranged from less than one year to 17 years; time as a paraprofessional was not reported for 22 paraprofessionals. In over half of the cases, paraprofessional race/ethnicity was not described nor was their prior training in systematic instruction.

Information about student characteristics related to gender, race/ethnicity, school level, and reported disability diagnosis is presented in Table 4. The 60 paraprofessionals delivered systematic instruction to over 47 students, with 13 paraprofessionals providing instruction to more than one student. Some authors did not report the number of students in their study (i.e., Hall et al., 2010; Slider, 2004). The majority of student participants were male, with ages ranging between 16 months and 16 years across the 45 students for whom age was reported. Most students attended either an early childhood or elementary program. In 70% of cases, student race/ethnicity could not be identified. Students primarily had disability.
<table>
<thead>
<tr>
<th>Study</th>
<th>n*</th>
<th>Student Disability</th>
<th>School Level</th>
<th>Training Type</th>
<th>Prompting Strategy</th>
<th>Paraprofessional Outcomes (n)</th>
<th>Student Outcomes (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bingham, Spooner, and Browder (2007)</td>
<td>3</td>
<td>ID; MD</td>
<td>Elementary; middle; high</td>
<td>Didactic</td>
<td>LMP</td>
<td>Positive (3)</td>
<td>Positive (3)</td>
</tr>
<tr>
<td>Bolton and Mayer (2008)</td>
<td>3</td>
<td>ASD</td>
<td>Early childhood; middle</td>
<td>Didactic</td>
<td>DTT</td>
<td>Positive (3)</td>
<td>NA</td>
</tr>
<tr>
<td>Britton, Collins, Jones, and Bausch (2017)</td>
<td>1</td>
<td>ID</td>
<td>High</td>
<td>Experiential</td>
<td>SP</td>
<td>Positive (1)</td>
<td>Positive (1)</td>
</tr>
<tr>
<td>Dib and Sturmey (2007)</td>
<td>3</td>
<td>ASD</td>
<td>CD</td>
<td>Experiential</td>
<td>DTT</td>
<td>Positive (3)</td>
<td>Positive (3)</td>
</tr>
<tr>
<td>Douglas, Light, and McNaughton (2013)</td>
<td>3</td>
<td>SI; DD; other</td>
<td>Early childhood</td>
<td>Didactic</td>
<td>NL</td>
<td>Positive (3)</td>
<td>Positive (3)</td>
</tr>
<tr>
<td>Hall, Grunton, Pope, and Romero (2010)</td>
<td>5</td>
<td>ASD</td>
<td>Early childhood</td>
<td>Didactic &amp; experiential</td>
<td>LMP; NL; PRT; CD</td>
<td>Positive (3); mixed (2)</td>
<td>NA</td>
</tr>
<tr>
<td>Kim, Koegel, and Koegel (2017)</td>
<td>3</td>
<td>ASD</td>
<td>Elementary</td>
<td>Didactic &amp; experiential</td>
<td>NL; PRT</td>
<td>Positive (3)</td>
<td>Positive (3)</td>
</tr>
<tr>
<td>Klein and Matos (2012)</td>
<td>3</td>
<td>ASD</td>
<td>Elementary</td>
<td>Didactic &amp; experiential</td>
<td>PRT</td>
<td>Positive (3)</td>
<td>Positive (3)</td>
</tr>
<tr>
<td>Leblanc, Ricciardi, and Luiselli (2005)</td>
<td>3</td>
<td>ASD</td>
<td>CD</td>
<td>Didactic &amp; experiential</td>
<td>CD</td>
<td>Positive (3)</td>
<td>NA</td>
</tr>
<tr>
<td>Martella, Marchand-Martella, Macfarlane, and Young (1993)</td>
<td>1</td>
<td>ID; other</td>
<td>High</td>
<td>Didactic &amp; experiential</td>
<td>TD</td>
<td>Positive (1)</td>
<td>Positive (1)</td>
</tr>
<tr>
<td>McCulloch and Noonan (2013)</td>
<td>3</td>
<td>ASD; DD</td>
<td>Elementary</td>
<td>Didactic</td>
<td>NL</td>
<td>Positive (2); mixed (1)</td>
<td>Positive (3)</td>
</tr>
<tr>
<td>O’Guin (2010)</td>
<td>5</td>
<td>ASD; OI; SLI</td>
<td>Middle</td>
<td>Didactic &amp; experiential</td>
<td>DTT; PRT</td>
<td>Positive (5); Positive (4)</td>
<td>Positive (5); Positive (4)</td>
</tr>
<tr>
<td>Robinson (2011)</td>
<td>4</td>
<td>ASD</td>
<td>Early childhood; elementary</td>
<td>Experiential</td>
<td>PRT</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Continued on next page)
TABLE 2

(Continued)

<table>
<thead>
<tr>
<th>Study</th>
<th>n*</th>
<th>Student Disability</th>
<th>School Level</th>
<th>Training Type</th>
<th>Prompting Strategy</th>
<th>Paraprofessional Outcomes (n)</th>
<th>Student Outcomes (n)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ryan, Hemmes, Sturmey, Jacobs, and Grommet (2008)</td>
<td>6</td>
<td>ASD</td>
<td>CD</td>
<td>Didactic</td>
<td>NL</td>
<td>Positive (5); no effect (1)</td>
<td>No effect (6)</td>
</tr>
<tr>
<td>Schepis, Ownbey, Parsons, and Reid (2000)</td>
<td>2</td>
<td>DD; other</td>
<td>Early childhood</td>
<td>Didactic &amp; experiential</td>
<td>LMP</td>
<td>Positive (2)</td>
<td>NA</td>
</tr>
<tr>
<td>Schepis, Ownbey, Parsons, and Reid (2001)</td>
<td>4</td>
<td>ASD; ID; other</td>
<td>Early childhood</td>
<td>Didactic &amp; experiential</td>
<td>LMP</td>
<td>Positive (4)</td>
<td>Positive (5)</td>
</tr>
<tr>
<td>Slider (2004)</td>
<td>4</td>
<td>SLI</td>
<td>Early childhood</td>
<td>Didactic &amp; experiential</td>
<td>TD</td>
<td>Positive (3); no effect (1)</td>
<td>NA</td>
</tr>
<tr>
<td>Walker and Snell (2017)</td>
<td>1</td>
<td>ASD; SLI</td>
<td>Elementary</td>
<td>Didactic &amp; experiential</td>
<td>LMP</td>
<td>Positive (1)</td>
<td>Positive (1)</td>
</tr>
<tr>
<td>Zanton (2015)</td>
<td>3</td>
<td>Other</td>
<td>Early childhood</td>
<td>Didactic &amp; experiential</td>
<td>MLP</td>
<td>Positive (3)</td>
<td>NA</td>
</tr>
</tbody>
</table>

Note. ID = intellectual disability; MD = multiple disabilities; ASD = autism spectrum disorder; SI = sensory impairment; DD = developmental delay; other = specific genetic disorder, cerebral palsy, seizure disorder; OH = orthopedic impairment; SLI = speech and language impairment; CD = cannot determine; LMP = least-to-most prompting; DTT = discrete trial training; SP = simultaneous prompting; NL = naturalistic language; PRT = pivotal response training; MLP = most-to-least prompting; TD = time delay; NA = not applicable

*nRefers to the number of qualifying student participants.
diagnoses of ASD and/or speech language impairment. There were no reports of students receiving previous instruction on targeted skills, though these data were provided for only 10% of students.

**Paraprofessional Training**

All of the paraprofessionals received training on systematic instruction or an intervention package containing systematic instruction. In a majority of cases (77%), a researcher trained paraprofessionals, with fewer paraprofessionals trained by special education teachers (17%). Five of the teachers delivered training in collaboration with the researcher and four with a program director. Only one paraprofessional received all training from a special education teacher (Britton et al., 2017), and one study with three paraprofessionals did not identify the trainer (Bolton & Mayer, 2008). Over half of the trainings (57%) included both didactic and experiential formats. Thirty percent of the trainings were didactic only and 13% were experiential only. When reported, the duration of training sessions varied from less than one day (one 20 min session) up to 10 days. Researchers assessed implementation fidelity for all of paraprofessionals using a multiple baseline or probe design to measure training effects.

**Paraprofessional-Implemented Systematic Instruction**

Paraprofessionals used systematic instruction to teach the following skills to student participants: social/communication skills (63%), challenging behavior (20%), academic skills (3%), and essential life skills (2%). Paraprofessionals provided instruction in both inclusive settings (52%) and non-inclusive settings (48%). Furthermore, over half of the paraprofessionals (53%) taught the students in a one-to-one instructional arrangement with only 7% providing instruction in a whole group arrangement; instructional arrangement

| TABLE 3 | Paraprofessional Characteristics |
|---|---|---|
| Characteristic | Number | Percentage |
| Gender | | |
| Female | 53 | 88 |
| Male | 4 | 7 |
| Cannot determine | 3 | 5 |
| Race/ethnicity | | |
| Cannot determine | 38 | 63 |
| White | 8 | 13 |
| Hispanic/Latino | 5 | 8 |
| Black | 4 | 7 |
| Native American/Pacific Islander | 3 | 5 |
| Asian | 1 | 2 |
| Other | 1 | 2 |
| Educational level | | |
| Cannot determine | 17 | 29 |
| Four-year college degree | 14 | 23 |
| Some college, no degree | 14 | 23 |
| High school degree | 12 | 20 |
| Two-year college degree | 3 | 5 |
| Training in systematic instruction | | |
| No prior training | 34 | 57 |
| Cannot determine | 14 | 23 |
| Prior training | 12 | 20 |

| TABLE 4 | Student Characteristics |
|---|---|---|
| Characteristic | Number | Percentage |
| Gender | | |
| Male | 41 | 65 |
| Cannot determine | 13 | 21 |
| Female | 9 | 14 |
| Race/ethnicity | | |
| Cannot determine | 42 | 70 |
| White | 5 | 8 |
| Hispanic/Latino | 5 | 8 |
| Black | 4 | 7 |
| Native American/Pacific Islander | 3 | 5 |
| Asian | 1 | 2 |
| School level | | |
| Early childhood | 24 | 40 |
| Elementary | 14 | 23 |
| Cannot determine | 12 | 20 |
| Middle school | 7 | 12 |
| High school | 3 | 5 |
| Reported disability diagnosis | | |
| Autism spectrum disorder | 38 | 63 |
| Speech language impairment | 10 | 17 |
| Other | 10 | 17 |
| Developmental delay | 5 | 8 |
| Intellectual disability | 4 | 7 |
| Multiple disabilities | 2 | 3 |
| Orthopedic impairment | 2 | 3 |

Note. Descriptive statistics reflect the number and percentage of paraprofessionals who worked with students with these reported characteristics. Other = specific genetic disorders, cerebral palsy, and seizure disorder.
could not be determined in 37% of cases. In all but one study, paraprofessionals implemented intervention without assistance from others. Britton et al. (2017) utilized an instructional model under which a peer implemented instruction along with the paraprofessional.

Across the 19 studies, paraprofessionals delivered a variety of systematic prompting strategies to implement with their students as follows: naturalistic language strategies (33%), least-to-most prompting (22%), PRT (20%), DTT (18%), time delay (8%), most-to-least prompting (5%), and simultaneous prompting (2%). For 5% of paraprofessionals, the specific prompting strategy could not be identified. Additionally, only one paraprofessional implemented a schedule of reinforcement with clear plans for thinning reinforcement during instruction (Britton et al., 2017). When reported, the paraprofessionals implemented the intervention with the student participant between one and 19 weeks.

All 19 studies included graphed data depicting paraprofessional implementation of systematic instruction or an intervention package including systematic instruction. Data for the majority of paraprofessionals (90%) illustrated a positive effect in paraprofessional behavior, with all studies using a multiple baseline or multiple probe design to measure paraprofessional behavior. The research design for effects of systematic instruction on student behavior was a multiple baseline or multiple probe design for 67% of the paraprofessionals with the other 33% not reporting a research design or student data. Eighty-five percent of the 41 student participants with performance data illustrated a positive effect.

Quality Measures
Researchers assessed dependent measure reliability on the paraprofessional’s behavior in all of the studies with 82% being at an acceptable level, 5% at an unacceptable level, and 13% could not be determined. Researchers collected generalization data for 17% of the paraprofessionals. These data illustrated a positive effect for all. Maintenance data were collected for 43% of paraprofessionals, with a positive effect demonstrated across all paraprofessionals. Finally, researchers assessed the social validity of the paraprofessional training across 50% of the paraprofessionals. All paraprofessionals confirmed the social validity of the training they received. Researchers assessed the reliability of student outcome measure across 68% of students, with reliability at an acceptable level in most cases (80%). Reliability was not reported for 12% and was unacceptable for 7%. Generalization was assessed for 15% of the students with a positive effect for 89%. Similar to the paraprofessional data, maintenance data were collected more often than generalization data. Researchers evaluated the ability to maintain the skill with 35% of the students, all of whom demonstrated skill maintenance. Social validity specific to the systematic instruction intervention was assessed across 30% of the student participants; for these students, systematic instruction was found to be a socially valid instructional approach.

Discussion
The purpose of this literature review was to summarize single-case intervention studies in which paraprofessionals used systematic instruction to teach students with disabilities in school settings. In particular, we gathered information from 19 studies to describe study characteristics related to student and paraprofessional participants, paraprofessional training, paraprofessional-implemented systematic instruction, and study quality. Findings from this review add to the knowledge-base of paraprofessional-implemented practices by expanding on the existing body of work that has focused on a range of paraprofessional-implemented interventions for students with disabilities (i.e., Brock & Carter, 2013; Rispoli et al., 2011; Walker & Smith, 2015). Our specific focus on systematic instruction allowed for a targeted analysis of paraprofessional training practices and implementation of a well-established EBP for students with ASD and extensive support needs. In the sections that follow, we outline key findings, implications for practice, and limitations that inform future research directions.

Key Findings and Implications for Practice
There are several key findings pertaining to paraprofessional training and the conditions
under which paraprofessionals implemented systematic instruction that have important implications for the field. Overall, paraprofessional training resulted in positive implementation outcomes for paraprofessional participants, suggesting that paraprofessionals can implement systematic instruction when given adequate support to do so. This outcome is promising given the increasing reliance on paraprofessionals in instructional roles (Carter et al., 2009) and the current emphasis on paraprofessional knowledge of EBPs (CEC, 2015). In terms of training dosage, paraprofessionals were successful in implementing systematic instruction after receiving training over a relatively short period of time which is consistent with previous research (e.g., Douglas, 2012). Similar to other reviews (Rispoli et al., 2011; Walker & Smith, 2015), our findings suggest that trainers used both didactic and experiential training approaches, with a majority of training activities involving a combination of the two approaches. However, it should be noted that over a quarter of paraprofessionals received didactic training alone, an approach that has been considered less effective than experiential methods. We encourage teachers and others who supervise paraprofessionals to consider experiential methods associated with stronger implementation outcomes when didactic approaches are not effective. For example, performance feedback (Fallon et al., 2015) can be an effective strategy for providing ongoing guidance on the extent to which paraprofessionals correctly implement systematic instruction procedural elements when working directly with students (e.g., Walker & Snell, 2017).

Across the reviewed studies, researchers primarily supported paraprofessionals during training activities, as has been reported in other reviews (e.g., Brock & Carter, 2013; Walker & Smith, 2015). Due to the central role that teachers play in training and supervising paraprofessionals (Biggs et al., 2019) and issues concerning the feasibility of schools relying on outside experts (Walker et al., 2017), it will be important for schools to consider teacher-delivered training to support paraprofessional implementation of EBPs. Unfortunately, special education teachers often report feeling underprepared to serve in this role (Biggs et al., 2019; Douglas et al., 2016) and, as such, might require additional support from administration to develop these knowledge and skills to be successful. Another important consideration relates to teacher preparation programs, where pre-service teachers receive limited coursework and applied experience involving paraprofessional training and supervision (Biggs et al., 2019). As novice teachers enter the workforce, their limited knowledge and experience can serve as a barrier to supervising paraprofessionals in their use of various instructional practices, potentially leading to poorly-implemented student supports. To address these issues, it is critical for the field to develop and follow guidelines for teacher preparation programs that outline recommendations for content and clinical experiences for paraprofessional training and supervision (e.g., Yates et al., 2019).

We found that paraprofessionals commonly implemented systematic instruction with students with ASD to address social/communication skills and challenging behavior. This is not surprising given that students with ASD have support needs related to social and communication interactions and restricted interests and repetitive behaviors (American Psychiatric Association, 2013). These findings provide additional evidence that paraprofessionals can implement EBPs for students with ASD (Wong et al., 2015) to address these particular areas of need (e.g., Rispoli et al., 2011). Paraprofessional-implemented systematic instruction took place in both non-inclusive and inclusive school settings and primarily within a one-to-one instructional arrangement. Because paraprofessionals often play a prominent role in inclusive classrooms, it is important to understand how systematic instruction can be implemented within inclusive environments where other educators, such as general education teachers, may not possess the skills needed to support students with disabilities (Morningstar et al., 2016). Whole group instructional arrangements are typical of general education classrooms and may offer paraprofessionals with opportunities to utilize systematic instruction in creative ways to target a range of skills (e.g., facilitating peer supports; Brock et al., 2016). Furthermore, interventions delivered within whole group instructional arrangements as compared to one-to-one arrangements may produce
more substantial outcomes for social/communication skills and/or challenging behavior (Walker et al., 2018).

Limitations and Future Research Directions

There are several limitations that should be taken into consideration, many of which can inform future research directions. First, the scope of paraprofessional-implemented systematic instruction was limited in terms of disability categories, skill domains, and prompting strategies. Most paraprofessionals delivered the intervention to students with ASD to address challenging behavior and social/communication skills using either naturalistic language strategies, least-to-most prompting, PRT, or DTT. As such, future research should address a more diverse group of students with varying disability diagnoses across a range of skill domains. For example, given the impetus for inclusive education for students with extensive support needs, the utility of systematic instruction in inclusive classrooms to teach academic content aligned to grade-level standards must be evaluated (Morningstar et al., 2016). In addition, there is a need to examine the extent to which paraprofessionals can implement a range of systematic prompting strategies for students with disabilities, including most-to-least prompting and simultaneous prompting.

Second, our findings revealed that a large percentage of studies failed to include measures of skill generalization and maintenance; this percentage was even lower for student outcomes. Understanding whether paraprofessional-implemented systematic instruction can be generalized across different conditions and whether such effects can be maintained over time without additional trainer support is critical, as this information may inform decisions that educators make about paraprofessional training approaches. Similarly, future research must include generalization and maintenance measures for student outcomes to assess more comprehensively the effects of paraprofessional implementation of systematic instruction for students. It also should be noted that only one study reported a reinforcement thinning procedure as part of the intervention. Training paraprofessionals to utilize strategies such as this to promote skill maintenance will be critical in future studies.

Third, social validity was assessed across fewer than half of the paraprofessional and student participants. When social validity was assessed, it focused on validity of the training approaches and the systematic instruction intervention after the intervention had ended. In the future, researchers must be sure to include paraprofessionals in social validity assessments to understand whether interventions are contextually appropriate and feasible. This might include exploring a range of social validity assessment methods to address the various aspects of social validity (Horner et al., 2005).

A fourth and final limitation involves our methodological approach. We limited our analyses to descriptive summaries of study characteristic rather than statistical analyses of intervention effect (Cooper et al., 2009). Our original aim was to provide a descriptive summary, and we also recognized the limitations of conducting a meta-analysis given that some studies included in the review involved multiple component interventions. In the future, when the literature base has expanded to include a more diverse participant group and range of target skills and prompting strategies, it will be valuable to use meta-analytic strategies to (a) estimate effect sizes associated with systematic instruction delivered by paraprofessionals and corresponding student outcomes and (b) identify moderating variables that influence outcomes. This will allow researchers to estimate the magnitude of intervention effect and explore the conditions under which paraprofessional-implemented systematic instruction is most effective (e.g., prompting strategies, settings, instructional arrangements).

Conclusion

Paraprofessionals play an important role in supporting students with disabilities. Our review focused on paraprofessional-implemented systematic instruction for students with disabilities. Findings suggest that paraprofessionals can successfully implement systematic instructional practices when provided with training, resulting in positive student outcomes. Given the utility that systematic instruction holds, it will be important to move forward with research efforts to explore effective paraprofessional training strategies and
implementation across a range of student participants, skill areas, and systematic instruction procedures.

References

Studies included in the review are marked with an asterisk.


Fisher, M., & Pleasants, S. L. (2012). Roles, responsi-


Received: 5 June 2019
Initial Acceptance: 7 August 2019
Final Acceptance: 20 September 2019
Effects of General Education Teacher-Delivered Embedded Instruction to Teach Students with Intellectual Disability to Solve Word Problems

Jessica A. Bowman, John McDonnell, Joanna Ryan, Olivia Fudge Coleman, Lyndsey Aiono Conradi, and Carrie Eichelberger
University of Utah

Abstract: Current research on teaching mathematics to students with intellectual disability (ID) has demonstrated that, given evidence-based instruction, these students can acquire a wide range of skills. However, very little research has been conducted in inclusive general education (GE) classrooms and no studies have utilized the GE teacher as the intervention agent (Spooner, Root, Saunders, & Browder, 2018). This study utilized a multiple baseline design to evaluate the effects of GE teacher-delivered embedded instruction on two main variables: (1) students’ ability to solve word problems, and (2) generalization across people and materials and to untaught word problems. The data indicate that the embedded mathematics instructional package was effective in teaching students to (1) solve additive word problems with sums less than five, and (2) to generalize to untaught word problems and across people and materials. Limitations and suggestions for future research are also discussed.

Over the past decade, the field of special education has grappled with the concept of providing access to the general education curriculum for students with moderate and severe intellectual disability (ID) as outlined in the Individuals with Disabilities Education Act (IDEA, 2004; Courtade et al., 2012). The field has established that given the opportunity, students with moderate and severe ID can learn complex academic skills, such as how to solve for \( x \) in an algebraic equation (Jimenez et al., 2008) and how to solve and distinguish between types of word problems (Browder et al., 2018). However, most of this research investigates how to teach these academic skills in self-contained special education settings (Spooner et al., 2018) which presents a conflict in informing the field on effective instructional strategies to support students in inclusive general education (GE) contexts.

Access to the GE Curriculum in GE Classrooms

Questions have been raised about whether or not instruction in the core curriculum is appropriate in self-contained settings (Ryndak et al., 2008/2009), where students are less likely to have access to a teacher who is a content area expert (Mason-Williams et al., 2017), and likely to have “limited access to communication supports and partners” (Kurth et al., 2016, p. 238). Ryndak et al. (2008/2009) argue that access to the GE curriculum does not just include access to GE content standards. They assert that access to the general curriculum includes access to GE contexts, teachers, and activities as well. On the other hand, presence in a GE classroom does not necessarily indicate that a student is receiving access to the GE curriculum (Copeland & Cosbey, 2008/2009). Providing true access to the GE curriculum requires that teachers collaborate to design high-quality differentiated instruction and support students to engage in classroom activities. This means that all students are given the opportunity to demonstrate mas-
tery of the content in a variety of ways and all educators present in the classroom are involved in providing individual student supports as needed (Causton & Theoharis, 2014; Copeland & Cosbey, 2008/2009).

One way that schools achieve effective inclusion is through shared responsibility of staff members. According to Hoppey and McLeskey (2014) effective inclusive schools may not adhere to traditional roles and responsibilities of general and special educators. In a case study investigating an exemplar inclusive middle school, Olson et al., (2016) found that general educators often took on the roles traditionally filled by special educators to support students with severe disabilities in GE classrooms. General educators were found to collaborate extensively with paraprofessionals to receive information related to individual student supports and needs. In this case, it simply was not feasible for special education teachers to provide direct support to every student with a disability in each GE classroom, which required other school staff to take on those responsibilities.

Embedded Instruction

One evidence-based method to provide access to students’ individualized learning goals related to accessing the general curriculum in inclusive settings without removing them from participating in whole group activities is embedded instruction (Hudson, et al., 2013; Jimenez & Kamei, 2015). Embedded instruction is defined as systematic instructional trials that are distributed across natural opportunities throughout the day in the performance setting (McDonnell et al., 2014) and has been shown to be effective when implemented by paraprofessionals, special education teachers, peers and GE teachers (Jimenez & Kamei, 2015). In a study by Polychronis et al. (2004), four GE teachers were trained to implement embedded instruction for four students with developmental disabilities in the areas of mathematics and social studies. The teachers used a three-second constant time delay procedure with error correction and reinforcement embedded during natural opportunities. All four students reached criterion and the GE teachers reported positive perceptions of the embedded instruction procedures as well as satisfaction with student progress. GE teachers can be a vital support to students with ID in their classrooms and their involvement in educating students with ID could help to alleviate the unintended consequences resulting from these students receiving exclusive support from special education staff (decreased opportunities for peer interaction) in GE classrooms (Carter et al., 2008).

Rationale and Purpose

While the literature on embedded instruction has demonstrated that it can be used to teach a variety of discrete skills, there is little published to date on using embedded instruction to teach more complex chained skills. In the area of mathematics, students with ID may need to have opportunities to work on the prerequisite skills necessary for them to engage in higher level mathematical content (Hunt et al., 2012). This would include in the area of problem solving, which is seen as the cornerstone of mathematical literacy (National Council of Teachers of Mathematics, 2000; Van de Walle et al., 2016) and has been a prominent focus in some of the most recently published applied research studies focused on teaching mathematics to students with ID (Browder et al., 2018; Spooner et al., 2018). Many of the studies teaching word problem solving use evidence-based practices for teaching students with ID mathematics, including the use of systematic instruction, graphic organizers, and manipulatives (Spooner et al., 2018). However, the effectiveness of many of these evidence-based practices have yet to be demonstrated in inclusive settings, where the structure of the classroom and the people who are available to deliver instruction differ significantly from a typical special education setting (Hudson et al., 2013).

The purpose of this study is to determine the impact of a general educator-delivered embedded mathematics instructional package on the word problem-solving ability of three students with ID. We also wanted to examine the extent to which students with ID would generalize what they learned about solving word problems to untaught examples and across people and materials. Evaluation of the embedded mathematics instructional package focused on (a) the percentage correct of task
analysis steps per probe, (b) the total number of correct problems solved, and (c) the mean performance of each task analysis step across phases. Lastly, this study seeks to determine whether GE teachers viewed the embedded instructional procedures as useful and feasible.

**Method**

**Participants**

Once institutional review board approval was obtained, the first author contacted the special education director in the cooperating school district and asked them to nominate elementary special education teachers who taught students who were included in GE classrooms for mathematics. Three elementary teachers were nominated and then contacted to set up a time to discuss the study and the prerequisite skills needed for each participant. To be included in the study, participants met the following criteria (1) have a moderate or severe ID; (2) be eligible for alternate assessment; (3) receive instruction in a GE classroom for mathematics; (4) consistently demonstrate the ability to (a) communicate a choice, (b) identify numbers 1–10, and (c) count to 10 with one-to-one correspondence; and (5) have a teacher report that solving word problems would be a valuable skill for them to learn. Each teacher then sent home consent forms to each student who fit the participation criteria. Ten consent forms were sent out and 10 consent forms were returned between the three special education teachers. Once consent forms were returned, the first author met with each student to explain the study and to gain assent. Once assent was obtained, each student was assessed to see if their mathematics skills fit the participation criteria. Once students were assessed and if they were found to have the prerequisite skills, their GE teacher was contacted to see if they would be interested in participating in the study. Four students demonstrated all prerequisite mathematics skills consistently, and three out of the four GE teachers indicated an interest in participating.

**Students.** Alex was a 7-year-old Hispanic male in first grade. He qualified for special education services under the developmental delay classification and participated in an early literacy alternate assessment. His home language was Spanish, and he communicated mostly in English at school. Alex spent about 30% of his day in GE settings without paraprofessional support. His IEP included mathematics goals related to counting and one-to-one correspondence, which he had already met when the study began. According to the Wechsler Nonverbal Scale of Full-Scale IQ, Alex had an IQ score of 61. According to the Adaptive Behavior Assessment System, 2nd Edition, (ABAS-II), his adaptive composite was 46.

Emma was an 11-year-old White female in the fifth grade. She qualified for special education services under the ID classification and currently participated in the state’s alternate assessment. Emma spent all of her time in GE settings with no paraprofessional support. She communicated in three to five-word phrases, and her IEP included a goal for mathematics to be able to add and subtract numbers up to 20. Emma’s IQ was 53, according to the Wechsler Preschool and Primary Scale of Intelligence, 3rd Edition (WPPSI-III). According to the Adaptive Behavior Assessment System, 2nd Edition (ABAS-II), her general adaptive composite was 64.

Brandon was a 10-year-old African American male in the fourth grade. He qualified for special education services under the autism classification and currently participated in the state’s alternate assessment. He spent about 30% of the day in GE settings, with the support of a paraprofessional due to some disruptive behavior. Brandon communicated in two-word phrases and had IEP goals related to solving one-step real-world problems with addition. The school was unable to obtain an IQ score when Brandon was assessed. According to the Gilliam Autism Rating Scale, 2nd Edition (GARS-2), Brandon scored a 125, indicating a high probability of autism. His general adaptive composite was 47 based on the Adaptive Behavior Assessment System-Second Edition (ABAS-II).

**Teachers.** Three GE teachers implemented all instructional trials and one special education teacher conducted generalization probes.
Ms. V was Alex’s first-grade GE teacher with 10 years of teaching experience. She had a bachelor’s degree in elementary education and had 19 students in her class. Ms. V had experience supporting students with ID in her classroom for four years. Ms. K was Brandon’s fourth-grade GE teacher with six years of teaching experience. She had a bachelor’s degree in elementary teaching (first through eighth grades) and had 25 students in her class. Ms. K had experience supporting students with ID in her classroom for four years. Ms. B was Emma’s fifth-grade GE teacher with 22 years of experience. She had a bachelor’s degree in elementary education and a master’s of education in teaching and learning. She also had an English and reading level one endorsements. She had 28 students in her classroom and has been supporting students with ID in her classroom for eight years. Ms. M was a special education teacher in her ninth year of teaching, eight of which she has taught students with ID. She had a bachelor’s degree in special education with an emphasis on mild/moderate disabilities and a minor in teaching English to speakers of other languages. She also had a master’s degree in special education and an administrator’s certificate. Ms. M currently had 16 students with disabilities on her caseload, including all three student participants in this study.

Setting

This study took place in one elementary school in a suburban school district in the Mountain West region of the United States. The school was inclusive and was in a district that had policies supportive of inclusive education. The elementary school served 455 students grades kindergarten through sixth grade. Twenty-three percent of students at this school qualified for free or reduced lunch and 12% qualified for special education services.

All teaching sessions took place in GE classrooms during the class mathematics block. Each teacher’s mathematics block usually consisted of whole group instruction followed by group and/or individual student practice. All GE classrooms were working on skills that were more complex than the skills taught in this study (e.g., skip counting, multiplying and dividing decimals, fractions), especially considering that the target skill most closely aligns to a kindergarten level skill (K.OA.A.2; National Governors Association Center for Best Practices & Council of Chief State School Officers, 2010). To ensure a quiet environment, the researcher conducted all probes in the special education classroom or hallway.

Target Skill

The target skill for this study was to solve word problems with sums up to five. In discussion with the special education teacher and in reviewing student IEP goals, this was reported to be a priority skill for all participants. In addition, it is aligned directly or as a prerequisite skill with the following state alternate achievement standards [otherwise referred to as the Dynamic Learning Maps Essential Elements (2013)] in first, fourth, and fifth grades:

- EE.1.OA.1.a. Represent addition and subtraction with objects, fingers, mental images, drawings, sounds (e.g., claps), or acting out situations.
- EE.4.OA.3. Solve one-step real-world problems using addition or subtraction within 100.
- EE.5.NBT.5. Multiply whole numbers up to $5 \times 5$.
- Demonstrate the concept of addition. (untested skill between the initial and distal precursor linkage levels).

Materials. Materials for this study included addition word problems with sums less than five, a graphic organizer, manipulatives, and an instructional calendar for each GE teacher. Ten change-addition and group word problems were developed for this study utilizing the names of the students and their classmates as well as high-interest themes (e.g., recess, class parties, going ice skating). The word problems were developed in a standardized four-sentence format (the first sentence provided the context, the second and third sentences provided the key noun(s) and quantities, and the last sentence provided the question that needed solving) with picture supports included above each noun, according to the guidelines set forth by Spooner et al., (2017; e.g., “It was recess. Luca brought two balls out to the playground. Sophia brought two balls out to the playground. How
many balls were on the playground?”). In addition, they were evaluated using the Flesch-Kincaid readability test to ensure that the reading level was developmentally appropriate for each student. All word problems were printed on 8 ½ × 11 paper that was cut in half, laminated for durability and were randomly assigned to the teaching set and the generalization set (five word problems in each set). Once they were assigned to a set, the teaching set of word problems was numbered to control for random presentation by the GE teacher.

The graphic organizer was an 8 ½ × 11 size laminated mat that included three circles, similar to the graphic organizer used by Browder and colleagues (2012, p. 313). From left to right, it included the first circle, an addition sign, the second circle, the equal sign, and the final circle (see Figure 1). The manipulatives used varied by each GE teacher (depending on what was utilized in their classrooms most frequently). Two GE teachers used base-10 blocks, and the other used counting bears. Probes conducted in the special education classroom used base-10 blocks, and situational generalization probes were conducted using Unifix cubes.

In addition to the word problems, graphic organizer, and manipulatives, each GE teacher was given an instructional calendar that indicated which problems to teach on each day. To determine the order, the numbers one through five were put into an online random number generator and plugged into the calendar. If two numbers appeared twice in one day, the next sequence of three numbers from the random number generator was used.

**Dependent Variable**

Three dependent variables were measured throughout the study. The primary dependent variable used to determine the effectiveness of the intervention package on students’ word problem-solving strategy acquisition and generalization was the percentage of correct steps on the five-step task analysis for solving one-step word problems with sums up to five. The task analysis was: (1) Count the first num-

---

1. Secure attention
2. Present word problem
3. Present word problem and stop after first fact.
4. Deliver the controlling prompt (model) – “We count 1. Watch me - 1.” (move counters back) “Your turn!”
5. Read the next portion of the story and stop after 2nd fact.
6. Deliver the controlling prompt (model) – “We count 4. Watch me – 1, 2, 3, 4.” (move counters back) “Your turn!”
7. Read the rest of the math story.
8. “We need to find out how many ____ (dollars)” (use the noun from the problem)
9. “Watch me! I move them all over and count - 1, 2, 3, 4, 5, 6, 7.” (move the pieces back) “Your turn!”
10. “There are 7 dollars altogether! How many dollars?” (“7 dollars”).

---

Figure 1. Teacher Script and Graphic Organizer.
ber of manipulatives into the first addend circle; (2) Count the second number of manipulatives into the second addend circle; (3) Move all the counters over to the sum circle; (4) Count how many altogether; and (5) State the sum with the label that matches the noun in the final question. To calculate the percentage of steps completed correctly, the number of correct steps completed was divided by the number of steps and multiplied by 100. The criterion for mastery and for changing phases was 80% steps completed correctly across two probe days.

The secondary dependent variables included the total number of word problems answered correctly (measured by having all TA steps completed correctly) and, to identify patterns of acquisition and errors across steps of the TA, the participants’ mean performance on each task analysis step across phases was also determined.

These dependent variables were measured by conducting three different probes twice a week: a teaching set probe, a situational generalization probe, and an untaught set probe. The teaching set probe, conducted by a member of the research team, assessed the same set of five word problems that were being taught by the GE teacher. The situational generalization probe included the same set of word problems but were conducted by the special education teacher with a different set of manipulatives. The researcher also conducted untaught set probes which assessed students’ generalization to a set of five word problems that the student had not been taught.

Research Design

This study utilized a multiple baseline design (Johnston, 2010). Each participant began baseline data collection together. When the first participant established a stable baseline trend, she began the intervention and was given a probe twice a week. The other two participants remained in baseline until the first participant established an upward trend in intervention data (i.e., percentage of task analysis steps completed correctly). At this point, the second participant entered intervention. Once the second participant demonstrated an upward trend in intervention data, the last participant began intervention. All participants remained in intervention until they reached 80% of the task analysis steps correct over two consecutive sessions. Once they reached criterion, they entered Phase II, where they were assessed on each dependent variable once a week until each participant had at least three weeks of data.

Procedure

Teacher training. Before each GE teacher began the intervention, they were trained to implement the procedures by the first author. The training took approximately 20 minutes. The training addressed the research questions and a rationale for the importance of the study, a description and research support for embedded instruction, the expectations of the student and teachers, and how to implement simultaneous prompting. Each teacher role-played the procedures with the first author. Then, on the first day of the intervention, the first author modeled how to implement the procedures with the student for the GE teacher, and then the first author watched the teacher implement it with the student, providing feedback as needed until the teacher was able to implement the intervention with 100% fidelity across three word problems. Each teacher was given a script to refer to if they needed it (see Figure 1).

Baseline and probe procedures. Baseline and probe procedures were conducted twice weekly using the same procedures in the special education classroom or at a table in the hallway. Each probe consisted of 15 word problem trials (five trials for the teaching set of word problems, five for situational generalization, and five for generalization to the untaught set of word problems). During each probe session, trials were assessed in a single-opportunity format (i.e., the trial was ended as soon as a student made an error on a step of the task analysis). This helped to prevent student frustration by not asking them to persist in a procedure they had not learned and also helped to avoid incidental learning. If the participant made an error on any step of the task analysis, no correction was provided, and the trial was ended.

During each probe, each student was provided with the manipulatives and the graphic organizer to use, but they received no
prompts, instruction, or feedback on how to use them. The word problem was read through completely one time by the researcher. Then the researcher would say, “Now you will solve the problem.” The student would be expected to follow the steps in the task analysis without prompting. Any portions of the task analysis completed correctly would be checked. If the student did not label the sum consistent with the noun in the word problem (e.g., “four balls”), the researcher would ask for clarification by saying “four what?”.

Instructional procedure. GE teachers implemented all embedded instructional trials in the GE classroom, in their mathematics block during naturally occurring breaks (usually during independent student practice or transitions) using simultaneous prompting (with corrective feedback if errors were made) and total task presentation. The teachers were instructed to have at least five minutes between each word problem and to do three word problems per day for five days a week.

To begin, the teacher would look at their instructional calendar for the day to see which three word problems to teach that day. Once the teacher secured student attention, they would present the entire word problem by reading it aloud. Then they would say, “Now it’s your turn to solve!” They would present the word problem again and stop after the second sentence and provide a controlling prompt (“George was at the pet store. He saw three dogs. We count three. Watch me – 1, 2, 3. Your turn!”). Once the student counted out three manipulatives into the circle, the GE teacher would read the next portion of the story, and stop after the third sentence (“Then he saw two cats. We count two. Watch me – 1, 2. Your turn!”). The participant would then count out two more manipulatives into the circle, and the GE teacher would read the rest of the word problem and provide another controlling prompt (“How many animals did George see? We need to find out how many altogether. Watch me! I move them all over and count – 1, 2, 3, 4, 5.” Then they would move the manipulatives back to their original position and say, “Your turn!”). Once the student moved all the manipulatives over to the sum circle and counted them, the GE teacher would repeat the last question with a controlling prompt (“How many animals did George see? He saw five animals. You say five animals.”). Once the participant verbalized the solution and the trial was finished, the GE teacher would provide social reinforcement. If any errors were made, the GE teacher provided a specific verbal prompt to ensure student success (“Try again. The word problem said ‘two cats.’ Count two cats.”). Then, they would complete the second and third word problems during natural opportunities in the mathematics block using the same procedures. Once the teacher completed each problem with the participant, they would put a check mark next to that number on the instructional calendar. If they were unable to complete any of the problems due to time restraints or absences, they would indicate that by circling the uncompleted word problems.

Generalization and Phase II probe procedures. Generalization procedures were consistent with the baseline and probe procedures described above. For generalization, all participants were continuously assessed for situational generalization across people and materials, where the special education teacher (different person) would provide probes using different manipulatives (Unifix cubes or wooden blocks) on the teaching set of five word problems. During the untaught word problem generalization probes, the researcher conducted the probes on a set of five word problems that the student had not been taught before.

Since the target skill is an essential prerequisite skill for many other more complex mathematics skills, this study sought to implement a maintenance procedure that more closely aligns with actual practice where students are supported to retain the knowledge and skill required with a reduced number of trials and with lower levels of instructional support. Once the participants met the criterion of 80% across two probes across both the acquisition and generalization measures, they entered Phase II of the intervention, which served to assess maintenance when the dosage of instruction was cut from 15 word problem trials a week to four word problem trials a week. Once in Phase II, the GE teachers faded from providing embedded instruction on three word problems every day using simultaneous prompting to providing embedded in-
struction on two word problems twice a week using the system of least to most prompts. The students were probed on each dependent variable once a week until the last participant had a minimum of three Phase II data points.

Social Validity

The special education teacher identified participants for whom learning how to solve word problems would be a valuable and appropriate skill. She was informally interviewed at the end of the study about her perceptions on the skill taught, the utility of the intervention, and the ease with which it was implemented.

GE teacher perceptions of the utility and feasibility of the intervention procedures were assessed at the end using a survey. The three-question survey asked a question about whether the intervention could be planned, prepared, and implemented in a reasonable amount of time, whether it was effective for the participant, and whether they expected they would use the intervention again before the end of the school year. GE teachers rated each statement on a 4-point scale (1 – strongly disagree – 4 – strongly agree).

Intervention Fidelity and Interobserver Agreement

GE teacher implementation fidelity probes were conducted through in-person observation by scoring whether the teacher correctly followed the steps in the task analysis. Fidelity probes were run during 40% of the instructional sessions, with a second observer also collecting interobserver agreement (IOA) during 38% of all of the fidelity probes. Teacher fidelity was calculated by dividing the number of correctly implemented steps by the number of correctly implemented steps plus the incorrectly implemented steps and multiplied by 100. Overall fidelity was 95.3% with a range of 80–100%. IOA on the teachers’ implementation of the procedures was calculated by dividing the number of agreements by the number of agreements plus the number of disagreements. In addition to teacher fidelity, the GE teachers also reported when trials were missed due to absences or lack of time on their instructional calendars. By the end of the study, teachers reported having completed 83.7% of the trials listed on the instructional calendar. This translated to an average of 12–15 GE teacher-implemented word problem trials per week (out of the prescribed 15).

IOA data were collected in-person on percentage correct of task analysis steps during 100% of the baseline probes, during 44% of the intervention probes, and during 53% of Phase II probes. Baseline probe IOA was 100%. Intervention probe IOA ranged from 86%–100% with a mean of 99.5%. Phase II probe IOA was ranged from 99.7–100%.

Results

Mathematical Word Problem-Solving

Percent of steps correct. Figure 2 shows the percentage of correct steps of the word problem-solving task analysis for all three participants across the probe and generalization measures. Probe data are reported across the teaching set of word problems (squares), the untaught set of word problems (circles), and the situational generalization of the teaching set (cross). The data show that the participants demonstrated no steps correct during baseline. At the start of the intervention, each participant’s data demonstrated a gradual change in trend on the taught set of word problems and on the generalization sets (across people and materials and to the untaught word problems). All students met criterion with all three measures and maintained performance during Phase II for three to five weeks. In the Phase II, all students maintained high fidelity to the task analysis steps. While there is the possibility of testing effects since participants were tested on the same 10 word problems during each probe, a vertical analysis of the data shows that participants who were in the baseline stage for a longer amount of time did not show improvement, so this is an unlikely threat to internal validity.

To supplement the visual analysis, the effect size of the intervention was calculated for the percent of steps correct for the untaught set of word problems by using an online calculator to determine the Tau-U (Parker et al., 2011; see http://www.singlecaseresearch.org/calculators/tau-u). The aggregated Tau-U for all students was 0.95 (very large effect; Vannest & Ninci, 2015). Emma and Brandon’s individual
data indicated a Tau-U of 1 (very large effect), and Alex’s data indicated a Tau-U of 0.88 (very large effect).

Number of word problems solved correctly. Figure 3 shows the total number of word problems answered correctly per probe. All students answered no word problems correctly during baseline, and Emma and Alex demonstrated gradual and consistent positive increases in trend once the intervention began. Brandon, however, saw some inconsistent increases in taught and untaught word problems answered correctly once the intervention began, but most probes showed that he answered no questions completely correct during the intervention phase. In Phase II, however, all three students had a large majority of the probes where they answered four or five
out of the five questions correctly, with untaught word problem generalization probes having some low outliers early on in the phase for Emma and Alex. One possible threat to internal validity is the possibility that participants memorized the five taught word problems. Additional analysis of the data indicates that the word problems participants solved correctly were not the same across probes, and that, in fact, student errors centered not...
around word problems at all, but around the step(s) with which they had the most difficulty.

Mean performance on task analysis steps across phases. The average number of times each student demonstrated each step of the task analysis correctly by phases is displayed in Table 1. All students showed marked increases in Steps 1 through 4 of the task analysis from baseline to the intervention phase. However, Step 5 of the task analysis (state the sum of the problem with the label that matches the noun in the final question) did not show large increases until Phase II for all three participants. In Phase II, all three students were performing each step close to 100% correct for each probe. However, Emma and Alex still showed lower mean performance on labeling the sum on untaught word problems in Phase II.

Social Validity

The social validity survey demonstrated consistently high teacher ratings. At the end of the study, all three teachers said they strongly agreed that “... the individualized academic lessons for <student> presented within the typical flow of my class can be effectively planned, prepared, and implemented within a reasonable amount of time.” Teachers agreed or strongly agreed that the intervention was effective and that they planned to use it with another student before the end of the school year. One teacher commented, “I thought this intervention process was a fantastic way to provide some individual instruction in word problems to (Alex) in the regular classroom. I was able to incorporate time to meet with him and still teach/monitor my class and include him in as many of the activities as possible. I feel that he enjoyed the intervention time and felt successful in his efforts.” In the informal in-

### Table 1

<table>
<thead>
<tr>
<th>Participants' Mean Performance on Task Analysis Steps across Phases</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Task Analysis Steps</strong></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td><strong>Emma</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td><strong>Brandon</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
<tr>
<td><strong>Alex</strong></td>
</tr>
<tr>
<td>1</td>
</tr>
<tr>
<td>2</td>
</tr>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
<tr>
<td>5</td>
</tr>
</tbody>
</table>

Note: B = baseline, I = intervention, I–II = Intervention Phase II. Task analysis steps: 1 = Count the first number of manipulatives into the circle; 2 = Count the second number of manipulatives into another circle; 3 = Move all the counters over to the sum circle; 4 = Count how many altogether; 5 = State the sum with the label that matches the noun in the final question.
terview with the special education teacher, she reported that the skill of solving word problems was very important for the participants to learn and that the intervention was very useful and somewhat easy to implement. Ms. M also noted that she saw evidence of skill transfer to the alternate assessment, especially for Brandon.

Discussion

The purpose of this study was to measure the effectiveness and social validity of a GE teacher-delivered embedded mathematics instructional package on the word problem-solving ability of three students with ID in inclusive mathematics classrooms. The mathematics instructional package was delivered by the students’ GE teachers and included embedded instructional trials with simultaneous prompting, distributed across the mathematics block in the GE classroom. All three participants learned to solve the teaching set of five word problems across people and materials and generalized to a set of five untaught word problems.

An essential component of mathematical proficiency is to be able to make sense of a mathematics problem (Van de Walle et al., 2016). While the participants solved word problems related to the experiences they encounter in their everyday lives and generalized to untaught word problems, it is still unclear whether students would be able to apply these strategies in real-world contexts. It was the intent of the intervention for students to not just get the ‘answer,’ but to be able to make sense of the problem as well as the solution by labeling their sum with what it is they were trying to find out. When inspecting the data related to the participants’ mean performance on each task analysis step across phases (Table 1), it became apparent that the participants were not labeling the sum for a majority of the problems assessed in the intervention phase. This explains why participants’ total word problems solved correctly by probe was lower in the intervention phase (Figure 3). If the student did not label the sum correctly, the problem was incorrect. However, as the participants moved to Phase II and instruction and probes slowed and simultaneous prompting was replaced by least-to-most intrusive prompting, performance on this step increased, possibly due to the increase in corrective feedback. In addition, all three participants’ generalization to untaught word problems still lagged slightly behind the other two measures (see Figure 2). Given the issues students had with labeling their sums and the lag in demonstrating mastery with the untaught word problems, we speculate that the teaching set may have been too small given the size of the stimulus class, and a more systematic approach to developing and selecting the problems included in the teaching set may have improved generalization.

Systematic instruction, use of graphic organizers, and use of manipulatives are recognized as evidence-based practices in teaching mathematics to students with ID (Spooner et al., 2018). The results of this study are consistent with those findings. The findings of this study are also consistent with studies that have used similar mathematics instructional packages (Browder et al. 2012; Browder et al., 2018). However, a few main differences apply. First, this study used simultaneous prompting (as opposed to least to most prompting) because of its ease of use and efficiency for GE teachers. In retrospect, simultaneous prompting may not have been the best strategy because once teachers were providing least to most prompting in Phase II, variability in student performance decreased, likely because students were given more opportunities to demonstrate the target skill with corrective feedback. Next, this study did not use any sort of student checklist, as we were not teaching students to discriminate between types of word problems, and wanted to keep the intervention as manageable as possible for the GE teachers to implement.

This study is also consistent with studies showing that GE teacher-delivered embedded instruction is effective in teaching students with ID academic skills within GE contexts (e.g., Johnson & McDonnell, 2004; Polychronis et al., 2004). Further, this study extends the literature by demonstrating that the instructional components (systematic instruction, graphic organizers and manipulatives) can be effective when delivered by a GE teacher in a GE setting to teach more complex chained skills using embedded instruction. In addition, this study demonstrates that embed-
ded instruction can be provided in smaller doses to support students to sustain academic skills, not just acquire them.

**Limitations**

This study had several limitations. First, because it was conducted in an inclusive school district, results will need replication in districts without such policies to improve external validity. Next, because multiple components were utilized in the intervention package, it is not possible to identify which components had an effect on student performance. Related to the acquisition of the mathematics concepts, this study had several limitations. One issue was that the intervention taught additive word problems with sums up to five, which could be solved using just one type of graphic organizer and required no discrimination between types of word problems. Students also had difficulty with labeling the sum, suggesting that their conceptual understanding of addition was not developed until Phase II. Since this study could not directly control for outside instruction in word problem solving, there is a possibility that student acquisition of the skill could have been due to classroom instruction even though that instruction was generally very different from the target skill. Lastly, this study did not assess students’ generalization to naturalistic settings or activities, which is essential to ensure that the skills learned can be used in meaningful ways outside of a probe setting.

**Conclusion**

One way to mitigate the perceived barriers of inclusive education could be to promote shared responsibility for supporting student progress in the GE curriculum and individualized learning goals (Hoppey & McLeskey, 2014). This study shows a possible way to increase the capacity of GE teachers to support students with ID in learning prerequisite mathematical problem-solving skills to promote access to the GE curriculum in GE settings. Future research should continue to investigate how to teach more complex skills using a variety of different formats (e.g., embedded instruction, cooperative learning, small groups; Copeland & Cosbey, 2008/2009) and intervention agents (e.g. GE teachers, peers) to provide effective and inclusive educational programs for students with ID.

**References**


Hunt, P., McDonnell, J., & Crockett, M. A. (2012). Reconciling an ecological curricular framework...
focusing on quality of life outcomes with the development and instruction of standards-based academic goals. Research and Practice for Persons with Severe Disabilities, 37, 139–152.


Received: 11 April 2019
Initial Acceptance: 10 June 2019
Final Acceptance: 24 August 2019
Dissemination and Decision-Making: Factors Related to Pre-Service Practitioners’ Selection of Practices for Students with Autism

Maria L. Hugh and LeAnne D. Johnson
University of Minnesota Twin Cities

Veronica P. Fleury
Florida State University

Abstract: Though increasing the teaching of evidence-based practices (EBPs) for students with autism spectrum disorder (ASD) within pre-service preparation programs is one approach to addressing the need for increased use of EBPs in the field, this approach likely oversimplifies the problem. We surveyed 60 pre-service practitioners’ decision-making around selection of EBPs and non-EBPs that are commonly used for students with ASD. We then explored factors related to participants’ familiarity with, approval of, and likelihood to select practices and evaluated the effects of source-trustworthiness on their decision-making. Participants were familiar with most practices but more often approved of and selected EBPs than non-EBPs. Interestingly, participants were more likely to select one EBP when it was presented by an untrustworthy source than when presented by a trustworthy source. Based on these findings, we discuss recommendations for pre-service programs to promote the use of EBPs through active dissemination and consideration of individual factors in an effort to move EBPs into routine practice.

As awareness and identification of autism spectrum disorder (ASD) continue to expand (Christensen et al., 2016), efforts to improve outcomes for individuals with ASD, whether based in science or good intention, are also expanding. Teaching students with ASD requires the use of intentional, explicit interventions and practices to remediate difficulties and support learning within an individual’s areas of need. To accomplish this, special educators are expected to use evidence-based practices (EBPs) that have an established research base supporting their effectiveness in supporting skill development for specific populations (IDEIA, 2004). Though there are identified EBPs for students with ASD (Green & Ricciardi, 2015; Wong et al., 2015), researchers have found that practitioners are not using them with sufficient frequency to achieve broad scale impact on the outcomes of students with ASD (Brock et al., 2014; Morrier et al., 2011).

The problem is not as simple as just focusing on the knowledge and skill needs of practitioners as a means by which to increase the use of EBPs when addressing the needs of students with ASD (Dingfelder & Mandell, 2011). Since ASD was first recognized by Dr. Leo Kanner in 1943, intervention usage has not always corresponded with scientific findings (Travers, 2017). Widespread, present day use of non-evidence-based practices such as sensory integration techniques and alternative diets (Green et al., 2006), establishes a need to further explore factors that relate to why practitioners use practices that do not have evidentiary support.
Frameworks Describing Practitioners’ Selection and Use of Practices

Understanding the processes by which EBPs and non-EBPs are adopted is an imperative. Emerging frameworks for examining implementation and decision-making offer insight into the processes by which practices are or are not used for students with disabilities (Cook & Odom, 2013). These frameworks point to several practitioner level factors that may influence the adoption of instructional strategies, such as their preferences, abilities, knowledge, and motivation. Wills and Holmes-Rovner (2006) established a simplified decision-making model for patients in healthcare, which can be useful in understanding the drivers behind educators’ adoption of practices. Building from that literature, we propose a model (Figure 1) with factors we hypothesize may be related to practitioners’ decisions to select and use certain practices.

Practitioners make decisions to adopt practices based on the information provided to them and how that information aligns with what they value (Wills & Holmes-Rovner, 2006). In his innovation-decision process, Rogers (2005) described the process by which individuals adopt a practice as a multi-step sequence that begins with gaining knowledge and becoming familiar with a practice. Upon gaining information, through personal experience or various forms of dissemination, the practitioner still must engage in decision-making regarding whether that information aligns with their values and beliefs in order to approve of and be persuaded to use a practice (Rogers, 2005; Wills & Holmes-Rovner, 2006). Each of these stages is likely influenced by different factors that comprise an individual’s decision-making context, which ultimately influences whether or not a practice is adopted. With the goal of increasing broad scale adoption and use of EBPs for students with ASD, it seems plausible that understanding the decision-making context of practitioners enrolled in pre-service programs may inform new opportunities to influence the likelihood of adoption of practices that have evidentiary support.

Decision-Making Context for Pre-service Practitioners

Pre-service practitioners’ familiarity with EBPs. Pre-service practitioners (PSPs) are in a critical period in their careers during which beliefs, preferences, and knowledge about practices are being formed (Scheeler et al., 2016). Familiarity with a practice represents knowledge that may be co-constructed through personal experience and information presented within a pre-service program. To date, researchers have examined PSPs’ familiarity with practices in a tangential way relative to the training they have received, rather than as a determinant of implementation (e.g., Hsiao & Peterson, 2018).
It is commonly believed that special educators need to be trained in content, pedagogy and scientifically supported practices (Leko et al., 2012; Smith et al., 2010). Teacher preparation programs, however, are found to vary widely by state and licensure area with only a small portion including ASD specific competencies for PSPs (Barnhill et al., 2016). In a recent study, Hsiao and Peterson (2018) found that only 12–55% of special educators serving students with ASD were directly taught by their professors how to use any EBPs for students with ASD. Concerns about the limited teaching of EBPs by professors in pre-service programs are magnified when we consider pre-service programs as the primary means by which practitioners also learn to discern between EBPs and non-EBPs and incorporate new information with their own personal experiences. Understanding ways to increase familiarity while building the value of EBPs as compared to non-EBPs is necessary to decisions to adopt practices for use (Guckert et al., 2016), particularly when some degree of familiarity and value may already exist through personal experiences. Knowing familiarity may be informed by different sources, understanding how different sources of information relate to PSPs’ perceptions of different types of practices and decisions about the use of practices is warranted.

Source trustworthiness. The source that presents information about a practice may influence perceptions of the practice, ultimately facilitating or impeding implementation of EBPs. Cook et al.’s (2013) model of active dissemination of EBPs in Special Education identifies credibility as an important message characteristic. That is, the source of information about EBPs must be credible, or trustworthy, from the perspective of the practitioner. A variety of information sources such as; celebrities, researchers, physicians, and practitioners are shown to be differentially influential when an individual determines whether or not to adopt a practice (Landrum et al., 2002; Merk et al., 2017; Metz et al., 2015; Van Boekel et al., 2017). Specific to educators, researchers found that educators rated practice-based sources (e.g., colleagues) and forms (e.g., workshops) as more usable and trustworthy than academic-based sources (e.g., professors) and forms (e.g., college courses, professional journals; Knight et al., 2018; Landrum et al., 2002). Extending these findings, Knight et al., found that educators reported frequently using what they learned from practice-based sources and never using practices learned from researchers. In a similar study of PSPs’ perceptions of source and form, Merk et al. found, in contrast, that PSPs reported higher practical value for content presented in the form of a scientific study. The researchers also found, however, that PSPs ascribed a higher practical value when the source was a practitioner that they assumed had a high degree of expertise, most often due to the PSPs’ perception that there was a high degree of alignment between the practitioner’s beliefs and their own. Rogers (2005) and Cook et al. (2013) suggest perceptions about the people who are the source of information matter when individuals, in this case PSPs, engage in their decision-making process to use a practice by assessing source trustworthiness and considering their approval of certain practices with and without evidentiary support.

Pre-Service Practitioners’ Approval of Practices

Approval is a construct that represents whether a practitioner finds a practice to be compatible with their values and beliefs such that they are likely to implement it (Montano & Kasprzyk, 2015; Rogers, 2005). Both the source and form through which information is obtained can greatly influence whether a practice is considered “compatible” with practitioners’ “existing values, past experiences and needs” (Rogers, 2005, p. 240). Confirmation biases, when an individual seeks or interprets information that aligns with their pre-existing notions, contribute to whether or not a practice may be considered believable or efficacious, and likely as a result, whether or not an individual approves of a practice (Lewandowsky et al., 2012). The degree of compatibility between the familiarity that PSPs’ have with a practice (co-constructed through experience and varying sources/forms of information) and perceptions of source and form through which information is presented in pre-service programs may serve an important role when examining whether or not a PSP approves of certain practices for use in the field. Further, that role may be influenced by how trustworthy a PSP perceives different
sources and forms of information. The interplay of confirmation biases and perceptions about trustworthiness may contribute to the inadvertent reinforcement within pre-service programs of PSPs’ approval or disapproval of certain practices irrespective of whether or not those practices are EBPs.

The Breakdown between Information, Dissemination, and Use of EBPs

In recent years, two organizations spearheaded efforts to enhance the quality of information available about practices that are intended for use with children and adolescents with ASD, the National Professional Development Center on ASD (NPDC; see Wong et al., 2015) and the National Autism Center (NAC; see NAC, 2009). Both organizations used quality indicators for research to ensure that only studies with scientific merit and rigorous methodology were evaluated to ensure conclusions about evidentiary support are based on reliable and valid data. The NPDC (Wong et al., 2015) and NAC (2009) reviews were instrumental in making information available about whether practices are evidence-based. The NPDC and NAC have used various dissemination strategies to familiarize practitioners with EBPs, including practice-based (e.g., webinars, videos, downloadable practice guides) and research-based (e.g., listings of scientific studies, expert summaries of findings) resources.

The availability of information that is disseminated through sources that researchers would agree are trustworthy, appears to be insufficient to encourage broad scale use of EBPs and discourage the use of non-EBPs (Brock et al., 2014; Morrier et al., 2011). Researchers are finding that very small portions of surveyed teachers report using EBPs (Morrier et al., 2011). EBPs are actually implemented infrequently even when reported to be used more frequently than non-EBPs (Knight et al., 2018), and some teachers are most frequently using practices known not to have evidentiary support, such as sensory integration (Hess et al., 2008). With practitioners’ decision-making and implementation context potentially limiting the use of EBPs, preparation programs may be in a position to facilitate greater use of EBPs in the field. Understand-}

---

**Study Purpose**

The purpose of this study was to explore factors that may be related to PSPs’ likelihood to adopt practices for students with ASD. We explored PSPs’ decision-making contexts, including whether they were familiar with a practice and if this familiarity was related to personal, educational or training experience. The model proposes that practitioner’s approval of a practice is shaped by their experience and familiarity with a practice, along with the messaging they receive about the practice, including the source that presents the given information (Landrum et al., 2002). The research questions were:

1. What percentage of PSPs are familiar with, approve of, and are likely to use or recommend certain practices intended for children with ASD (EBP and non-EBP), and are there differences in relations between evidentiary support for a practice and PSPs’ evaluations of the practice?
2. Are there relations between PSPs’ decision-making context and their approval of or likelihood to use or recommend practices?
3. Is the trustworthiness of the source providing information about a practice related to PSPs’ approval of or likelihood to use or recommend practices?

**Method**

**Participants and Recruitment**

With approval from the University Institutional Review Board, participants were recruited at the beginning of a fall semester at one large, public midwestern university. Two methods were used to recruit PSPs for participation. First, a department administrator sent recruitment emails that included the survey link to first-year students enrolled in Master’s and Bachelor’s special education teacher
preparation programs \((n = 150)\). Next, we contacted instructors from five introductory special education courses offered in a single semester \((n=143\) students, many of whom were also first-year students). All course instructors forwarded an email from the research team to their students. The email explained the study and contained instructions and links to complete the voluntary survey. The email instructed participants to complete one of two forms, depending on the first letter of their last name, and asked them to complete the survey on their own time. Participants were provided an opportunity to opt into a lottery to win a small gift card. No financial incentives were offered for participating. The first author visited three of the five classes to introduce the study in person and invite participation. Students who received multiple invitations to participate in the study were instructed to only complete the survey once.

The initial sample included 93 participant surveys. A total of 33 surveys were excluded due to missing responses or failed attention checks built into the survey as described below. The final sample included 60 completed surveys. Participant demographic data are summarized in Table 1. The majority of participants were white (85%) and female (86.7%). The participants’ mean age was 25, with range of 19–58 years old. In regard to program of study, 43 participants were in pre-service teacher training programs, 80.1% of whom were in special education and 20.9% in general education preparation programs. The majority of participants knew a person with ASD (80%).

**Study Design**

This was a quasi-experimental study employing a multifactorial four (practices) by two (evidence-based status) by two (source-trustworthiness) design \((4 \times 2 \times 2)\). Two forms of the survey were created. Source trustworthiness (i.e., “source condition”) was manipulated in Forms One and Two by counterbalancing pairings of trustworthy and untrustworthy sources with EBPs and non-EBPs (Fleury et al., 2019). Participants were assigned a form (Form One or Two) based on the first letter of their last name (Form One: A-K, \(n = 26\), Form Two: L-Z, \(n = 34\)). This survey distribution led to two groups with comparable demographics. The survey was distributed using Qualtrics™ software (Qualtrics, Provo, UT), allowing the researchers to randomize the order in which practices were presented across participants to minimize any potential order effects.

**Survey Instrument**

Survey items were drawn from a larger study on the public’s beliefs about practices for people with ASD (Fleury et al., 2019). Two attention checks were embedded into the survey. Participants who failed an attention check

### Table 1

<table>
<thead>
<tr>
<th>Participant Demographics</th>
<th>%</th>
<th>N</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Gender</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>13.3%</td>
<td>8</td>
</tr>
<tr>
<td>Female</td>
<td>86.7%</td>
<td>52</td>
</tr>
<tr>
<td><strong>Race/Ethnicity</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>White</td>
<td>85%</td>
<td>51</td>
</tr>
<tr>
<td>Asian</td>
<td>10%</td>
<td>6</td>
</tr>
<tr>
<td>Hawaiian or Pacific Islander</td>
<td>1.7%</td>
<td>1</td>
</tr>
<tr>
<td>American Indian/Alaska Native</td>
<td>1.7%</td>
<td>1</td>
</tr>
<tr>
<td>Black or African American</td>
<td>1.7%</td>
<td>1</td>
</tr>
<tr>
<td><strong>Highest level of education</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>High school degree</td>
<td>3.3%</td>
<td>2</td>
</tr>
<tr>
<td>Associate’s degree</td>
<td>1.7%</td>
<td>1</td>
</tr>
<tr>
<td>Some college, no degree</td>
<td>48%</td>
<td>29</td>
</tr>
<tr>
<td>Bachelor’s degree</td>
<td>36.7%</td>
<td>22</td>
</tr>
<tr>
<td>Master’s degree</td>
<td>5%</td>
<td>3</td>
</tr>
<tr>
<td>Doctoral or professional degree</td>
<td>5%</td>
<td>3</td>
</tr>
<tr>
<td><strong>Year in program</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>First</td>
<td>15.5%</td>
<td>9</td>
</tr>
<tr>
<td>Second</td>
<td>15.5%</td>
<td>9</td>
</tr>
<tr>
<td>Third</td>
<td>25.9%</td>
<td>15</td>
</tr>
<tr>
<td>Fourth</td>
<td>34.5%</td>
<td>20</td>
</tr>
<tr>
<td>Fifth/Sixth</td>
<td>8.6%</td>
<td>5</td>
</tr>
<tr>
<td><strong>Degree seeking</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Bachelor’s</td>
<td>41.7%</td>
<td>25</td>
</tr>
<tr>
<td>Master’s</td>
<td>48.3%</td>
<td>29</td>
</tr>
<tr>
<td>Doctorate</td>
<td>8.3%</td>
<td>5</td>
</tr>
<tr>
<td><strong>Currently Working in School</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>General Education Teacher</td>
<td>20%</td>
<td>4</td>
</tr>
<tr>
<td>Special Education Teacher</td>
<td>15%</td>
<td>3</td>
</tr>
<tr>
<td>Special Education Assistant</td>
<td>30%</td>
<td>6</td>
</tr>
<tr>
<td>Other</td>
<td>35%</td>
<td>7</td>
</tr>
</tbody>
</table>

*Note.  \(n = 60\).*
were exited from the survey and their data were discarded from analyses. The survey consisted of 112 items in four areas: (a) demographics, (b) causes of ASD, (c) formation of beliefs, and (d) familiarity, approval and likelihood to use or recommend practices. For the purposes of this study, only participants’ demographic information and their familiarity with, approval of, and likelihood to use or recommend practices are described and analyzed.

The survey consisted of descriptions of eight common practices (four EBPs, four non-EBPs) for children with ASD. We used practices identified as EBPs and non-EBPs by the NPDC (Wong et al., 2015) and did not provide explicit information relating to evidence base to participants. Likewise, each practice was described as being presented by one of eight sources (four trustworthy and four untrustworthy as identified by researchers in a recent study, Van Boekel et al., 2017). Practices were presented with a corresponding image and brief paragraph describing how the practice is used and why it is thought to help children with ASD (see Table 2). Each practice was described by either a trustworthy or untrustworthy source. Trustworthy sources included (a) professor, (b) physician, (c) textbook, (d) online collection of peer-reviewed journal (Merk et al., 2017). Untrustworthy sources included (a) magazine, (b) celebrity, (c) blog, and (d) television sitcom. EBPs included (a) Discrete Trial Teaching (DTT), (b) Picture Exchange Communication Systems (PECS), (c) Social Stories™, and (d) Self-Management (Wong et al., 2015). Non-EBPs included (a) Gluten/Casein Free Diet (GFCF), (b) Pressure/Weighted Vests, (c) Brushing, and (d) Chelation. Table 2 provides each practice description along with the counterbalanced pairings of practices with sources in the two versions of the survey. Each practice description was designed to include enough information for the participant to respond to questions regardless of whether or not they had prior knowledge of or experience with a practice.

After the presentation of each practice, participants were asked to respond to the following statements: (a) I am familiar with (practice) (dichotomous, yes/no), (b) I approve of (practice) as a practice for individuals with ASD (5-point rating scale anchored with strongly disagree to strongly agree), and (c) If applicable, I would use or recommend to a friend to use (practice) as a practice for individuals with ASD (5-point rating scale anchored with strongly disagree to strongly agree).

Questions pertaining to both personal and professional demographic information were asked at the end of the survey. In addition to expected demographic questions describing participants, questions relating to participants’ decision-making context included their highest educational level (high school, associate’s degree, bachelor’s degree, graduate degree or higher) and if the participant knows someone with ASD (dichotomous yes/no).

Data Analysis

We used SPSS (Version 24.0) software to perform all statistical analyses. To analyze relations with source trustworthiness, responses from both forms were combined by practice with a “source condition” variable that indicated which form each participant completed (Form One or Form Two) in order to evaluate the source and practice pairings. We calculated response frequencies for demographic information, familiarity, approval, and likelihood to use or recommend practice. Percentage of the sample familiar with each practice was also computed. Due to the ordinal nature of the data for participants’ ratings of approval and their likelihood to use or recommend a practice, we ran Fisher’s Exact Test (Kim, 2017) to evaluate pairwise relations. Fisher’s Exact Test (FE) uses contingency tables from small samples (n<30), and includes no assumptions regarding the distributions (Kim, 2017). Using this method for each practice we performed pairwise analyses of (a) familiarity with the practice, (b) approval of the practice, (c) likelihood to use or recommend the practice, (d) participant’s education level, (e) participant’s knowing someone with ASD, and (f) source condition.

To evaluate if there were differences between participants’ overall ratings of approval and likelihood to use or recommend EBPs compared to non-EBPs, we calculated mean ratings of agreement for approval and likelihood to use/recommend each practice by tak-
### TABLE 2

<table>
<thead>
<tr>
<th>Description and Form 1 Source</th>
<th>Form 2 Source</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Evidence-Based Practices</strong></td>
<td></td>
</tr>
<tr>
<td>An online collection of peer-reviewed journal articles converged on discrete trial teaching (DTT) as a one-to-one instructional approach used to teach skills in a planned, controlled, and systematic manner. This strategy involves breaking down a complex task into smaller skills. DTT is characterized by repeated, or massed trials to ensure that the individual with ASD has lots of opportunities to practice and learn the target skill. The instructional trial begins when the adult presents a clear direction to elicit a target behavior. Positive praise and/or tangible rewards are used to reinforce desired skills or behaviors. The online collection of peer-reviewed journal articles described that data collection is an important part of DTT as it provides teachers/practitioners with information about beginning skill level, progress and challenges, the extent to which the individual has learned the skill, and whether the learner can use the skill in a variety of settings (e.g., classroom, home, community).</td>
<td>Blog</td>
</tr>
<tr>
<td>A professor specializing in communication sciences explained that the Picture Exchange Communication System (PECS) is a technique used to teach individuals who are unable to speak to communicate using picture symbols in a social context. Using PECS, learners are initially taught to give a picture a communication partner picture symbol(s) that represents what they would like to communicate. The initial stages of PECS involve teaching the individual how to request for basic needs (i.e., desired food item or activity). In later phases, the individual is taught to communicate for other social purposes (i.e., asking questions, making comments). The professor described that there are six phases of PECS instruction: (1) “how” to communicate, (2) distance and persistence, (3) picture discrimination, (4) sentence structure, (5) responsive requesting, and (6) commenting.</td>
<td>Sitcom</td>
</tr>
<tr>
<td>A blog explained that social narratives (SN) are interventions that describe social situations in some detail by highlighting relevant cues and offering examples of appropriate responding. Individuals with ASD often have difficulty understanding what behaviors are expected of them in different social situations. Social narratives can help individuals with ASD by establishing clear expectations of how they should behave in a particular situation. Usually written in first person from the perspective of the learner, social narratives include sentences that detail the situation, provide suggestions for appropriate learner responses, and describe the thoughts and feelings of other people involved in the situation. The blog went on to describe that social narratives are individualized according to learner needs and typically are quite short and may include pictures or other visual aids.</td>
<td>Online collection of peer-reviewed journal articles</td>
</tr>
<tr>
<td>A television sitcom focused on self-management (SM) as an intervention package that teaches learners to independently regulate their own behavior. Self-management involves teaching learners to discriminate between appropriate and inappropriate behaviors, accurately monitor and record their own behaviors, and reinforce themselves for behaving appropriately. The television sitcom demonstrated that although learners may initially require adult support to accurately record behaviors and provide self-reinforcement, this support is faded over time.</td>
<td>Professor specializing in communication sciences</td>
</tr>
</tbody>
</table>

(Continued on next page)
TABLE 2
(Continued)

<table>
<thead>
<tr>
<th>Description and Form 1 Source</th>
<th>Form 2 Source</th>
</tr>
</thead>
</table>
| **Non-Evidence-Based Practices**
  A textbook explained that **Gluten-free (GF), Casein-free (CF), and Gluten-Casein Free (GF/CF)** exclude foods containing gluten or casein. It has been suggested that individuals with autism spectrum disorders often have food intolerance, allergies, and sensitivities related to both Gluten and Casein, and that often these intolerances, allergies, and sensitivities will exacerbate symptoms of ASD. Adopting a GF/CF diet, may show improvements with their symptoms over time. Foods containing Gluten: pastas, other noodles, breads, crackers, baked goods, cereals, breakfast foods containing wheat, croutons, sauces and gravies, flour tortillas, and others. The textbook also described foods containing Casein: milk, yogurt, sour cream, cheese, butter, ice cream, puddings, and others.

  A **magazine** explained that **weighted/pressure vests** are therapeutic garments designed to be worn by children on their torsos. Typically, these vests are recommended for children who are hyperactive, have difficulty focusing and concentrating, or are oversensitive to environmental stimuli. Weighted vests can range between 2% to 10% of the child’s body weight, depending on age, and are designed to have weight distributed evenly around the child’s body. Pressure vests are tightly wrapped around the child’s body to provide consistent, even pressure around the child’s body. The physician described that the specified amount of time the vests are worn depends on the child’s age and sensory needs.

  A **celebrity** explained that **brushing techniques** are programs that involve brushing an individual’s extremities and core, excluding the face, chest, and stomach, with a soft, plastic brush using firm pressure. These techniques are used with individuals that display fear or resistance to touch, have difficulty with transition, or are often lethargic. The brushing sessions, usually done every two hours, help to relax the nervous system and reduce the symptoms of sensory defensiveness and anxiety, or activate the nervous system and reduce lethargy, depending on the individual. The celebrity described that brushing is often paired with a course of joint compressions, which are light compressions on the joints of the body done rapidly, to contribute to the deep pressure aspect of the overall program.

  A **textbook** explained that many of the behavioral symptoms associated with ASD resemble symptoms of mercury poisoning. **Chelation** is the medical process of removing excessive toxic metals, such as mercury, from the body using chemicals that bind with the metals (called chelators). Once the chelators bind to the toxic metal, the body is able to excrete it from the body. There are several substances that can be used as chelators, both natural (i.e., cilantro) and synthetic (i.e., DMSA, lipoic acid). The recommended chelation therapy includes the following steps: (1) stopping any ongoing exposure by eliminating fish or seafood from the diet, replacing amalgam tooth fillings with white composite material, and using only thimerosal-free vaccines; (2) removing loosely-bound body mercury; (3) getting rid of more tightly-bound mercury; (4) appropriate nutritional support that includes the use of antioxidants; (5) finally, the magazine concluded that close monitoring of blood count, liver and kidney function, and mineral levels is needed to allow the doctor to gauge how much mercury is being executed.

*Note.* Bold emphasis (practices), underlining (trustworthy sources), and italics (untrustworthy sources) were added for distinction for this publication.
ing the mean of participant’s ratings for each respective set of the four EBPs and the four non-EBPs. The distributions of these mean ratings by evidence-based status did not meet assumptions for paired t-tests, thus differences could not be explored.

Results

Participants’ Familiarity, Approval, and Likelihood to Use or Recommend Practices

Table 3 depicts the frequencies and percentages of responses in each category for participants’ approval of each practice and familiarity with each practice. Participants were most often familiar with GFCF Diet (88% familiar), and least often familiar with Chelation (32% familiar). For EBPs, the distributions of approval responses were skewed, with more participants agreeing or strongly agreeing than disagreeing or strongly disagreeing, with the exception of DTT. The majority of participants strongly agreed that they approved of PECS (55%) and Social Narratives (55%) with no participants strongly disagreeing. Similarly, the plurality of participants responded that they agreed (40%) and none strongly disagreed that they approved of Self-Management. For DTT, the responses were more normally distributed, with the plurality of respondents neither agreeing nor disagreeing (36.7%) that they approved of the practice. For non-EBPs, the plurality for Brushing (43.3%), GFCF Diet (45%), and Pressure/Weighted Vests (35%) neither agreed nor disagreed that they approved of the practice and the distribution of responses were more normal. Chelation differed; the plurality of participants strongly disagreed that they were likely to use or recommend it (46.7%), none strongly agreed, and one agreed. Overall, participants more often indicated they agreed or strongly agreed that they were likely to use or recommend practices that were evidence-based than those that were not.

Relations between Decision-Making Context, Approval, and Likelihood to Use or Recommend Practices

Figure 2 depicts the results of pairwise analyses from Fisher’s Exact Test indicating which factors (knowing someone with ASD, education level, and source-trustworthiness) were associated with participants’ familiarity, approval and likelihood to use or recommend practices. The findings are mixed across practices, regardless of evidence-based status.

Personal experience and familiarity. Education level was significantly related to participants’ familiarity with PECS (p < .05, FE), DTT (p < .001, FE), and Weighted Vests (p < .05, FE) such that the majority of people who were familiar with each of these practices held a bachelor’s degree. Participants who knew someone with ASD significantly more often disagreed that they approved of Chelation (p < .05, FE) and more often neither agreed nor disagreed that they would use or recommend GFCF diet (p < .05, FE). Knowing someone with ASD was not related to any other factors. A participants’ familiarity with a practice was also found to be related to approval and likelihood to use or recommend the practice. Familiarity with the practice was significantly associated with approval of DTT (p < .001, FE) and PECS (p < .05, FE). Familiarity was more often significantly related to a participant’s likelihood to use or recommend the practice for EBPs; DTT (p < .05, FE), PECS (p < .05, FE).
## TABLE 3
Frequency of responses to Familiarity, Approval and Likelihood to Recommend Practices

<table>
<thead>
<tr>
<th></th>
<th>Familiar</th>
<th>Approve</th>
<th>Likely to Use or Recommend</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>Strongly Disagree</td>
<td>Disagree</td>
</tr>
<tr>
<td></td>
<td>n %</td>
<td>n %</td>
<td>n %</td>
</tr>
<tr>
<td><strong>EBP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PECS</td>
<td>46</td>
<td>77</td>
<td>0</td>
</tr>
<tr>
<td>DTT</td>
<td>35</td>
<td>58</td>
<td>1</td>
</tr>
<tr>
<td>Self-Management</td>
<td>16</td>
<td>27</td>
<td>0</td>
</tr>
<tr>
<td>Social Narratives</td>
<td>52</td>
<td>87</td>
<td>0</td>
</tr>
<tr>
<td><strong>Non-EBP</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Brushing</td>
<td>27</td>
<td>45</td>
<td>5</td>
</tr>
<tr>
<td>GFCF Diet</td>
<td>53</td>
<td>88</td>
<td>11</td>
</tr>
<tr>
<td>Chelation</td>
<td>19</td>
<td>32</td>
<td>24</td>
</tr>
<tr>
<td>Pressure/Weighted Vests</td>
<td>48</td>
<td>80</td>
<td>5</td>
</tr>
</tbody>
</table>

Note. $n = 60$, EBP = evidence-based practice, PECS = Picture Exchange Communication System, DTT = Discrete Trial Teaching, GFCF = Gluten Free/Casein Free Diet
Social Narratives ($p < .05$, $FE$), than for non-EBPs (Chelation, $p < .05$, $FE$).

Source trustworthiness. The source of treatment information did not significantly relate to PSP’s decisions to use or recommend most practices. DTT is the only practice for which source-trustworthiness was significantly associated with any other factors; participants were...
more likely to use or recommend DTT when it was presented by a blog rather than a peer-reviewed journal article \((p < .05, FE)\).

**Approval.** In addition to the previously described factors, participants’ likelihood to use or recommend a practice was significantly related to approval for all EBPs; DTT \((p < .001, FE)\), PECS \((p < .05, FE)\), Social Narratives \((p < .001, FE)\), and Self-Management \((p < .001, FE)\), and most non-EBPs; Chelation \((p < .001, FE)\), GFCF Diet \((p < .001, FE)\), and Pressure/Weighted Vests \((p < .001, FE)\).

**Discussion**

Though EBPs have been identified and targeted for broad dissemination (NAC, 2009; Wong et al., 2015), some researchers suggest that the lack of full scale implementation of EBPs may be attributed to a lack of familiarity educators have at the onset of their careers given a lack of information being disseminated through preparation programs (Morrier et al., 2011; Rosenberg et al., 2004). Given our findings that the majority of PSPs were, by and large, familiar with EBPs and non-EBPs, efforts to enhance the dissemination of information in preparation programs to include EBPs may oversimplify the implementation problem that exists.

**Implications for Preparation Programs**

We offer an initial framework for understanding what factors may relate to PSPs’ approval and likelihood of recommending use of EBPs and non-EBPs. This framework is predicated on the notion that if a PSP values certain EBPs and establishes an intention to recommend use of those EBPs, they may be more likely to actually use and recommend EBPs to others once working in the field (Montano & Kasprzyk, 2015). The means by which approval and intent to use an EBP are facilitated within a preparation program becomes particularly salient. Consistent with demonstrations that the source of the information being provided relates to practitioners’ perceptions about information (Landrum et al., 2002; Merk et al., 2017; Metz et al., 2015), we explored the relation between source–trustworthiness and PSPs’ approval of and likelihood to adopt practices. Our findings related to Discrete Trial Teaching (DTT) highlight opportunities that may exist within preparation programs to facilitate an intention to implement a practice as well as challenges that may exist for enacting implementation once a PSP completes their program and enters the field.

PSPs more often reported they were likely to use or recommend DTT if the source presenting information was a blog rather than a peer-reviewed journal, which is a source that may not match practitioners’ values of (Cook et al., 2013; Johnson et al., 2018). Specific to this study, PSPs may have been more susceptible to source influence for DTT because they were less familiar with (58% familiar) and neutral in their approval ratings of this practice (i.e., the fewest respondents agreed that they approved of DTT). This suggests that the PSPs in this study may not have had existing knowledge about DTT. In addition, the observed influence of source may reflect differences in PSPs’ perceptions of the complexity of DTT relative to the source given an absence of familiarity. This interpretation is consistent with assertions by researchers that practices may seem less complex and more personal to a reader when presented through popular media, such as a blog (Metz et al., 2015). In light of these findings, preparation programs may find that assessing PSPs’ familiarity with practices and disseminating research-based information through media that match PSPs’ values may facilitate a greater intention to implement EBPs. Future research exploring methods for tailoring sources to the characteristics of PSPs is warranted, particularly when considering the induction of PSPs into a field in which implementation of EBPs is limited.

The challenges for enacting an intention to implement an EBP once a PSP enters the field reflects increasingly complex and dynamic interactions between the individual implementer and the context in which they make decisions and take actions. Knight et al. (2018) found that educators in the field reported that their use of practices was rarely influenced by an expert or research, so future efforts to identify what and how information should be shared will be pivotal in increasing implementation of EBPs. As suggested by Rogers (2005) and echoed by Schreck and Mazur (2008), information promotes adoption and
use when practices are perceived to offer a relative advantage, be compatible, have a complexity that is within the skills of the implementer, and allow for trialability and observability such that an implementer may evaluate aspects of the practice prior to full implementation. Unfortunately, presenting information that promotes adoption and use may be effective irrespective of the evidentiary base for the actual practices. Given that proponents of non-EBPs must rely on dissemination forms and sources that may be more aligned with sources that educators in the field find more informative for their practice, use of non-EBPs by practitioners is widespread.

Using weighted vests, one method of sensory integration, is one such non-EBP that is touted as a “quick-fix” for students with ASD based on pseudoscientific claims (Barton et al., 2015). Disseminated information about weighted vests may make them appear less complex than EBPs that have multi-step implementation procedures (Metz et al., 2015; Travers, 2017). Given the low response effort required on their part to implement the practice, practitioners may perceive a relative advantage of this non-EBP and may therefore persist with, and recommend to others, use of this practice for providing a relative advantage to students. To counteract this and ensure students truly do experience a relative advantage, researchers and preparation programs should seize on other dissemination sources to reduce perceptions of complexity, perhaps through observability and trialability, such that practitioners’ stated interest in doing what is effective for students (Knight et al., 2018) may be enacted through use of EBPs. Further, PSPs in this study indicated differences in their ratings of approval and recommendations for use of non-EBPs, including weighted vests specifically, when compared to EBPs. This suggests that PSPs were discerning between practice types, with a tendency towards greater approval and use of EBPs. This may contrast the findings for in-service practitioners for whom Hudson et al., (2016) reported that practitioners were ambivalent about research and may rely on other factors to inform their adoption of practices, such as professional judgement (Knight et al., 2018) or recommendations from a colleague they believe holds high expertise (Merk et al., 2017). Preventing ambivalence and strengthening reliance on research-based information may be an important aspect of preparation programs.

**Implications for Future Research and Limitations**

Research findings showing that practitioners more often rely on their own professional opinion and do not find published research helpful (Hudson et al., 2016; Knight et al., 2018) contrast many current efforts to enhance implementation of EBPs (Wong et al., 2015). It may be the case though, that researchers also need to reconsider their own presentations of their work. In a review of studies that helped to establish several practices as evidence based for young children with ASD, Johnson and colleagues (2018) found that though the reporting practices of researchers met the quality indicators for research, the information that is needed to promote implementation by practitioners was largely absent from published studies. From that perspective, it is not surprising that in-service practitioners are not getting the information they need and find valuable from scientific work presented in research journals. Within preparation programs, PSPs are supported through the translational process of reading a study and planning for implementation in their future classrooms by skilled instructors. That type of implementation support (Powell et al., 2015) rarely exists for in-service practitioners.

Preparation programs may be positioned to offer the greatest impact on implementation of EBPs as part of educators’ routine practice by exploring new ways to strengthen PSPs’ familiarity with, approval of, and intention to implement EBPs upon completion of their preparation. Though the findings presented in this study may be unique to the specific preparation program, the personal and professional experiences of these PSPs combined with trustworthy sources’ (e.g., professors) dissemination of information about EBPs within university coursework such that intentions to use and recommend EBPs outnumbered non-EBPs. Given findings from other studies identifying a lack of information and training on EBPs within preparation programs (Hsiao & Peterson, 2018; Rosenberg et al., 2004), these
programs should evaluate the degree to which the individuals who are responsible for facilitating experience and disseminating information are themselves accessing research-based information about practices and serving as a trustworthy source for their own PSPs. Further, evaluating the dissemination sources that are utilized and the degree to which those sources are functioning as determinants of PSPs’ intentions to implement EBPs is warranted.

This survey was designed for the general public and piloted with a convenience sample of PSPs for this study. As such, there are some limitations as well as opportunities for future research that maintains a singular focus on the preparation of pre-service practitioners. First, this sample included participants enrolled at different points in their programs and in different programs (e.g., master’s, bachelors, general education, special education), so variability in their knowledge base may have been present, but was outside of the scope of this study. Second, this study included both educational and medical practices. In future studies, including only those practices that are appropriate for implementation in educational environments may offer greater insights. In the current study, retention of the medically based practices was perceived to be acceptable given that the future work of the PSPs in this study may include counseling families of children with ASD about practices that may be adopted and implemented by the family rather than directly by the practitioner. Next, this study relied on an evaluation of source influence given accepted identifications of certain sources as trustworthy versus untrustworthy (Van Boekel et al., 2017). With that approach, the individual who is the source of the information and the form through which information is disseminated are commingled. Future work that focuses exclusively on practitioners should not only separate those factors, but also include practitioner-sources (e.g., a special educator) so as to establish who should be providing information about certain types of practices (Carrington et al., 2016; Landrum et al., 2002; Merk et al., 2017). Finally, research suggests there may be additional factors that influence both how practitioners establish an intention to adopt a practice and how intention is turned into actual implementation. Expanding this survey to evaluate implementation supports (Powell et al., 2015) may prove useful to identifying new means of building experience and knowledge that may be incorporated in to, (a) pre-service preparation when knowledge and beliefs are still malleable, (b) induction systems that support PSPs as they transition into in-service roles, and (c) the ongoing professional work of educators. With a more comprehensive understanding of how early belief systems develop for PSPs around certain practices, personalized professional development systems may be designed for educators to promote sustained and routine use of effective practices.

References


Received: 13 March 2019
Initial Acceptance: 16 May 2019
Final Acceptance: 21 August 2019
Increasing the Provision of Choices within an Adult Transition Program

Rachel J. Schwartz, Rachel E. Robertson, and Sarah Westerfield
University of Pittsburgh

Abstract: Opportunities for choice often become limited as individuals with disabilities age. Previous research reports that staff practices directly influence choice-making opportunities for adults with disabilities. Given the link between choice-making and staff behavior, more research is needed on training staff in postsecondary settings in providing choice-making opportunities. The present study trained staff in a community transition program to facilitate diverse choice opportunities for consumers. Researchers conducted a multiple-baseline-across-participants single-subject design to evaluate the effects of a staff-training package on the frequency and diversity of choices offered by staff to consumers. Findings show that staff increased the frequency and diversity of choices offered. The potential implications of this training package for special education are discussed.

The opportunity to make independent choices marks the transition from adolescence into adulthood. Choice-making forms the core of self-determination, a construct which promotes individuals as causal agents in their lives rather than dependent on others to direct or control their actions (Shogren et al., 2004; Wehmeyer & Abery, 2013). The value of choice, notably the independence and identity affirmation associated with it, has reportedly led to greater self-determination and participation with the broader community (Agran et al., 2010; Wehmeyer & Abery, 2013). Adulthood symbolizes autonomy to make self-directed choices, yet prior findings demonstrate that adults with intellectual and/or developmental disabilities (I/DD) often fail to access choices in even minor aspects of their lives (Heller et al., 2011; Neely-Barnes et al., 2008b).

Since the rise of normalization, postsecondary services have largely embraced the concept of choice within their service models (Neely-Barnes et al., 2008a). These models traditionally include consumer participation in choice during person-centered planning and annual goal setting, but struggle to incorporate choice during general, everyday decisions (Brown & Brown, 2009). Consumers of adult disability services often lack opportunities to exert choice in even minor aspects of their day to day living such as what to eat or wear, and their lives often become determined by others around them—namely, staff and service providers (Agran et al., 2010).

Staff behaviors and perceptions of consumer needs can directly encourage or impede consumer-directed choice-making (Agran et al., 2010; Zakrjasek et al., 2014). During an investigation to determine vocational interests of adults with disabilities, Cobigo et al. (2010) compared staff perceptions of consumer job preferences with jobs ultimately chosen by consumers. Findings revealed a misalliance between choices hypothesized by staff and choices made by workers. These results support similar research findings that suggest reliance on a proxy to make consumer choices often yields inaccurate results (Agran et al., 2010; Smith et al., 2005). In order to shift control from service providers to adults with disabilities, disability organizations must recognize the relationship between staff practices and availability of consumer choices.

This study was completed with a grant from the Organization for Autism Research and a University of Pittsburgh School of Education Grant. Correspondence concerning this article should be addressed to Rachel J. Schwartz, The Watson Institute, 301 Camp Meeting Road, Sewickley, PA 15143. E-mail: Rachells@thewatsoninstitute.org
ing of choice and the importance of consumer-directed choices. For example, Parsons et al. (1997) examined staff training in choice procedures to increase the number of choices and, to a lesser degree, the type of choices. Researchers provided staff members in a residential facility with a description of choice procedures, engaged in role-plays with staff, and provided performance feedback. Findings resulted in an increase in the number of choices staff members offered, and indicated a slight increase in choice diversity across three choice types. Salmento and Bambara (2000) also trained staff in a residential facility to offer single-stimulus choices to adults with profound ID during routines of daily living. Staff training included a rationale for choice, how to identify choice opportunities during routines, a specific choice-offering sequence, and performance feedback. Similarly, Reid et al. (2003) successfully trained job coaches in a vocational setting to offer on-the-job choices. Reid and colleagues implemented comparable training components to Parsons et al. (1997) and Salmento and Bambara (2000) including a rationale, description, demonstration of choice procedures, and performance feedback with equally positive results.

Prior research in staff training produced encouraging outcomes and affirmed the importance of choice in postsecondary settings. However, the definition and measurement of choice within staff trainings requires further examination. Previous research primarily defined choice as a single-stimulus or paired-stimulus presentation of options and measured the effects of training on the frequency of choices provided by staff (Cooper & Browder, 2001; Parsons et al., 1997; Wilson et al., 2006). Although increasing the frequency of choices presumably generates more opportunities for choice-making, reducing choice to a frequency measure alone potentially minimizes the importance of diversity in choice type and corresponding choice content (Stalker & Harris, 1998). Numerous researchers exploring the construct of choice have discussed the importance of providing diverse choice opportunities (e.g., Agran et al., 2010). Yet, the majority of staff training literature distills choice to a quantitative measure when investigating the quality of the choice seems equally important.

Although prior literature demonstrates increases in choice offerings, researchers must continue to broaden the definition and measurement of choice by incorporating frequent opportunities for diverse types of choices. Through the present study, researchers attempted to extend choice literature by examining whether a staff training package in choice-making would (a) increase frequency of choices staff members offered consumers using a designated choice sequence, (b) decrease frequency of incorrect choices staff members offered consumers, and (c) increase the diversity of choices staff members offered consumers.

Method

Participants, Setting, and Materials

Participants. A total of three staff members employed as direct-care staff consented to participate in this research. Researchers recruited participants who worked full-time in a disability organization serving adults with disabilities, and had not received a previous staff training relating to choice. The director of the program identified potential staff members who met the inclusion requirements, and the primary researcher approached each of them for potential participation. One participant dropped out after one week due to an inconsistent work schedule, and the research team added another staff member to the study. Pseudonyms were used to protect participant identity. Kamal was a 56-year-old man from Northern Africa who worked with the present disability organization for over 13 years. Blair was a 29-year-old Caucasian female who had worked for the community program for three years. Prior to her experience in the present organization, Blair worked for one year at a residential home for adults with mental illness. Cathy was a Caucasian female in her late 50s who had been employed with the community program for three years. She did not have prior experience working with disabilities, but does have a niece with a disability. Cathy joined the study a few days later than the other participants due to another staff member dropping out prior to data collection. None of the participants had received any prior training in providing choices, identifying choice opportunities, or choice diversity.
Setting. The community transition program was located within a larger organization providing rehabilitation and vocational services to individuals with vision loss. In addition to vision services, this organization housed a vocational program and a community transition program on different floors of the main building. The current study took place in the area of the community transition program. The transition program occupied one large room that included a kitchen, living room area with couches and a television, and a series of five long wooden tables with chairs. Staff and consumers shared locker space near the entrance of the program. One wall held a row of cupboards that contained various vocational and leisure activities including beading, puzzles, block stackers, and fine motor activities. The opposite wall contained art supplies and a series of in-progress or completed art projects. A long hallway ran through the middle of the room and connected the main program space to three bathrooms, a small sensory room past the kitchen, and a staff office. The daily schedule of the community program consisted of a community outing, cooking, and a mix of vocational and leisure tasks. These tasks included stringing beads, putting pennies in a jar, and puzzles. Staff and a small group of consumers left the program between 9:30 and 10:00 a.m. for community outings and returned during or after lunch, approximately around 12:00 p.m. During the study, staff members worked with their assigned consumers in the transition program but were not always in the program at the same time. Researchers did not collect data during any times when staff were working at the same or nearby tables, or completing the same activity (i.e. mealtime). This protocol guarded against potential history threats.

Fourteen adults with intellectual and developmental disabilities (I/DD) attended the community transition program. Five of the consumers had co-morbid diagnosis of vision loss and ID and four had a co-morbid diagnosis of autism spectrum disorders and ID. Consumers ranged in ages from 22 to 74 years old with a mean age of 44 years old. All participants either lived at a relative’s home (N = 3) or in various residential programs (N = 11).

Materials. A Sony ICD PX333 digital voice recorder and Sony ECMCS3 omnidirectional stereo microphone were used to audio record each observation session. To ensure that each observation lasted strictly 30 minutes, researchers used the timer function on an iPhone6. Individual data were collected using a data sheet of the choice sequence and a pen or pencil.

Experimental Design

Researchers implemented a multiple-baseline-across-participants single-subject design to evaluate the effects of staff training and feedback (Wolery et al., 2010). Initial observations of staff occurred during baseline. Once the first staff member reached a stable baseline, researchers introduced the choice training program and began performance feedback for the first staff member while the other participants remained in baseline. Researchers introduced the choice training sequentially across staff, therefore training for each staff member began after the previous staff reached criteria on correct choice opportunities during the feedback phase.

Response Definitions and Data Collection

The research team collected data during each condition on (a) frequency of correct choices offered by staff to consumers, (b) frequency of incorrect choices, and (c) frequency of choice diversity (e.g., different choice types). Researchers identified a correct choice as participants following each step of the choice sequence described in Table 1. If staff skipped or did not correctly complete one or more steps, the entire choice was marked as incorrect. Frequency (number) was used as the dependent measure because choice represents a potentially limitless action. In other words, choice does not have a ceiling for the amount that staff should offer within a single activity or time frame. By capturing the frequency of correct and incorrect choice offerings, the data represents the total number of choices attempts and if staff improved in their implementation of a choice sequence.

For this study, researchers conceptualized choice as providing verbal options for any single action (Brown et al., 1993; Stancliffe, 2001). Researchers only counted choices if the choices were appropriate (i.e. not harmful to consum-
TABLE 1
Checklist of Correct Choice Sequence

<table>
<thead>
<tr>
<th>Choice Sequence</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gain attention</td>
<td>Staff member makes contact with consumer in a way such that the consumer orients towards the staff member (e.g., eye contact, body movement, vocal response).</td>
</tr>
<tr>
<td>Present choice options</td>
<td>Presents choice options vocally, visually, through tactile means, or a combination of methods.</td>
</tr>
<tr>
<td>Wait at least 3s</td>
<td>Staff member waits at least three seconds before moving on. If consumer chooses within this time, move to next step.</td>
</tr>
<tr>
<td>If consumer chooses, immediately give desired stimulus within 5s</td>
<td>Staff member provides consumer with choice response or honors the choice within a 5 second latency. If the choice is an activity, immediately start preparing.</td>
</tr>
<tr>
<td>If consumer refuses the options provided, staff will:</td>
<td></td>
</tr>
<tr>
<td>● Make a choice for consumer if item is necessary for health or societal reasons.</td>
<td></td>
</tr>
<tr>
<td>● If not applicable, move on in activity.</td>
<td></td>
</tr>
</tbody>
</table>

Inter-Observer Agreement and Procedural Fidelity

Training. The primary investigator trained a doctoral student as a secondary coder to identify a correct choice sequence and to differentiate between choice types. The secondary coder scored sample audio recordings for following the correct choice sequence and identification of choice types. The primary investigator and the secondary coder calculated their IOA on this sample data until they scored 80% reliability (range: 86–100%).

IOA and procedural fidelity. The research team collected interobserver agreement (IOA) and procedural fidelity data for 33% of all sessions within each phase of the study. Two members of the research team independently transcribed choice offerings and collected data on staff performance of the choice sequence. The primary investigator (PI) calculated IOA on the percentage of agreement on the type of choice offered and whether staff members completed the choice sequence correctly. An agreement was defined as when both observers agree that a behavior did or did not occur during an observation. A disagreement was defined as when the observers did not agree on the occurrence of a behavior during an observation. The mean IOA calculated totaled 92% (range: 82–100%). During coaching and feedback sessions, the second-
ary coder collected procedural fidelity on the following behaviors a) coaching staff members to provide a choice opportunity, b) providing corrective and positive feedback at the end of observation, and c) not directly modeling choice opportunities with consumers. A secondary observer collected procedural fidelity measures and calculated fidelity by dividing the number of correctly implemented behaviors by the researcher by the number of planned behaviors multiplied by 100. Researchers calculated mean procedural fidelity to be 92% (range: 80% – 100%).

**General Procedure**

**Baseline.** Data collection occurred approximately twice per week per participant (range two to four observations) for a minimum of at least five observations. Each observation lasted 30 minutes and occurred during the individual morning sessions (8:00 am to 12:00 pm) since afternoon sessions consisted of group activities. Researchers counterbalanced observation times across participants to ensure that they observed participants equally across different times of the morning sessions. Morning sessions always involved leisure activities and tasks that were available in the observation setting (see above in “Setting”). The availability of activities remained consistent throughout the study. During baseline sessions, staff completed their daily routine as usual without any direction from the researchers. This phase replicated what typically occurred in the environment. Researchers gave participants a pocket recorder connected to a microphone and instructed each participant to engage in their typical daily schedule. Baseline continued until the first staff member’s behavior reached stable responding. At this time, training began in a staggered fashion for the first staff member while the other participants remained in baseline.

**Skills training.** Each staff received 1:1 training in choice by the PI in a small, private room located within the community program. The PI implemented Behavior Skills Training (BST) during the choice training with staff (Parsons et al., 2012). An evidence-based prac-

<table>
<thead>
<tr>
<th>Choice Type</th>
<th>Definition</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Within</td>
<td>A choice that functions to provide different materials within one activity. Speaker responses are typically specific items.</td>
<td>“Which color of paper do you want?”</td>
</tr>
<tr>
<td>Between</td>
<td>A choice that functions to provide options between two or more activities. Activities could be present in the environment or not. Speaker responses are typically specific activities.</td>
<td>“What do you want to do today?”</td>
</tr>
<tr>
<td>Refusal</td>
<td>A choice that provides an option to refuse what is being offered. Speaker responses limited to a “Yes” or “No”.</td>
<td>“Do you want a drink of water?”</td>
</tr>
<tr>
<td>Termination</td>
<td>A choice that functions to end a current activity. Speaker responses typically a “Yes” or “No”.</td>
<td>“Are you finished?”</td>
</tr>
<tr>
<td>Who</td>
<td>A choice that functions to provide options for persons involved in an activity. Speaker responses may be a specific names or a “Yes” or “No”.</td>
<td>“Who do you want to work with today?”</td>
</tr>
<tr>
<td>Where</td>
<td>A choice that functions to provide options of locations for an activity. Speaker responses typically name a specific location.</td>
<td>“Do you want to eat lunch outside or inside?”</td>
</tr>
<tr>
<td>When</td>
<td>A choice that functions to provide options for the timing of an activity. Speaker responses may include a specific time, delay of time (e.g., in 5 minutes), or a “Yes” or “No”.</td>
<td>“Do you want to work on the art project now or after lunch?” “When do you want to watch TV?”</td>
</tr>
</tbody>
</table>
tice, BST promotes the performance and competency of a specific skill by direct care staff (Ward-Horner & Sturmey, 2012). BST protocol consists of a rationale, written description, demonstration, practice, and feedback with practice and feedback continuing until the trainee demonstrates a set level of competency (Parsons et al., 2012). The current training included a) descriptions and demonstrations of choice, choice-making sequence and choice diversity, b) modeling of the choice-making sequence and choice types, c) role-plays, and d) performance feedback. Feedback included a) what components of the choice sequence the staff member did well, b) on what steps the staff member erred, c) the different types of choices the staff member offered, and, d) considerations for how to offer other choice types. A PowerPoint presentation outlining each component guided the training and researchers responded to any questions during the training with further explanations, examples, and demonstrations. Each training session lasted approximately 45 minutes.

Similar to the training utilized in Salmento and Bambara’s study (2000), researchers taught staff a sequence for offering and responding to choices (see Table 1). This sequence was adapted from previous research (Salmento & Bambara, 2000; Sigafoos & Dempsey, 1992; Sigafoos et al., 1993) and modified to align with Brown and associates’ Model of Choice Diversity (1993) (see Table 2).

Performance feedback and coaching. After completing the training and demonstrating the choice sequence during role-plays, staff returned to their typical shift and implemented the choice sequence to consumers as described during training. Prior to beginning the first three observations post-training, the researchers provided a brief reminder of the training components by saying “In the training, we discussed the importance of choice and offering a range of choices and choice types.” The researchers then coached participants as well as provided specific positive and corrective feedback following every observation throughout intervention. Staff feedback mirrored feedback structure and content from training. During the observation, the researcher provided coaching to staff members if the inter-response time between the last choice offerings was at least 5 min or if staff did not offer a choice within the first 5 min of an observation session. Coaching included first asking staff to consider potential choice opportunities during the current activity (e.g., “Bob seems to be sitting alone on the coach, are there any choices you could offer him right now?”) and how the staff could turn typical commands or questions into choice opportunities (e.g., “Instead of saying “Go to the bathroom”, could you turn that into a choice?”). If the staff member did not respond, voiced confusion, or did not attempt a choice offering, the researchers then provided an example of a potential choice to offer.

For example, Kamal usually gave a consumer a glass of lemonade without the consumer requesting it. When coached, Kamal said he knew the consumer wanted it because he always had a glass of lemonade in the morning. The researcher asked Kamal if he thought that there was a potential choice involved in offering a drink. After Kamal stated he didn’t believe a choice was needed, the researcher provided examples of offering the consumer a choice to refuse a drink (e.g., “Do you want a drink?”) and/or a choice between available drink options (e.g., “What do you want to drink?”; “Do you want lemonade, water, coffee, or hot chocolate?”). Kamal then crafted his own choice offering to present to the consumer (“Do you want some lemonade?”). Researchers continued to provide coaching during sessions and feedback after sessions ended until staff members reached criterion. Criterion involved staff achieving above their highest baseline level in the frequency of correct choices they offered consumers for three consecutive observations. Researchers chose this criterion based off of prior research (e.g., Reid et al., 2003; Salmento & Bambara, 2000). Once staff reached criterion, they moved into maintenance and the next staff participant began the initial skills training.

Maintenance and follow-up. After the feedback phase, researchers continued to collect data on choice offerings. This phase replicated baseline procedures. If at any time staff fell at or below baseline levels in the frequency of correct choice offerings, the researchers provided a booster training session. This ses-
sion consisted of reminding staff the choice sequence, and the different types of choices that could be offered to consumers. Only one booster training was given during this study. Researchers returned to the study site four weeks after the last maintenance probe to collect follow-up data. The follow-up phase replicated baseline procedures.

Results

Figures 1–2 illustrate the frequency of correct and incorrect choice offerings, and the diversity of choices offered by participants to consumers. The results are presented for each staff member.

Kamal

In baseline, Kamal’s frequency of correct choices was relatively stable. He offered approximately three correct choices (range 1–5), two incorrect choices (range 1–2), and two choice types (range 1–2; refusal and termination). During choice training, Kamal actively engaged with the content and willingly discussed choice as a strategy.

Immediately after receiving skills training, Kamal displayed a slight increase in frequency of correct (M = 8) and incorrect choices (M = 5) but did not change in his diversity of choice types (M = 2). He met criteria in correctly offering choices after three days of performance feedback with an average of seven correct choices (range: 6–8). During this time, Kamal engaged the researchers with several questions regarding choice, in particular how choice may or may not be relevant depending on the consumer. Although his frequency of correct choices remained only slightly above baseline and his choice diversity remained at baseline levels (M = 2), researchers continued performance feedback and coaching until his level increased and stabilized. On the fourth day of feedback and coaching, Kamal doubled his provision of correct choices and diversity of choices.

During the maintenance phase, Kamal displayed some variability in the frequency of correct choice offerings although he continued to offer above baseline levels. His correct choice offerings and diversity of choices steadily decreased during the second half of maintenance until he reached baseline levels (day 47). Given the contratherapeutic trend of the data, the primary researcher provided Kamal with a booster session that included a description of the choice sequence, an overview of the importance of choice, and a description of the diversity of choices. The next observation (day 52), both the frequency and diversity of choices increased (from 5 correct choices to 12 correct; from 2 choice types to 3). Kamal’s data stabilized above baseline levels during the second half of maintenance to an average of nine choices (range 5–13) and his incorrect choices dropped immediately and maintained at zero levels. Approximately one month after the last maintenance observations, researchers returned to collect follow-up data. Kamal correctly provided six choices, three different choice types, and had no incorrect attempts.

Blair

During baseline, Blair offered an average of eight correct choice opportunities (range: 3–14), two incorrectly offered choices (range: 0–7), and an average of one choice type (range: 1–3). After skills training, Blair increased her correct choice offerings to an average of 33 choices (range 29–36), six choice types (range: 5–7), and only one incorrectly offered choice. She showed an increasing trend in her correct choices with a slight decrease in trend for choice diversity during intervention sessions. Her feedback and coaching sessions lasted for three days until she met criterion to move into maintenance. During maintenance sessions, Blair provided an average of 24 correct choice opportunities (range: 17–30), four choice types (range 3–5), and provided one incorrect choice. (range: 0–1) Although Blair maintained higher than baseline levels, she displayed a decreasing trend in frequency of correct choices but showed little variability in choice types. Blair resigned from the organization directly after the maintenance condition ended; therefore researchers could not collect follow-up data.

Cathy

Cathy joined the study a week after researchers collected initial baseline data on Kamal.
and Blair. During baseline, Cathy correctly provided choices an average of four times (range: 0–14), incorrectly provided choices an average of one (range: 0–3), and offered approximately one choice type (range: 0–2). After skills training, Cathy immediately increased her correct choices to 41 choice opportunities on the first day of coaching and

Figure 1. Frequency of correct and incorrect choices. Squares represent correct and circles represent incorrect.
feedback and four choice types. She provided, on average, 27 correct choices (range: 18–41), three incorrect choices (range: 0–3), and four different choice types (range: 3–5). Cathy exhibited a decreasing trend in correct choices after her first day of feedback but maintained well above her baseline levels. Cathy remained in coaching and feedback for

Figure 2. Frequency of choice types per observation.
three consecutive days until she met criterion, and then moved into maintenance. Cathy’s frequency of correct choices increased during maintenance to an average of 33 choices (range: 23–41) although she sustained offering an average of four choice types (range: 3–6). Her incorrect choices decreased to an average of zero (range: 0–2). Although Cathy displayed relatively high variability within maintenance, the average frequency of correct choices she offered between the first ($M = 34$) and second half of the maintenance phase ($M = 32$) demonstrated only a slight decrease. During follow-up, Cathy provided 12 correct choice offerings and three choice types. No incorrect choices occurred. Although Cathy surpassed her average frequency of baseline choices, she did not maintain her level of performance from maintenance to follow-up. It should be noted Cathy was undergoing medical testing at the time and this may have impacted her results.

**Social Validity**

Researchers assessed acceptability of the intervention upon conclusion of the study by having staff members complete a questionnaire using a 5-point Likert scale with higher numbers representing greater acceptability and use of choice-making. There were a total of ten questions addressing staff perceptions of the skills training, the acceptability of the skills training, the usefulness of understanding choice in their setting, and their perceptions of choice-making with consumers. Final scores ranged from 45 to 49 with a mean score of 47 out of a possible 50. In addition, Kamal made the following comment on his post-study survey “The training was very beneficial to both consumers and caregiver. I have learned new skills and the consumers are happier due to the fact that they are in charge of their daily life.”

During the follow-up visit, Kamal commented on how the choice training influenced his interactions with consumers and positively impacted all of the staff in the program. Kamal described how previously, staff left the oldest consumer in the program alone to sit on the couch all day. Since the training, Kamal noted that he and the other staff (even those not involved in the training) have begun offering the consumer activities to complete on the couch or choices of activities occurring in other locations. Cathy commented in her post-training questionnaire that all of the strategies would be more effective if the activities they had in the center were more challenging.

Prior to the start of the study, staff rated common strategies in terms of their effectiveness in working with individuals with disabilities using a Likert scale (1 = Not effective at all and 5 = Very effective). Staff rated the following strategies as most effective: offering choices (cumulative score = 15), reminder of rules (cumulative score = 14), and modeling appropriate behavior or skill (cumulative score = 13). Staff found time-out (cumulative score = 9) and verbal reprimands (cumulative score = 7) to be the least effective strategies. This questionnaire provided the researchers with a better understanding of how staff perceived common behavior strategies before receiving intervention. Staff completed an identical rating scale at the end of the study. Staff remained consistent from the initial questionnaire rating offering choices as the most effective (score = 15).

**Discussion**

All adults strive to achieve control over their lives. For adults with intellectual and developmental disabilities, the struggle to obtain control often rests in the hands of staff (Mansell et al., 2008). Researchers in the present study examined the effects of a staff-training package on the provision and diversity of choices offered by staff to adults with disabilities in a community transition program. The research team implemented a training protocol that included description, role-play, individualized coaching, and performance feedback (Salmanto & Bambara, 2000). The current study extended prior training and choice frameworks by expanding the definition of choice and including additional dependent variables to capture the types of choices offered and errors made by staff during choice sequences. Visual analysis of the data indicates a functional relationship between staff training and frequency of choices and choice types. By receiving explicit training in the importance of offering choices, participants increased their correct choice opportunities, decreased incor-
rect choices, and increased the diversity of choices within and across different activities with consumers.

The first research question sought to determine if training staff in a designated choice sequence would increase the frequency of choices staff offered consumers. Similar to recent research (e.g., Wilson et al., 2006), all participants immediately improved in their provision of correct choices. This study differentiates from prior literature in multicomponent trainings for staff in an adult services setting by the immediacy of effect and high performance of correct choices offered. In comparison with previous research, the provision of choices offered by Blair (N = 36) and Cathy (N = 41) produced a much higher frequency of correct choices than participants in similar research studies using the same or similar performance criterion (e.g., Reid et al., 2003; Salmento & Bambara, 2000).

Although all participants yielded higher levels of choice offerings after training, they also demonstrated some inconsistencies in their data from maintenance to follow-up. The frequency of choices offered by Blair and Cathy immediately after training allowed both to reach target performance within three observations, therefore they received limited coaching and feedback from researchers. Behavioral literature argues the importance of practicing key components of skills to high frequencies in order to attain fluency (Binder et al., 1996; Kubina & Yurich, 2009). The criteria set in this study may have restricted opportunities for participants, particularly Blair and Cathy, to receive ongoing practice and feedback. Given the overall variability and slight decline in correct choices within the maintenance and follow-up conditions, a higher performance criterion in conjunction with extended coaching and feedback may have facilitated all participants in achieving and sustaining a higher frequency of correct choices across the study (Reid et al., 2003).

Previous literature reported correct choices offered by staff, but did not collect or analyze data on if staff failed to follow all steps of a designated choice sequence. In addition to increasing correct choice opportunities, all participants decreased and maintained low levels of incorrect choices after training. The results of the current study provide introductory analysis of the types of errors staff make when offering choices and a comparison of the frequency correct and incorrect choices over time. When staff members did error on the choice sequence, they made the identical mistake of neglecting to wait after presenting consumers a choice. This error occurred across staff regardless of consumer or choice type.

One explanation for this finding could relate to staff perceptions of consumer needs and wants. Two participants in this study expressed that they believed consumers do not always need to be asked for what they want to do because as staff, they already know consumer preferences and routines. This anecdotal information combined with staff errors may provide evidence about how preconceived opinions of consumers may act as a barrier to choice opportunities (Cobigo et al., 2010). Although staff may recognize the importance of choice and demonstrate commitment to providing choices, subjective views of consumers may lead to erroneous choice opportunities (Cobigo et al., 2010; Stancliffe et al., 2011). Future training and research should continue to include the collection of incorrect choices and the type of errors made by staff to offer further insight into staff responding.

The final research question explored the impact of staff training on choice diversity. Building off of research by Parsons et al. (1997) and Salmento and Bambara (2000), the present study attempted to address prior restrictions in choice definitions and measurement by including choice diversity as a dependent variable. Examining choice diversity in conjunction with choice frequency provided a comprehensive evaluation of the effects of the training and allowed researchers to quantify choice in a manner better connected to the conceptual framework of choice (Brown & Brown, 2009; Stancliffe & Aber, 2000). This procedure offers an alternative method for how experimental researchers may approach choice.

In the current study, all participants increased their diversity of choices although no staff member implemented all choice types consistently. In baseline, each participant offered consumers almost exclusively refusal choices. After receiving training, participants...
increased their use of within, between, and termination choices although they continued to offer a majority of refusal choices. It is possible that staff deferred to refusal choices because this type of choice may require less planning and effort than other choice types. For example, asking, “Do you want to sit down?” involves little planning and resources versus “When do you want to eat lunch?” or “What do you want to do today?” Staff may have required more explicit training in how to include all choice types within a single activity or day.

Additional factors that may have contributed to the overall results of this study include (a) prerequisite skills of staff; (b) the addition of both choice diversity and choice sequencing during training; (c) consumer responding; (d) the culture of the disability organization. Staff may need to possess a certain level of understanding of choice and self-determination in order to increase and diversify choices. The staff members in this study all rated choice as a valuable strategy during the initial survey; therefore they may have been more receptive to implementing what they learned during training. Depending on initial screenings, future researchers may consider scaffolding training to facilitate the acquisition of the choice sequence before incorporating choice types. Training one skill to fluency and then adding subsequent skills may produce high frequency behavior that maintains over time (Binder et al., 1996). In addition, future researchers cannot dismiss the importance of organizational culture on staff behavior (Stancliffe, 2001). If an organization does not explicitly discuss choice as a proactive strategy with staff members or encourage a range of choices to be offered to consumers, researchers may face staff unwilling to engage or provide choices.

Limitations

This study had several limitations. The first limitation is the small sample involved in the study. The sample size decreases the external validity of study findings and the broader implications of results. The rapid acquisition of skills during training and feedback represents another potential limitation in this study. In the training and feedback phase, both Blair and Cathy reached criteria after three data points. This did not allow researchers to evaluate the trend or variability in the data although the immediate level change provided indication of the effectiveness of the intervention.

In addition, observations were restricted to one setting that had a limited variety of materials, activities, and locations. The opportunity to observe staff in other locations (i.e. community) may provide further information regarding staff allocation of choices. It may be important to observe participants across settings to analyze whether participants could generalize choice procedures to unknown settings and stimuli. Finally, researchers did not analyze data on consumer responses to choice. The use of audio recordings contributed to the decision to analyze only staff behavior because there was not a way to capture non-vocal consumer responses as researchers did not receive permission to video record observations. The results of this research address the impact of training on staff behavior, but do not provide further information on how choice influences consumer responses and whether an increase in choice offerings corresponds with a similar increase in consumer choice responses. Consumer responses to choice may reinforce staff behavior and enable greater understanding of consumers' voices and values, therefore future research should include consumer feedback as a dependent variable.

Conclusion

All individuals reserve the right to exercise choice in their lives. For adults with disabilities, specifically adults with significant needs, choices are often restricted in number and diversity (Agran et al., 2010; Smith et al., 2005). The results of the current study indicate that training staff in choice procedures may increase choice opportunities and diversity for adults with disabilities as well as influence staff perspectives on the impact of choice. Choices are not guaranteed nor are they often freely provided to adults with I/DD (Agran et al., 2010; Brown & Brown, 2009; Reid et al., 2007). Schools, postsecondary organizations, and researchers should continue to investigate and emphasize the positive out-
comes related to increasing consumer access to choices.

References


Received: 3 April 2019
Initial Acceptance: 10 June 2019
Final Acceptance: 30 August 2019
The eighth volume of the CEC Division on Autism and Developmental Disabilities' Prism series, Friendship 101 focuses on building social competence, friendship making, and recreation and leisure skills among students with autism spectrum disorder and other developmental disabilities. Chapters in this evidence-based, user-friendly guide address the needs of students in different developmental periods (from pre-K through young adulthood), providing teachers, parents, faculty and teacher educators with tools and strategies for enhancing the social skill development of these children and youth. Presented through an ecological perspective, together these chapters emphasize building social competence within and across school, home, and community contexts.

Available from CEC Publications
https://www.cec.sped.org/Publications

The transition from high school to adulthood is challenging for many young people, and often particularly difficult for those with disabilities. Transition To Adulthood: Work, Community, and Educational Success provides a blueprint for supporting youth with disabilities in achieving their postsecondary goals in a variety of adult settings – education and training, employment, and the community.

This publication of the Council for Exceptional Children’s Division on Autism and Developmental Disabilities is the latest in its successful Prism series (Prism 11) and covers a wide range of topics, from assessing students’ interests and abilities to fine-tuning their education plans and goals, ensuring that students with disabilities are included in a variety of settings, and building community relationships to ensure their continuing inclusion. It provides a valuable resource for transition personnel, special and general educators, and special education administrators at the school and state level, as well as adult service professionals.

With eight chapters written by 20 authors, Transition to Adulthood covers the breadth of research delineating best practices and proven instructional strategies for ensuring that students with disabilities reach their full potential and achieve their goals.

Visit the CEC online catalog at pubs.cec.sped.org to order this book. The direct link to the book listing is http://pubs.cec.sped.org/p6296/
Education and Training in Autism and Developmental Disabilities

Editorial Policy

*Education and Training in Autism and Developmental Disabilities* focuses on the education and welfare of persons with autism and developmental disabilities. *ETADD* invites research and expository manuscripts and critical review of the literature. Major emphasis is on identification and assessment, educational programming, characteristics, training of instructional personnel, habilitation, prevention, community understanding and provisions, and legislation.

Each manuscript is evaluated anonymously by three reviewers. Criteria for acceptance include the following: relevance, reader interest, quality, applicability, contribution to the field, and economy and smoothness of expression. The review process requires two to four months.

Viewpoints expressed are those of the authors and do not necessarily conform to positions of the editors or of the officers of the Division.

Submission of Manuscripts

1. Manuscript submission is a representation that the manuscript is the author’s own work, has not been published, and is not currently under consideration for publication elsewhere.


3. Each manuscript must have a cover sheet giving the names and affiliations of all authors and the address of the principal author.

4. Research studies, including experimental (group and single-subject methodologies), quasi-experimental, surveys, and qualitative designs should be no more than 20–30 typewritten, double-spaced pages, including references, tables, figures, and an abstract.

5. Graphs and figures should be originals or sharp, high quality photographic prints suitable, if necessary, for a 50% reduction in size.

6. Manuscripts along with a transmittal letter and cover sheet that gives the names, affiliations, and address of all authors should be emailed (etadd@asu.edu) to the Editor: Stanley H. Zucker, Mary Lou Fulton Teachers College, Box 871811, Arizona State University, Tempe, AZ 85287-1811.

7. Upon receipt, each manuscript will be screened by the editor. Appropriate manuscripts will then be sent to consulting editors. Principal authors will receive notification of receipt of manuscript.

8. The Editor reserves the right to make minor editorial changes which do not materially affect the meaning of the text.

9. Manuscripts are the property of *ETADD* for a minimum period of six months. All articles accepted for publication are copyrighted in the name of the Division on Autism and Developmental Disabilities.

10. Please describe subjects (or any other references to persons with disabilities) with a people first orientation. Also, use the term "intellectual disability" (singular) to replace any previous term used to describe the population of students with significant limitations in intellectual functioning and adaptive behavior as manifested in the developmental period.
22nd International Conference on Autism, Intellectual Disability and Developmental Disabilities

“Research-Informed Practice”

January 20 – 22, 2021 in Clearwater Beach, Florida

CEC’s Division on Autism and Developmental Disabilities (DADD) welcomes the submission of proposals for innovative and engaging presentations that explicitly link research to practice within the following or related topical areas:

- Evidence-based Practice
- Assessment & Instruction
- Communication & Social Skills
- Self-determination
- Life Skills & Sexuality
- Assistive Technology
- Behavioral Supports
- General Curriculum Access
- Transition Across the Lifespan
- Diversity & Inclusion
- Mental Health
- Coaching & Training

For further information, please contact Cindy Perras, DADD’s Conference Co-ordinator: cindy.perras@gmail.com.